Essays on the term structure of interest rates

Dissertation
submitted by
Anna Cieślak
for the degree of
Ph.D. in Economics
at the Faculty of Economics
University of Lugano (USI)
Switzerland

Thesis Committee (in alphabetical order):
Prof. Jules van Binsbergen, Stanford University, Graduate School of Business *and* Northwestern University, Kellogg School of Management
Prof. Patrick Gagliardini, University of Lugano (SFI)
Prof. Fabio Trojani, Chair, University of Lugano (SFI)
Prof. Pietro Veronesi, University of Chicago, Booth School of Business

August 18, 2011

To Tomek

Acknowledgment

I am indebted to many people who supported me during the preparation of this thesis. First of all, I thank my thesis advisor Professor Fabio Trojani at the University of Lugano for giving me the opportunity to do research at his Chair and inducing my interests in term structure modeling. I thank Professor Paul Söderlind (University of St. Gallen) for the inspiration he gave me at the early stage of my PhD.

Special thanks go to Professor Torben Andersen (Northwestern Kellogg) and Professor Jules van Binsbergen (Northwestern Kellogg and Stanford GSB) for many insightful discussions, their invaluable support and guidance. I am also thankful to my mentor at the University of Chicago, Professor Pietro Veronesi, for sponsoring my visit at the Booth School of Business that has proved to be a unique research experience. I gratefully acknowledge the financial support of the Swiss National Science Foundation (SNSF) that made possible my stay in Chicago.

I am particularly grateful to my co-author and friend Pavol Povala for sharing with me the passion and the desire for understanding the workings of fixed income markets.

Finally, I am still trying to find words to express gratitude to my Husband and my Parents for the unconditional and constant support they have shown during my PhD process.

Contents

1 Understanding bond risk premia **17**

1.1 Data sources . . . 23

1.2 Components in the yield curve . . . 24

1.2.A Basic example and intuition . . . 24

1.2.B Identifying the persistent component τ_t . . . 25

1.2.C Cycles as deviations from the long-run relationship between yields and short rate expectations . . . 28

1.3 The predictability of bond excess returns revisited . . . 29

1.3.A First look at predictive regressions . . . 29

1.3.B Anatomy of the cycle . . . 31

1.3.C The single returns forecasting factor: distilling the term premium frequency 33

1.3.D The Cochrane-Piazzesi factor . . . 34

1.4 Roles of factors in the cross section of yields . . . 36

1.4.A Quantifying the cross-sectional impact of factors on yields . . . 36

1.4.B Link to the level, slope and curvature . . . 38

1.5 Macroeconomic fundamentals and $\widehat{cf}$ . . . 40

1.5.A Do macro variables predict returns beyond $\widehat{cf}$? . . . 40

1.5.B What is special about the return of a two-year bond? . . . 42

1.6 Robustness . . . 44

1.6.A Out-of-sample predictability of bond returns . . . 44

1.6.B Other data sets . . . 46

1.7 Conclusions . . . 47

A **67**

A.1 Cointegration . . . 67
A.2 Data . . . 67
A.2.A Comparison of excess returns from different data sets . . . 70
A.3 Basic expression for the long-term yield . . . 71
A.4 Small sample standard errors . . . 73
A.5 Constructing the single factor . . . 74
A.5.A Exploiting information about future returns . . . 74
A.5.B One-step NLS estimation . . . 74
A.5.C Common factor by eigenvalue decomposition . . . 75
A.5.D Comparing the results . . . 75
A.6 Predictability of bond excess returns at different horizons . . . 76
A.7 Long-run inflation expectations: the persistent component . . . 76
A.7.A Model of the term structure of inflation expectations . . . 77
A.7.B Data . . . 79
A.7.C Estimation results . . . 80
A.7.D Sensitivity of predictive results to τ_t^{CPI} . . . 80
A.8 Predictability within a macro-finance model . . . 82
A.8.A Incorporating τ_t into a Taylor rule . . . 82
A.8.B Model setup . . . 84
A.8.C Model estimation . . . 84
A.8.D Predictability of bond excess returns with filtered states . . . 85
A.9 Out-of-sample tests . . . 91

2 Understanding the term structure of yield curve volatility **93**

2.1 Empirical facts about the term structure of yield volatilities . . . 99
2.1.A Data . . . 99
2.1.B Realized yield covariances . . . 100
2.1.C Information in second moments of yields . . . 102
2.2 The model . . . 105
2.2.A Discussion . . . 108
2.3 Model estimation . . . 109
2.3.A Transition dynamics . . . 109

2.3.B Measurements . . . 110
2.3.C Pseudo-maximum likelihood estimation . . . 111
2.4 Estimation results . . . 112
2.4.A Model performance . . . 112
2.4.B Filtered states . . . 116
2.4.C Short and long-end volatility: two episodes . . . 117
2.4.D Are volatility factors revealed by the yield curve? . . . 118
2.4.E Filtered factors versus principal components . . . 119
2.5 Yield volatility states and macroeconomic conditions . . . 120
2.5.A Expectations, uncertainties, and the duration of volatilities . . . 121
2.6 Yield volatility and market-wide liquidity . . . 124
2.7 Discussion and robustness . . . 126
2.7.A Realized covariance matrix estimation . . . 126
2.7.B Alternatives to the realized covolatility estimator . . . 127
2.7.C Specification of bond premiums . . . 128
2.8 Conclusions . . . 130

B **155**

B.1 Data appendix . . . 155
B.1.A GovPX and BrokerTec . . . 155
B.1.B Testing for microstructure noise . . . 156
B.1.C Extracting zero coupon yield curve from high-frequency data . . . 156
B.1.D Survey data . . . 157
B.2 Technical appendix . . . 158
B.2.A Moments of the V_t process . . . 158
B.2.B Moments of the Y_t dynamics . . . 159
B.3 General solution . . . 162
B.3.A Dependence between X and V factors . . . 162
B.3.B General form of the market prices of risk . . . 162
B.3.C Solution for bond prices . . . 163
B.3.D Instantaneous volatility of yields . . . 164
B.3.E Conditional covariance of X and V . . . 165
B.3.F Discrete approximation to the unconditional covariance matrix of X and V 165

B.4 Transition dynamics . . . 166
B.5 Identification . . . 167
B.6 Estimation Appendix . . . 168

3 Correlation risk and the term structure of interest rates 173
3.1 The Economy . . . 177
3.1.A The Short Interest Rate and the Market Price of Risk . . . 179
3.1.B The Term Structure of Interest Rates . . . 181
3.1.C Yields . . . 183
3.1.D Excess Bond Returns . . . 185
3.1.E Forward Interest Rate . . . 186
3.1.F Interest Rate Derivatives . . . 187
3.2 The Model-Implied Factor Dynamics . . . 188
3.2.A Empirical approach . . . 188
3.2.B Properties of Risk Factors . . . 192
3.2.C Factors in Yields . . . 194
3.3 Yield Curve Puzzles . . . 195
3.3.A Excess Returns on Bonds . . . 196
3.3.B The Failure of the Expectations Hypothesis . . . 196
3.3.C Second Moments of Yields . . . 199
3.3.D Aspects in Derivative Pricing . . . 202
3.4 Extensions . . . 204
3.4.A Forward-rate factor . . . 204
3.4.B Conditional Hedge Ratio . . . 207
3.4.C Unspanned Factors . . . 209
3.5 Conclusions . . . 211

C 227
C.1 Proofs: Term Structure of Interest Rates . . . 227
C.1.A Second Moments of the Short Interest Rate . . . 227
C.1.B Proof of Proposition 3: Solution for the Term Structure of Interest Rates . . . 228
C.1.C Bond Returns . . . 231
C.1.D Dynamics of the Forward Rate . . . 232

C.1.E Pricing of Zero-Bond Options . . . 233
C.2 Relation to the Quadratic Term Structure Models (QTSMs) . . . 236
C.2.A Unique invertibility of the state . . . 238
C.3 Moments of the factors and yields . . . 238
C.3.A Moments of the discrete time Wishart process . . . 238
C.3.B Moments of yields . . . 243
C.4 Useful Results for the Wishart Process . . . 244
C.5 Details on the Estimation Approach . . . 247
C.5.A The 2×2 Model . . . 247
C.5.B The 3×3 Model . . . 248

List of Figures

1.1 Fit of the modified Taylor rule (OLS) . . . 49
1.2 The persistent factor, τ_t^{CPI} . . . 50
1.3 The anatomy of the cycle . . . 51
1.4 Single factor . . . 52
1.5 Contributions of τ_t^{CPI}, $c_t^{(1)}$ and $\widehat{cf}_t$ to explained variance of PCs . . . 52
1.6 Comparing the R^2's . . . 53
1.7 Decomposing the Cochrane-Piazzesi factor . . . 54
1.8 Cross-sectional impact of factors on yields . . . 55
1.9 Comparing the cross-sectional impact of factors and PCs . . . 56

A.1 Comparison of realized excess returns across data sets . . . 71
A.2 Long-horizon inflation expectations . . . 81
A.3 Sensitivity of the predictability evidence to v and N parameters . . . 87
A.4 Sensitivity of predictability evidence to the window size . . . 88
A.5 Macro-finance model: filtered yield curve factors . . . 89
A.6 Macro-finance model: fit to yields . . . 90

2.1 Yield and volatility curves . . . 132
2.2 Evolution of yields and volatilities . . . 133
2.3 Realized correlations of yield factors . . . 134
2.4 Factors in yield volatilities . . . 135
2.5 Yield level versus yield volatility . . . 136
2.6 Regressions of realized volatility on the yield level . . . 137
2.7 In-sample fit to yields . . . 138
2.8 In-sample fit to the second moment dynamics of yields . . . 139

2.9 Filtered factor dynamics . . . 140
2.10 Factor shocks and yield curve responses . . . 141
2.11 Contribution of X_t and V_t to model-implied yields . . . 142
2.12 Impulse-response functions of volatility factors V_{11} and V_{22} to macroeconomic uncertainty proxies . . . 143

B.1 Volatility signature plot . . . 157

3.1 Term spreads and yield volatilities during 1994/95 and 2004/05 tightenings . . . 214
3.2 Instantaneous correlations of positive factors for different k's in the 2×2 WTSM . 215
3.3 Unconditional distribution of the 5-year yield for different k in the 2×2 WTSM . 216
3.4 Actual and model-implied unconditional distributions of the 5-year yield . . . 217
3.5 Loadings of yields on principal components: 2×2 WTSM vs. data . . . 218
3.6 Variance explained by principal components, conditional on factor correlation . . . 219
3.7 Conditional principal components of yields in the 2×2 WTSM . . . 220
3.8 Properties of expected excess bond returns in the 2×2 WTSM . . . 221
3.9 Properties of realized excess bond returns . . . 222
3.10 Campbell-Shiller regression coefficients . . . 223
3.11 Term structure of forward interest rate volatilities . . . 224
3.12 Term structure of cap implied volatilities . . . 225
3.13 Cochrane-Piazzesi projection coefficients . . . 226

List of Tables

1.1 Modified Taylor rule (OLS) . . . 49
1.2 Estimates of the vector error correction model . . . 57
1.3 Bond excess returns: summary statistics . . . 58
1.4 First look at predictive regressions of bond returns . . . 59
1.5 Predicting returns with the single forecasting factor . . . 60
1.6 The link between the level and the return forecasting factor . . . 61
1.7 Decomposing the forward-rate predictive regressions . . . 62
1.8 Marginal predictability of bonds excess returns by macro and liquidity factors . . . 63
1.9 Macro risks and predictability: the case of the two-year bond . . . 64
1.10 Out-of-sample tests . . . 65
1.11 Comparing predictive R^2 in different data sets . . . 66

A.1 Unit root test . . . 68
A.2 Comparison of one-year holding period excess returns: CMT, GSW and FB data . . . 70
A.3 Comparing alternative constructions of the single factor . . . 76
A.4 Predictability of bond excess returns across horizons . . . 77
A.5 Bond premia predictability by filtered states s_t and f_t from the macro-finance model 86

2.1 Descriptive statistics of weekly yields and realized volatilities . . . 144
2.2 Parameter estimates . . . 146
2.3 In-sample model fit . . . 147
2.4 Out-of-sample yield forecasting . . . 148
2.5 Regressions of yield realized second moments on filtered states . . . 149
2.6 Regressions of filtered states on yield curve principal components . . . 150

2.7 Major moves in the U.S. Treasury yield and volatility curves in the period 2000–2004 (1st part) . . . 151
2.7 Major moves in the US Treasury yield and volatility curves in the period 2000–2004 (2nd part) . . . 152
2.8 Regressions of filtered states on macroeconomic surveys . . . 153
2.9 Liquidity and the volatility curve . . . 154

B.1 Average number of quotes/trades per day in the GovPX and BrokerTec databases 155
B.2 Autocorrelation of high-frequency yield changes . . . 156
B.3 Correlation of zero coupon yields with CMT and GSW yields . . . 157

3.1 Regressions of the yield changes onto the slope of the term structure . . . 198
3.2 GARCH(1,1) parameters for the model-implied and historical 5-year yield . . . 201
3.3 Single forward rate factor regressions . . . 207
3.4 Properties of the conditional hedge ratio . . . 208
3.5 Factors in the derivatives market . . . 210

C.1 Quadratic versus Wishart factor model . . . 237
C.2 Fitting errors for the 2×2 model . . . 248
C.3 Fitting errors for the 3×3 model . . . 249

Introduction

This doctoral thesis focuses on fixed income markets. I explore two basic questions in this area: (*i*) the risk compensation for holding Treasury bonds (bond risk premia, expected returns) and (*ii*) the behavior of interest rate volatilities across different maturities. My objective is to understand the effect of changes in risk premia and in yield volatilities on bond prices, and to propose a set of new modeling approaches to accommodate these features. As such, my work relies both on empirical and theoretical methods.

These research interests are reflected in three chapters of this dissertation. The question of bond risk premia and bond return predictability is the central theme of Chapter 1: **"Understanding bond risk premia."** Chapter 2 **"Understanding the term structure of yield curve volatility"** analyzes the dynamic features of yield volatilities and undertakes their modeling. Both chapters are part of a broader research agenda on bond markets, which I pursue in collaboration with Pavol Povala. Finally, Chapter 3 **"Correlation risk and the term structure of interest rates,"** based on joint work with Andrea Buraschi and Fabio Trojani, explores term structure models, especially their flexibility in matching the first and second moments of yield distribution, from the perspective of the state space that characterizes the yield factor dynamics.

The behavior of expected excess bond returns and their relationship with the real economy has long been an active area of research (Fama and Bliss, 1987; Campbell and Shiller, 1991). While traditional yield curve models use principal components as a convenient statistical representation of yields (Litterman and Scheinkman, 1991), recent evidence suggests that bond risk premia are driven by economic forces that cannot be fully captured by the level, slope and curvature (Cochrane and Piazzesi, 2005, 2008).

Thus, the literature has evolved two approaches to understanding risk premia on bonds. The first approach suggests that the yield curve itself contains a small component that is hard to detect in the cross-section of yields, but has a power for forecasting future bond returns. This

important variable reveals itself through a particular combination of forward rates or through higher-order principal components (Cochrane and Piazzesi, 2005; Duffee, 2011). The second approach maintains that macroeconomic variables such as real activity, unemployment or inflation contribute to the predictability of bond returns beyond what is explained by the yield curve itself (Ludvigson and Ng, 2009; Joslin, Priebsch, and Singleton, 2010). Combining the two domains into a coherent view of risk premia and yields continues to present an important open question.

Chapter 1 provides strong evidence that a separation between the two domains—yield curve factors and macro variables—is neither empirically justified nor needed from a theoretical point of view. The novel idea behind this result is a decomposition of the yield curve into two economic frequencies: sluggish generational adjustments related with long-horizon inflation expectations, and higher-frequency transitory fluctuations—cycles. Predictive regressions of one-year excess bond returns on a common factor constructed from the cycles give R^2's up to 60% across maturities. The result holds true in different data sets, passes a range of out-of-sample tests, and is not sensitive to the inclusion of the monetary experiment (1979/83), or the recent crisis (2007/09). We identify a simple economic mechanism that underlies this robust feature of the data: Cycles represent deviations from the long-run relationship between yields and the slow-moving component of expected inflation.

This single observation extends to a number of new insights. First, we show that a key element for return predictability is contained in the first principal component of yields—the level. Once we account for this information, there is surprisingly little we can learn about term premia from other principal components. Second, we interpret the standard predictive regression using forward rates—the Cochrane-Piazzesi regression—as a constrained case of a more general return forecasting factor that could have been constructed by bond investors in real time. Third, using a simple dynamic term structure model, we quantify the cross-sectional impact of that encompassing factor on yields. We find that the factor has a nontrivial effect on yields which increases with the maturity of the bond. Finally, conditional on those findings, we revisit the additional predictive content of macroeconomic fundamentals for bond returns. By rendering most popular predictors insignificant, our forecasting factor aggregates a variety of macro-finance risks into a single quantity.

Chapter 2 studies the structure, economic content and pricing implications of the fluctuating covariance matrix of interest rates. Using almost two decades of high-frequency bond data, we obtain a so far unexplored view of the links between the yield and volatility states, and their interactions with macro conditions and liquidity.

Recent interest rate environment in the US and worldwide emphasizes the role for a consistent modeling of the yield and volatility curves across maturities (Kim and Singleton, 2010). Now that the short-term interest rates have been close to zero for two years, long-term interest rates have become increasingly autonomous, with correlation between the three-month and the ten-year yield hitting lowest scores on record. Having reached heights not seen since the Volcker period, yield volatility continues to display a considerable degree of cross-sectional variation: Indeed, even though the crystalizing near-term outlook for the monetary policy path has tamed the front-end of the curve, it has also raised new questions about the long-run policy impact on the economy, thus leading to an unprecedented unease at the long maturity range.

Interest rate volatility provides complementary information about the economic landscape that cannot be learned from observing yields at infrequent intervals. We decompose the dynamics of the two curves within a comprehensive yield curve model, and identify volatility factors with the support of realized covolatility proxies and filtering. The model separates three economically distinct volatility states: (*i*) an erratic short-end state, (*ii*) a smooth long-end state, and (*iii*) a covolatility factor capturing interactions between the long and the intermediate region of the curve. Those components reveal the duration structure of economic uncertainties concerned with the monetary policy, inflation and real activity, as well as different aspects of liquidity. At the short-end, yield volatility is related with temporary dry-ups in market-wide liquidity (Hu, Pan, and Wang, 2010); at the long end, instead, it reflects a more persistent variation in funding liquidity associated with the tightness of financial conditions (Fontaine and Garcia, 2010).

Chapter 3 studies the implications of a "completely affine" term structure model that introduces a new element of flexibility in the joint modeling of market prices of risk and conditional second moments of risk factors. This approach allows for stochastic correlation among the priced risk factors, for a market price of risk that can be negative in some states of the world, and for a simple equilibrium interpretation.

A vast literature has explored the ability of term structure models to account for the time-series and cross-sectional properties of bond market dynamics. The research has focused on analytically tractable models that ensure economically meaningful behavior of yields and bond returns. This combination of theoretical and empirical requirements poses a significant challenge. In affine term structure models (ATSMs), for instance, the tractability in pricing and estimation comes with restrictions that guarantee admissibility of the underlying state processes and their econometric identification.

In order to match the physical dynamics of the yield curve, reduced-form models have exploited different specifications of the market price of risk whose most prominent examples are known as "completely affine" (Dai and Singleton, 2000), "essentially affine" (Duffee, 2002) and "extended affine" (Cheridito, Filipovic, and Kimmel, 2007). While these extensions have proved increasingly successful in fitting the data, they are not innocuous in terms of investors' preferences they imply.

Therefore, Chapter 3 takes a different approach focussing on a non-standard assumption about the state space rather than preference structure. Specifically, we assume the risk factors in the economy to follow a continuous-time affine process of positive definite matrices whose transition probability is Wishart. By construction, such factors allow correlations and volatilities both to be stochastic. This approach builds on the work of Gourieroux and Sufana (2003) who propose the Wishart process as a convenient theoretical framework to represent yield factors. We start from their insight and take the first attempt to investigate the properties of a continuous-time Wishart yield curve model. Exploring the properties of the state space, we do not introduce a new form of the market price of risk, but rather we resurrect the parsimonious completely affine class. We study the properties of the term structure, bond returns and standard interest rate derivatives, and document the ability of this setting to match several features of the data.

Specifically, we demonstrate that a parsimonious three-factor specification of the model replicates empirical regularities (Piazzesi, 2003) such as the predictability of excess bond returns, the persistence of conditional volatilities and correlations of yields, or the hump in the term structure of forward rate volatilities and implied volatilities of caps. Furthermore, we show that the model can accommodate such features of the bond market such as the Cochrane-Piazzesi (2005) return-forecasting factor, or the unspanned dynamics of interest rate derivatives (Collin-Dufresne and Goldstein, 2002).

The structure of this dissertation does not follow the chronological order in which the chapters were written. In fact, I have deliberately inverted the chronology: To reflect the generality of questions that are tackled in subsequent chapters, I start with the most broadly studied one—of expected bond returns or first moments of yields. Tables and figures are collected at the end of each chapter, so are the appendices that contain additional results, derivations and proofs.

Chapter 1

Understanding bond risk premia

Understanding the behavior of expected excess bond returns[1] and their relationship with the economy has long been an active area of research. Many popular models of the yield curve are motivated by the principal components (PCs) as a convenient and parsimonious representation of yields. However, recent evidence suggests that bond premia are driven by economic forces that cannot be fully captured by the level, slope and curvature alone.[2]

One way of modeling yields and term premia jointly, then, is to augment the standard trio of the yield curve factors with additional variables that forecast returns. Such models provide a tractable framework for thinking about the dynamics and the sources of risk compensation in the bond market, but they also implicitly take as given the assumption that a separation between the cross-sectional variation in yields and the variation in expected bond returns is needed.

Term premium factors come in at least two forms. First, the yield curve itself seems to contain a component that, being hard to detect in the cross-section, has a strong forecasting power for future bond returns. This important variable reveals itself through a particular combination of forward rates or through higher-order principal components, thus making its economic interpre-

[1]This chapter is based on the paper under the same title written in collaboration with Pavol Povala from the University of Lugano. Part of this research was conducted when I was visiting the University of Chicago Booth School of Business. We thank Torben Andersen, Ravi Bansal, Jules van Binsbergen, Greg Duffee, Jean-Sébastien Fontaine, Ralph Koijen, Arvind Krishnamurthy, Robert McDonald, Kenneth Singleton, Ivan Shaliastovich, Fabio Trojani, Pietro Veronesi, Liuren Wu, and seminar participants at the NBER Asset Pricing Meetings, WFA Meetings, Stanford GSB, Columbia Business School, Berkeley Haas, Northwestern Kellogg, Toronto Rotman, NY Fed, Fed Board, Blackrock, University of Texas at Austin McCombs, Dartmouth Tuck, Boston University, Economic Dynamics Working Group at the University of Chicago, University of Lugano, Bank of Canada, University of Geneva, HEC Lausanne SFI, and NCCR Finrisk Review Panel Zurich for comments.

[2]Whenever we label factors as the "level", "slope", and "curvature", we refer to the first three principal components of the yield curve.

tation complicated. Second, and independently, macroeconomic variables such as real activity, unemployment or inflation appear to contribute to the predictability of bond returns beyond what is explained by factors in the curve. Combining these two domains into a coherent view of term premia and yields continues to present an important open question. This is the question we address with the current paper.

We propose a new approach to analyzing the linkages between factors pricing bonds and those determining expected bond returns. A crucial observation is that interest rates move on at least two different economic frequencies. Specifically, we decompose the yield curve into a persistent component and shorter-lived fluctuations particular to each maturity, which we term cycles. The persistent component captures smooth adjustments in short rate expectations that may take decades to unfold, and are related both economically and statistically with the shifting long-run mean of inflation. To provide a measurement that is instantaneously available to investors, our approach remains intentionally simple: Borrowing from the adaptive learning literature, we proxy for the persistent factor using the discounted moving average of past core inflation data. This single variable explains 87% of variation in the ten-year yield. Cycles, as we show, represent stationary deviations from the long-term relationship between yields and that slow-moving factor.

Working from the basic notion of a n-period yield ($y_t^{(n)}$) as the sum of short rate (r_t) expectations and the risk premium ($rpy_t^{(n)}$) (Appendix A.3):

$$y_t^{(n)} = \frac{1}{n} E_t \sum_{i=0}^{n-1} r_{t+i} + rpy_t^{(n)}, \tag{1.1}$$

we exploit the cross-sectional composition of the cycles to construct a powerful predictor of excess bond returns. The underlying economic intuition is as follows: Being derived from a one-period risk-free bond, the cycle with the shortest maturity inherits stationary variation in short rate expectations but not in premia. As maturity increases, however, the transitory short rate expectations subside, and the variation in premia becomes more apparent. In combination, we are able to trace out a term structure pattern of risk compensation throughout the yield curve. This result serves to unearth new findings along three dimensions: (*i*) attainable bond return predictability, (*ii*) cross-sectional effect of risk premia on bond prices, and (*iii*) macroeconomic risks in term premia.

We start by revisiting the empirical predictability of bond excess returns. From cycles, we construct a common factor that forecasts bond returns for all maturities. We label this factor $\widehat{cf}$.

Predictive regressions of one-year excess bond returns on $\widehat{cf}$ give R^2's up to 60% in the period 1971–2009. Given the typical range of predictive R^2's between 30–35%, the numbers we report may appear excessive. Identifying the source of this improvement, we find that the standard level factor of yields combines distinct economic effects—short rate expectations and term premia—into one variable. We distill these effect into three economic frequencies: generational frequency related to persistent inflation expectations, business cycle frequency related to transitory short rate expectations, and the term premium frequency.

As a consequence of this view, we are able to discern the mechanism that makes forward rates a successful predictor of bond excess returns. We show that the commonly used forward rate factor (Cochrane and Piazzesi, 2005) can be interpreted as a specific linear combination of interest rate cycles, whose predictive power is constrained by the persistence of yields. If the information set of the market participants contains only past history of forward rates, then the forward rate factor is the best measure of the term premia that both econometricians and investors could obtain. However, because our proxy for the persistent inflation expectations in yields is known in real time, $\widehat{cf}$ provides a viable benchmark for the attainable degree of bond return predictability.

How does the return-forecasting reveal itself in the cross-section of yields? To answer this question, we project yields on three observable variables: the persistent and transitory factors underlying the short rate expectations, and the term premium factor, $\widehat{cf}$. These three factors explain on average 99.7% of variation in yields for maturities of one year through 20 years, compared to 99.9% captured by the traditional level, slope and curvature. The deterioration in the fit relative to the PCs comes with the benefit of an economic interpretation. The persistent short rate expectations component propagates itself uniformly across maturities, mimicking the impact of the usual level factor. The effect of the transitory short rate expectations decays with the maturity of yields, and is superseded by an increasing importance of the term premium factor $\widehat{cf}$. Notably, we find that variation in the term premium is reflected in the cross-section of yields. The one standard deviation change in $\widehat{cf}$ induces an average response of 54 basis points across the yield curve. This number exceeds the comparable impact of both the slope and the curvature.

One is ultimately interested in understanding the link between the term premia and macro-finance conditions. Taking $\widehat{cf}$ as a benchmark, we can assess the marginal predictive content of macroeconomic fundamentals for bond returns. The presence of $\widehat{cf}$ in the predictive regression renders most macro-finance variables insignificant, suggesting that our factor successfully aggregates a variety of economic risks into a single quantity. With a comprehensive set of macro-finance predictors, we are able to increase the R^2's relative to $\widehat{cf}$ just by two percentage points

at maturities from five to 20 years, and by five percentage points at the two-year maturity. This evidence points to a heterogeneity of economic factors driving term premia. Moreover, the half-life of $\widehat{cf}$ of about ten months suggests that term premia vary at a frequency higher than the business cycle. While correlated, many of the large moves in bond excess returns and in $\widehat{cf}$ appear in otherwise normal times, giving rise to an interest rate-specific cycle.

As an interesting by-product of this analysis, we emphasize the particular role of two key macroeconomic variables, unemployment and inflation, for predicting realized bond returns at the shortest maturities. Decomposing the realized excess return on a two-year bond into the expected return and the forecast error that investors make about the future path of monetary policy, we attribute the additional predictive power of fundamentals to the latter component. As such, unexpected returns suggest themselves as one possible channel through which fundamentals can predict realized excess bond returns at short maturities.

We illustrate the merit of our approach with an example of a slightly modified Taylor rule. Imagine that the Fed sets the policy rule having a similar decomposition in mind to the one we propose. Specifically, suppose that investors and the Fed alike perceive separate roles for two components of the inflation process: the slow moving long-run expectation of core inflation (τ_t^{CPI}), and its cyclical fluctuations (CPI_t^c). The transient inflation is controlled by the monetary policy actions. In contrast, the market's conditional long-run inflation forecast, τ_t^{CPI}, is largely determined by the central bank's credibility and investors' perceptions of the inflation target. Beside the two components of inflation, assume that unemployment, $UNEMPL_t$, is the only additional factor that enters the policy rule. How well are we able to explain the behavior of the Fed funds rate in the last four decades? Is the separation between τ_t^{CPI} and CPI_t^c ("the modified rule") more appealing than the Taylor rule that uses inflation as a compound number ("the restricted rule")? Figure 1.1 plots the fit of the modified rule for the 1971–2009 and 1985–2009 period, and Table 1.1 juxtaposes its estimates with the standard rule.

[Table 1.1 and Figure 1.1]

By comparing the respective R^2's, our decomposition does well in explaining the behavior of the short interest rate. The modified rule explains 79%, 61% and 91% of variation in the short rate, respectively, in the full 1971–2009 sample, 70s-to-mid-80s and post-Volcker samples relative to 56%, 30% and 75% captured by the standard rule in the same periods. This fit is remarkably good given that it is obtained from a small set of macro fundamentals only. Most importantly, the estimated coefficients in the modified rule are stable across the three periods, while those of

the restricted rule are not.[3] This observation suggests that the two types of economic shocks—transitory versus persistent—play different roles in determining interest rates. Disentangling them provides the basis for our conclusions about the linkages between term premia and the yield curve.

Related literature

An important part of the term structure literature has focused on studying the predictability of bond returns. Cochrane and Piazzesi (2005, CP) have drawn attention to this question by showing that a single linear combination of forward rates—the CP factor—predicts bond excess returns across a range of maturities. Importantly, that factor has a low correlation with the standard principal components (PCs) of yields. To uncover macroeconomic sources of bond return predictability, Ludvigson and Ng (2009) exploit information in 132 realized macroeconomic and financial series. The main PCs extracted from this panel are statistically significant in the presence of the CP factor and substantially improve the predictability. In a similar vein, Cooper and Priestley (2009) show that the output gap helps predict bond returns. Applying a statistical technique of supervised adaptive group lasso, Huang and Shi (2010) argue that the predictability of bond returns with macro variables is higher than previously documented. Recently, Fontaine and Garcia (2010) show that a factor identified from the spread between on- and off-the-run Treasury bonds drives a substantial part of bond premia that cannot be explained by the traditional PCs, nor the CP factor. In contrast to those studies, we focus on explaining variation of bond excess returns using a predictor, $\widehat{cf}$, formed from the basic zero yield curve and one inflation variable that plays the role of a level factor in yields. We show that the $\widehat{cf}$ factor encompasses many usual predictors of bond returns, and we are able to reconcile this result with the predictive power of forward rates.

Recent literature extends the classical Gaussian macro-finance framework of Ang and Piazzesi (2003) to study bond premia. Duffee (2007) develops a model with a set of latent factors impacting only the premia and studies their links to inflation and growth. Joslin, Priebsch, and Singleton (2010, JPS) propose a setting in which a portion of macro risks, related to inflation and real activity, is unspanned by the yield curve, but has an impact on excess returns. Wright (2009) studies international term premia within the JPS setup and relates much of the fall in forward rates to decreasing inflation uncertainty. Similarly, Jotikasthira, Le, and Lundblad (2010) apply the JPS setting to model the co-movement of the term structures across currencies with risk

[3]The instability of the Taylor rule coefficient is well documented in a number of studies, see e.g. Ang, Dong, and Piazzesi (2007), Clarida, Galí, and Gertler (2000).

premia being one of the channels. To account for the variation in the term premia, authors have gone beyond the standard three-factor setup. Cochrane and Piazzesi (2008) integrate their return-forecasting factor together with the level, slope and curvature into an affine term structure model. [4] Duffee (2011) estimates a five-factor model, and extracts a state that is largely hidden from the cross-section of yields but has an effect on future rates and excess bond returns. In our setting, three observable factors are enough to account jointly for the variation in premia and in yields. Especially, we demonstrate that all three, including the return-forecasting factor, play a role in explaining the cross-section of yields.

The identification of the persistent component in yields has attracted attention in the earlier literature. Roma and Torous (1997) study how real interest rates vary with the business cycle. They view business cycle as stationary deviations from a stochastic trend. Accounting for the trending and cyclical components in real consumption improves the fit of a consumption-based model to real returns on short-maturity bills. As a source of persistence in yields, Kozicki and Tinsley (1998, 2001a,b) point to sluggish changes in the market perceptions of the long-run monetary policy target for inflation. They introduce the concept of shifting endpoints that describe the behavior of the central tendency in long-term yields. From a methodological perspective, shifting endpoints reconcile observed long-term yields with the limiting behavior of conditional short rate forecasts.

In a related fashion, Fama (2006) shows that the predictability of the short rate for horizons beyond one year comes from its reversion toward a time-varying rather than constant long-term mean, which he proxies with a moving average of past one-year yield. Following similar intuition, several authors adopt slow-moving means of variables to generate persistent long-term yields. Important examples include reduced-form models of Rudebusch and Wu (2008), Dewachter and Lyrio (2006), Orphanides and Wei (2010), and Dewachter and Iania (2010) or a structural setting with adaptive learning as proposed by Piazzesi and Schneider (2011). Koijen, Van Hemert, and Van Nieuwerburgh (2009) extract the term premium as the difference between the long-term yield and the moving average of the past short rate to study mortgage choice. To the best of our knowledge, our study is the first to establish the link between long-horizon inflation expectations, persistent and transitory short rate expectations, and the predictability of bond excess returns.

[4] Their study emphasizes a particularly parsimonious form of market prices of risk: While bond premia move with the return-forecasting factor, they compensate only for the level shocks. The distinction between the physical (premia) and risk-neutral (pricing the cross-section) dynamics in those models is thoroughly discussed in Joslin, Singleton, and Zhu (2011).

1.1 Data sources

We use end-of-month yield data obtained from the H.15 statistical release of the Fed. Since we want to cover a broad spectrum of maturities over a possibly long sample, we consider constant maturity Treasury (CMT) yields. The available maturities comprise six months and one, two, three, five, seven, ten and 20 years in the post Bretton Woods period from November 1971 through December 2009. We bootstrap the zero coupon curve by treating the CMTs as par yields. In Appendix A.2.A, we provide a comparison of our zero curve and realized excess bond returns with other data sets (Fama-Bliss and Gürkaynak, Sack, and Wright (2006)); additionally, the robustness section discusses the sensitivity of the predictive results across these data sets. The comparison confirms a very close match between the different data sets at maturities that overlap. However, for the core results of this paper we rely on the CMT zero curve to account for the information that long maturity bonds contain about the term premia.

The inflation data which we use to construct the persistent component is from the FRED database. We use core CPI, which is not subject to revisions and excludes volatile food and energy prices. There are two main reasons for using core CPI rather than the CPI including all items. First, core CPI has been at the center of attention of the monetary policy makers.[5] Second and related, it is more suitable to compute the long-run expectations of inflation by excluding volatile components of prices. Nevertheless, we verify that our results remain robust to both core and all-items CPI measures.

Inflation data for a given month are released in the middle of next month. To account for the publication lag, when constructing the persistent inflation component we use the CPI data that are available as of month end. For example, the estimate of the persistent component for April 2000 uses inflation data until March 2000 only. We also check that our results are not sensitive to whether or not we allow the publication lag. Appendix A.2 provides additional details about the data we use in the subsequent analysis.

[5]Fed officials rely on core inflation to gauge price trends. As one recent example, this view has been expressed by the Fed chairman Ben Bernanke in his semiannual testimony before the Senate Banking Committee on March 2, 2011: "Inflation can vary considerably in the short run. [...] Our objective is to hit low and stable inflation in the medium term." Core inflation is a good predictor of the overall inflation over the next several years, which is the horizon of focus for the monetary policy makers.

1.2 Components in the yield curve

1.2.A Basic example and intuition

We motivate our decomposition with a stylized example. The yield of an n-period bond can be expressed as the average expected future short rate r_t and the term premium, $rpy_t^{(n)}$ (assuming log normality, see Appendix A.3). Reiterating equation (1.1):

$$y_t^{(n)} = \frac{1}{n} E_t \sum_{i=0}^{n-1} r_{t+i} + rpy_t^{(n)}. \tag{1.2}$$

Suppose that the short rate is determined according to:

$$r_t = \rho_0 + \rho_\tau \tau_t + \rho_x x_t, \tag{1.3}$$

where $\rho_0, \rho_x, \rho_\tau$ are constant parameters, and τ_t and x_t are two generic factors that differ by their persistence. Specifically, assume for simplicity that τ_t is unit root and x_t has quickly mean reverting stationary AR(1) dynamics with an autoregressive coefficient ϕ_x and standard normal innovations ε_{t+1}^x: $x_{t+1} = \mu_x + \phi_x x_t + \sigma_x \varepsilon_{t+1}^x$. We label τ_t as the generational frequency, and x_t as the business cycle frequency.

Solving for the expectations in (1.2), it is convenient to represent the n-period yield as:

$$y_t^{(n)} = b_0^{(n)} + b_\tau^{(n)} \tau_t + b_x^{(n)} x_t + rpy_t^{(n)}, \tag{1.4}$$

where $b_0^{(n)}$ is a maturity dependent constant, $b_\tau^{(n)} = \rho_\tau$ and $b_x^{(n)} = \frac{1}{n}\rho_x \left(\phi_x^n - 1\right)\left(\phi_x - 1\right)^{-1}$. We will refer to the sum of the transitory short rate expectations and the risk premium in (1.4) simply as "the cycle," defined as:

$$\tilde{c}_t^{(n)} = b_x^{(n)} x_t + rpy_t^{(n)}. \tag{1.5}$$

The composition of $\tilde{c}_t^{(n)}$ changes with the maturity of the bond. For one-period investment horizon, $n = 1$, $\tilde{c}_t^{(1)}$ captures variation in short rate expectations ($b_x^{(1)} x_t$), but not in premia because $rpy_t^{(1)}$ is zero in nominal terms. As the maturity n increases, the transitory short rate expectations decay due to the mean reversion in the dynamics of x_t. Thus, cycles extracted from the long end of the yield curve should provide the most valuable information about expected excess returns. This intuition underlies the predictability of bond returns that we document below.

A reduced-form specification for the short rate in spirit of equation (1.3) has been discussed by Fama (2006) and Rudebusch and Wu (2008), among others. It is compatible with models that explicitly account for the short rate persistence. First, τ_t can be interpreted as a level factor reflecting movements in the Fed's inflation target; x_t captures the endogenous response of the Fed to business cycle fluctuation in risks (Atkeson and Kehoe, 2008). Second, in an asymmetric information setup, τ_t can be seen as an outcome of investors' learning process about the unobserved Fed's inflation target (Kozicki and Tinsley, 2001a; Gürkaynak, Sack, and Swanson, 2005). Third, a persistent component τ_t associated with the trend in inflation can be generated in a New Keynesian model with credible central bank and symmetric information (Goodfriend and King, 2009). We think of (1.3) in the following context: In setting the policy rate, the Fed watches slow-moving changes in the economy that take place at a generational frequency, i.e. those spanning several decades such as central bank credibility, demographic changes, or changes in the savings behavior. At the same time, it also reacts to more cyclical swings reflected in the transitory variation of unemployment or realized inflation.[6] As shown in the Introduction, the Taylor rule that distinguishes between these two frequencies is able to explain a large part of variation in the US Fed funds rate over the last four decades. For completeness, in Appendix A.8 we estimate and study the implications of a macro-finance term structure model that incorporates such a Taylor rule.

Before we move on, in the remainder of this section we label τ_t, discuss more formally its relation to yields, and describe our strategy for identifying cycles.

1.2.B Identifying the persistent component τ_t

As summarized above, the literature suggests that inflation and especially the movements in its long-run mean have been a major determinant of the persistent rise and decline of US yields in the last four decades. Such results are intuitive in an economy characterized by fiat money, and one that did not experience other significant permanent shocks.[7] To accommodate the slow moving nature of the long run inflation expectations, we borrow from the extensive literature on adaptive learning in macroeconomics (e.g., Branch and Evans, 2006; Evans and Honkapohja, 2009). We make the common assumption that the data generating process for annual inflation

[6] This interpretation is consistent with the so-called Jackson Hole pre-crisis consensus on monetary policy, as recently summarized by Bean, Paustian, Penalver, and Taylor (2010), and referred to by Clarida (2010).

[7] In general, the yield curve can be subject to permanent shocks stemming from the political events (e.g. the German reunification), or changes in the monetary system such as the eurozone.

CPI_t is composed of the persistent ($\mathcal{T}_t$) and transitory (CPI_t^c) variation (e.g., Stock and Watson, 2007):

$$CPI_t = \mathcal{T}_t + CPI_t^c \tag{1.6}$$
$$\mathcal{T}_t = \mathcal{T}_{t-1} + \varepsilon_t^{\mathcal{T}}, \tag{1.7}$$

where $\varepsilon_t^{\mathcal{T}}$ is a shock uncorrelated with CPI_t^c. One can think of $\mathcal{T}_t$ in equation (1.6) as a time-varying inflation endpoint: $\lim_{s\to\infty} E_t(CPI_{t+s}) = \mathcal{T}_t$ (Kozicki and Tinsley, 2001a, 2006). Investors do not observe $\mathcal{T}_t$ and estimate its movements by means of constant gain learning. According to the constant gain rule, and unlike classical recursive least squares, recent observations are overweighed relative to those from the distant past. This feature makes the rule suitable for learning about time-varying parameters. From the definition of constant gain least squares applied to our setting, we form a proxy for $\mathcal{T}_t$ as a discounted moving average of the past realized core CPI:

$$\tau_t^{CPI} = \frac{\sum_{i=0}^{t-1} v^i CPI_{t-i}}{\sum_{i=0}^{t-1} v^i}, \tag{1.8}$$

where $(1-v)$ is the constant gain. The above equation can be rewritten as a learning recursion (e.g., Carceles-Poveda and Giannitsarou, 2007):

$$\tau_t^{CPI} = \tau_{t-1}^{CPI} + (1-v)\left(CPI_t - \tau_{t-1}^{CPI}\right). \tag{1.9}$$

Thus, at every time step, investors update their perceptions of τ_t^{CPI} by a small fixed portion of the deviation of current inflation from the previous long-run mean. Using inflation surveys, we estimate the gain parameter at $v = 0.9868$ (standard error 0.0025), and truncate the sums in equation (1.8) at $N = 120$ months. Appendix A.7 provides details of the estimation of v.[8] With those parameters, an observation from ten years ago still receives a weight of approximately 0.2.

[8] A number of papers argue for a similar gain parameter for inflation: Kozicki and Tinsley (2001a) use $v = 0.985$ for monthly data, Piazzesi and Schneider (2011) and Orphanides and Wei (2010) use $v = 0.95$ and $v = 0.98$ for quarterly data, respectively. Kozicki and Tinsley (2005) estimate $v = 0.96$ and find that discounting past data at about 4% per quarter gives inflation forecasts that closely track the long-run inflation expectations from the Survey of Professional Forecasters. The truncation parameter $N = 120$ months is motivated by the recent research of Malmendier and Nagel (2009) who argue that individuals form their inflation expectations using an adaptive rule and learn from the data experienced over their lifetimes rather than from all the available history. We stress that the parameters v and N are not a knife edge choice that would determine our subsequent findings. A sensitivity analysis shows that varying N between 100 and 150 months and v between 0.975 and 0.995 leads to negligible quantitative differences in results and does not change our interpretation. These results are available in Appendix A.7.D.

The application of the rule (1.9) to our context has a direct economic motivation. Evans, Honkapohja, and Williams (2010) show that the constant gain learning algorithm provides a maximally robust optimal prediction rule when investors are uncertain about the true data generating process, and want to employ an estimator that performs well across alternative models. This property makes the estimator (1.8) a justified choice in the presence of structural breaks and drifting coefficients. As an important feature, τ_t^{CPI} uses data only up to time t, hence it relies on the information available to investors in real-time.

We find that τ_t^{CPI} explains 86% of variation in yields on average across maturities from one to 20 years, with the lowest R^2 of 68% recorded for the one-year rate. Figure 1.2, panel a, superimposes the one- and ten-year yield with τ_t^{CPI} showing that the low-frequency variation in interest rates coincides with the smooth dynamics of our measure. For comparison, in panel b, we plot the median inflation forecast from the Livingston survey one year ahead, collected in June and December each year. The limited forecast horizon drives shorter-lived variation in the survey-based measure especially in the volatile periods; still, τ_t^{CPI} and surveys share a largely similar behavior over time.

[Figure 1.2 here.]

Our approach to constructing τ_t is deliberately simple, as we aim to obtain a measure of the low frequency factor in yields that is readily available to a bond investor. Still, it is informative to compare different specifications for τ_t and their implications for the subsequent predictability results. One alternative would be to use the moving average of past short rates. Intuitively, however, the moving average of past short rates faces a trade-off between smoothing over the business cycle frequency in the short rate and simultaneously extracting a timely measure of the generational frequency: In terms of equation (1.3) this tradeoff is represented by x_t and τ_t. For completeness, Appendix A.7.D analyzes the case in which the local mean reversion of yields is measured with the moving average of the past short rate. It also investigates the sensitivity of our findings to the way we construct the moving average. The results provide a robustness check for our predictability evidence and stress the importance of using an economically motivated variable—inflation—to explain the short rate behavior. Next, we show that τ_t^{CPI} has an interpretation in the context of its cointegrating relation with yields.

1.2.C Cycles as deviations from the long-run relationship between yields and short rate expectations

The high persistence of interest rates observed in historical samples suggests their close-to non-stationary dynamics. Indeed, many studies fail to reject the null hypothesis of a unit root in the US data (e.g. Jardet, Monfort, and Pegoraro, 2010; Joslin, Priebsch, and Singleton, 2010).[9] To the extent that our measure of τ_t explains a vast part of slow movements in yields, one can expect that yields and τ_t^{CPI} are cointegrated. Cointegration provides an econometric argument for our initial intuition that cycles should predict bond excess returns.

In our sample, yields and τ_t^{CPI} both feature nonstationary dynamics, as indicated by unit root tests. Following the standard approach (Engle and Granger, 1987), we regress yields on a contemporaneous value of τ_t^{CPI}:

$$y_t^{(n)} = b_0^{(n)} + b_\tau^{(n)} \tau_t^{CPI} + \epsilon_t^{(n)}, \tag{1.10}$$

and test for stationarity of the fitted residual. We denote the fitted residual of (1.10) as $c_t^{(n)}$ for individual yields, and $\overline{c}_t$ for the average yield across maturities, i.e. $\overline{y}_t = \frac{1}{20}\sum_{i=1}^{20} y_t^{(i)}$. To summarize their properties, we provide point estimates of (1.10) for $\overline{y}_t$ together with Newey-West corrected t-statistics (in brackets):

$$\overline{c}_t = \overline{y}_t - \underbrace{\widehat{b}_0}_{0.02\ [4.7]} - \underbrace{\widehat{b}_\tau}_{1.24\ [14.2]} \tau_t^{CPI}, \quad R^2 = 0.86. \tag{1.11}$$

We report detailed results of stationarity tests in Appendix A.1, and here just state the main conclusions. We consistently reject the null hypothesis that $c_t^{(n)}$ contains a unit root for maturities from one to 20 years at the 1% level. Thus, the data strongly supports the cointegrating relation.

Note that $c_t^{(n)}$ gives an empirical content to the notion of cycles we have introduced in equation (1.5). By cointegration, cycles represent stationary deviations from the long-run relationship between yields and the slow moving component of inflation expectations. Therefore, invoking the Granger representation theorem, they should forecast either Δy_t or $\Delta \tau_t$, or both. To verify this prediction, we estimate the error correction representation for yield changes. We allow one lag of variable changes to account for short-run deviations from (1.10):

[9] Even if the assumption of nonstationary interest rates may raise objections, the results of Campbell and Perron (1991) suggest that a near-integrated stationary variables are, in a finite sample, better modeled as containing a unit root, despite having an asymptotically stationary distribution.

$$\Delta y_t^{(n)} = a_c c_{t-\Delta t}^{(n)} + a_y \Delta y_{t-\Delta t}^{(n)} + a_\tau \Delta \tau_{t-\Delta t} + a_0 + \varepsilon_t, \quad \Delta t = 1 \text{ month}. \tag{1.12}$$

We focus on $\Delta y_t^{(n)}$ because we are interested in transitory adjustments of asset prices. Indeed, the error correction term, $c_{t-\Delta t}^{(n)}$, turns out significant precisely for this part of the system.

Table 1.2 presents the estimates of equation (1.12) for monthly data. The essence of the results is that cycles are highly significant predictors of monthly yield changes. The negative sign of a_c coefficients for all maturities suggests that a higher value of the cycle today predicts lower yields and thus higher excess bond returns in the future. As such, it conforms with the intuition of equation (1.5) that cycles and term premia should be positively related.

[Table 1.2 here.]

We build on this observation to explore the predictability of excess bond returns by the cycles. Beside formal motivation, cointegration provides a useful property that facilitates our subsequent analysis: the OLS estimates of equation (1.10) are "superconsistent" and converge to the true values at the rapid rate T^{-1} (Stock, 1987). Therefore, using cycles as predictors, we circumvent the problem of generated regressors.

1.3 The predictability of bond excess returns revisited

In this section, we discuss the predictability of bond excess returns and construct the return forecasting factor. We show that the predictable variation in bond returns is larger than reported so far, and quantify the amount of transitory movements in yields due to varying short rate expectations and risk premia, respectively.

1.3.A First look at predictive regressions

We regress bond excess returns on the cycles, and discuss the results in the context of the common predictive regressions using forward rates (Cochrane and Piazzesi, 2005; Fama and Bliss, 1987; Stambaugh, 1988). Following much of the contemporaneous literature, we focus on one-year holding period bond excess returns, and defer the analysis of other holding periods to Appendix A.6.

To fix notation, a one-year holding period excess log return on a bond with n years to maturity is defined as: $rx_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)}$, where $p_t^{(n)}$ is the log price of a zero bond, $p_t^{(n)} = -ny_t^{(n)}$, and $y_t^{(1)}$ is the one-year continuously compounded rate. The one-year forward rate locked in for

the time between $t+n-1$ and $t+n$ is given by: $f_t^{(n)} = p_t^{(n-1)} - p_t^{(n)}$. In Table 1.3, we report the descriptive statistics for bond excess returns.

[Table 1.3 here.]

We obtain cycles as fitted residuals from the regressions of yields on the persistent inflation factor in equation (1.10), i.e.:

$$c_t^{(n)} = y_t^{(n)} - \widehat{b}_0^{(n)} - \widehat{b}_\tau^{(n)} \tau_t^{CPI}, \tag{1.13}$$

and estimate the predictive regression:

$$rx_{t+1}^{(n)} = \delta_0 + \sum_i \delta_i c_t^{(i)} + \varepsilon_{t+1}^{(n)}, \tag{1.14}$$

where $i = \{1, 2, 5, 7, 10, 20\}$ years. This choice of maturities summarizes the relevant information in c_t's. To provide a benchmark for our results, we also estimate an analogous equation using forward rates instead of cycles:

$$rx_{t+1}^{(n)} = d_0 + \sum_i d_i f_t^{(i)} + \varepsilon_{t+1}^{(n)}. \tag{1.15}$$

For excess returns, we single out interesting points along the yield curve with maturities of two, five, seven, ten, 15 and 20 years. Sparing the detailed results, we note that in terms of its predictive power, regression (1.14) is equivalent to using a set of yields and τ_t^{CPI} as the explanatory variables. We follow the representation in terms of cycles because it offers a convenient interpretation of factors underlying the yield curve which we exploit below.

Table 1.4 summarizes the estimation results. We report the adjusted R^2 values and the Wald test statistics for the null hypothesis that all coefficients in (1.14) are jointly zero. The individual coefficient loadings are not reported, as by themselves they do not yield an interesting economic interpretation (Section 1.3.D explains why). It is evident that c_t's forecast a remarkable portion of variation in excess bond returns. In our sample, R^2's increase from 42% up to 57% across maturities. On average, these numbers more than double the predictability achieved with forward rates.

[Table 1.4 here.]

The Wald test strongly rejects that all coefficient on c_t's are zero, using both the Hansen-Hodrick and the Newey-West method. However, since both tests are known to overreject the null

hypothesis in small samples (e.g., Ang and Bekaert, 2007), we additionally provide a conservative test based on the reverse regression delta method recently proposed by Wei and Wright (2010). This approach amounts to regressing short-horizon (one-month) returns on the long-run (twelve-month) mean of the cycles, and is less prone to size distortions.[10] Although the reverse regression test statistics are by design more moderate, we consistently reject the null of no predictability by the cycles at the conventional significance levels. We compare the standard errors obtained with the cycles to those of the forward rate regressions. In both samples and across all maturities, cycles give much stronger evidence of predictability than do forward rates. Increasing the number of forward rates or choosing different maturities does not materially change the conclusions.

One may be worried about the small-sample reliability of our findings. For this reason, Table 1.4 provides small sample (SS) confidence bounds on R^2's computed with the block bootstrap. Even though $c_t^{(n)}$ is estimated with a high precision, the bootstrap procedure automatically accounts for its uncertainty (see Appendix A.4 for details). Importantly, the lower 5% confidence bound for the R^2's obtained with the cycles consistently exceeds the large-sample R^2 obtained with forward rates. A similar discrepancy holds true for the reported values of the Wald test.

In the remainder of this section, we look into the anatomy of the cycles to better understand the sources of this predictability. We connect our findings with two well-documented results in the literature: (*i*) that a single linear combination of forward rates predicts excess bonds returns (the Cochrane-Piazzesi factor), and (*ii*) that this predictability cannot be attained by the three principal components of yields.

1.3.B Anatomy of the cycle

At different maturities n, $c_t^{(n)}$ give rise to the term structure of interest rate cycles. We use its cross-sectional dynamics to further decompose the yield curve. Building on the intuition of equation (1.5), the cycle with the shortest maturity, $c_t^{(1)}$, mirrors a transitory business cycle movement in short rate expectations, but not in term premia: For an investor with a one-year horizon, $y_t^{(1)}$ is risk-free in nominal terms. Therefore, a natural way to decompose the transitory variation in the yield curve into the expectations part and the premium part is by estimating:

$$rx_{t+1}^{(n)} = \alpha_0^{(n)} + \alpha_1^{(n)} c_t^{(1)} + \alpha_2^{(n)} c_t^{(n)} + \varepsilon_{t+1}^{(n)}, \quad n \geq 2. \tag{1.16}$$

[10] Wei and Wright (2010) extend the reverse regressions proposed by Hodrick (1992) beyond just testing the null hypothesis of no predictability. In constructing one-month excess returns on bonds, we follow Campbell and Shiller (1991), approximating the log price of a $(n-1/12)$-maturity bond as $-(n-1/12)y_t^{(n)}$.

We use this regression to gauge the extent of variation in $c_t^{(n)}$ due to the expectations ($R_{ex}^{2,(n)}$) and premia ($R_p^{2,(n)}$), respectively, as:

$$R_{ex}^{2,(n)} := \left(\frac{\alpha_1^{(n)}}{\alpha_2^{(n)}}\right)^2 \frac{Var\left(c_t^{(1)}\right)}{Var\left(c_t^{(n)}\right)} \quad \text{and} \quad R_p^{2,(n)} := 1 - R_{ex}^{2,(n)}. \tag{1.17}$$

Figure 1.3 looks into this decomposition more closely. In panel a, we start by showing how much of the variation in individual excess returns can be explained by the individual cycles, i.e. we run a univariate regression of excess returns on cycles one-by-one: $rx_{t+1}^{(n)} = a_{i,n} + b_{i,n}c_t^{(i)} + \varepsilon_{t+1}^{(i,n)}$. The monotonic pattern of the plot verifies the intuition that the premium component of $c_t^{(n)}$ increases with the maturity, but it is zero for $c_t^{(1)}$.

[Figure 1.3 here.]

Panel b of Figure 1.3 shows the gain in our ability to explain returns when estimating equation (1.16) over the univariate regressions in panel a. The source of this gain is intuitive. In equation (1.16), we allow the OLS to prune the transitory short rate expectations component from $c_t^{(n)}$. Accordingly, we find that the estimated $\alpha_1^{(n)}$ coefficients are consistently negative across maturities, while $\alpha_2^{(n)}$ coefficient are positive and larger in absolute value than the corresponding $\alpha_1^{(n)}$ estimates (the individual coefficients are not reported). Separating the premium part of the cycle in that way leads to a significant increase in the R^2's, especially at the shorter maturities. The predictability obtained with (1.16) is only slightly weaker than the one reported in Table 1.4, in which six cycles are used. The deterioration is most pronounced at shorter maturities.

Panel c of Figure 1.3 applies the decomposition (1.17) to quantify the premium and expectations shares in the cycles, $c_t^{(n)}$. The premium-to-expectations split varies from 11%-to-89% for the two-year bond, through 52%-to-48% for the ten-year bond, up to 70%-to-30% for the 20-year bond. These numbers correspond to an average cycle variation due to term premium of 15, 43 and 60 basis points at the respective maturities.[11]

The economic interpretation of the cycles as sum of transitory short rate expectations and term premia is coupled with an interesting pattern of mean reversion across maturities: The persistence of the cycles declines gradually from above 13 months half-life for the $c_t^{(1)}$ to 10.5 months for the 5-year cycle $c_t^{(5)}$, at which level it approximately stabilizes for longer maturities.

[11] The numbers are obtained as: $R_p^{2,(n)} \times \text{std}(c_t^{(n)})$, where $\text{std}(c_t^{(n)})$ is the sample standard deviation of the n-maturity cycle.

1.3.C The single returns forecasting factor: distilling the term premium frequency

Cochrane and Piazzesi (2005) show that a single factor, which they make observable through a linear combination of forward rates, captures almost complete variation in expected excess returns on bonds with different maturities. In the next two sections we relate our findings to their result.

The predictive regressions above suggest that we can construct the single forecasting factor in two steps, which we summarize as follows:

Step 1. Obtain the cycles $c_t^{(n)}$ as residuals from regressing yields across maturities on τ_t^{CPI} as in equation (1.10) and (1.13), i.e. remove the generational frequency from yields.

Step 2. Project the average cycle onto the transitory short rate expectations factor $c_t^{(1)}$, thus remove the business cycle frequency due to the short rate expectations. The residual presents the return forecasting factor, which we call $\widehat{cf}_t$:

$$\bar{c}_t = \gamma_1 c_t^{(1)} + \bar{\varepsilon}_t, \quad \text{where} \quad \bar{c}_t = \frac{1}{m-1}\sum_{i=2}^{m} c_t^{(i)} \tag{1.18}$$

$$\widehat{cf}_t = \bar{c}_t - \hat{\gamma}_1 c_t^{(1)} \tag{1.19}$$

Empirically, $\widehat{cf}_t$ has a faster mean reversion than the factors related to the short rate expectations: its half-life is slightly below 10 months, giving rise to a term premium frequency. Figure 1.4 displays the evolution of the forecasting factor over time.

Note that both steps 1 and 2 involve only contemporaneous time-t variables on the left- and right-hand sides of the regressions. No information about future excess returns is used.

[Figure 1.4 here.]

In Table 1.5, panel A, we report the estimates of equation (1.18). The positive sign of γ_1 is consistent with the decomposition of cycles into the premium and expectations components in equation (1.16). A 100 basis points change in the transitory short rate expectations factor generates a 42 basis points reaction in the cycles, on average. Moreover, low standard errors on the estimated coefficients indicate that we are able to identify a robust feature of the data.

[Table 1.5 here.]

With $\widehat{cf}_t$, we forecast individual excess bond returns:

$$rx_{t+1}^{(i)} = \beta_0^{(i)} + \beta_1^{(i)}\widehat{cf}_t + \varepsilon_{t+1}^{(i)}. \tag{1.20}$$

Panel B of Table 1.5 reports the predictability of individual bond returns achieved with the single factor. On average, during the 1971–2009 period, $\widehat{cf}_t$ explains around 54% of variation in excess returns. The results are no significantly worse than those of the unrestricted regression in equation (1.14): That comparison is reflected in the row "ΔR^2."

Appendix A.5 provides several robustness checks, and discusses alternative ways of constructing the single factor: (i) by exploiting information about future excess returns in analogy to the construction of the forward rate factor, (ii) in one step via non-linear least squares, and (iii) by means of the eigenvalue decomposition of the covariance matrix of expected returns. We show that different approaches to constructing $\widehat{cf}_t$ produce essentially an identical outcome.

1.3.D The Cochrane-Piazzesi factor

It is useful to connect our findings with the single linear combination of forward rates—the Cochrane-Piazzesi (CP) factor, which has proved itself as the most successful in-sample predictor of bond returns. To this end, let us run the usual predictive regression of an average (across maturities) holding period excess return $\overline{rx}_{t+1}$ on a set of m forward rates with maturities 1 to m years at time t:

$$\begin{aligned} \overline{rx}_{t+1} &= \gamma_0 + \sum_{i=1}^{m} \gamma_i f_t^{(i)} + \overline{\varepsilon}_{t+1} && (1.21) \\ &= \gamma_0 + \gamma' \mathbf{f}_t + \overline{\varepsilon}_{t+1}. && (1.22) \end{aligned}$$

$\gamma'\mathbf{f}_t$ constructs the return forecasting factor of Cochrane and Piazzesi (2005). From decomposition (1.10) and the definition of the forward rate, it follows:[12]

$$\begin{aligned} \overline{rx}_{t+1} &= \widetilde{\gamma}_0 + \tau_t^{CPI}(\sum_{i=1}^{m} \overline{\gamma}_i) + \sum_{i=1}^{m} \widetilde{\gamma}_i c_t^{(i)} + \overline{\varepsilon}_{t+1}, && (1.23) \\ &= \widetilde{\gamma}_0 + \overline{\gamma}'\mathbf{1}\tau_t^{CPI} + \widetilde{\gamma}'\mathbf{c}_t + \overline{\varepsilon}_{t+1}, && (1.24) \end{aligned}$$

[12] Assuming $y_t^{(n)} = b_0^{(n)} + b_\tau^{(n)}\tau_t^{CPI} + c_t^{(n)}$ the forward rate can be expressed as:

$$f_t^{(n)} = \left[-(n-1)b_\tau^{(n-1)} + nb_\tau^{(n)}\right]\tau_t^{CPI} - (n-1)c_t^{(n-1)} + nc_t^{(n)} - (n-1)b_0^{(n-1)} + nb_0^{(n)}.$$

where

$$\overline{\gamma}_k = \gamma_k \left[-(k-1) b_\tau^{(k-1)} + k b_\tau^{(k)} \right] \tag{1.25}$$

$$\widetilde{\gamma}_k = \begin{cases} k(\gamma_k - \gamma_{k+1}) & \text{for } 1 \leq k < m \\ k\gamma_k & \text{for } k = m, \end{cases} \tag{1.26}$$

and $\mathbf{1}$ is an m-dimensional vector of ones, $\gamma, \overline{\gamma}, \widetilde{\gamma}$ are respective $m \times 1$ vectors of loadings, and $\mathbf{c}_t = \left(c_t^{(1)}, \ldots, c_t^{(m)} \right)'$. We can apply the same logic to forecasting an excess return of any maturity.

By reexpressing equation (1.22), we gain an understanding of how forward rate regressions work. As a typical pattern in regression (1.22), the γ_i coefficients have a neutralizing effect on each other: Independent of the data set used or the particular shape of the loadings, γ_i's (an so $\overline{\gamma}_i$'s) roughly sum to a number close to zero. This is intuitive since only the cyclical part of yield variation matters for forecasting $\overline{rx}$. Equation (1.23) tells us that the OLS tries to remove the common τ_t^{CPI} from forward rates, while preserving a linear combination of the cycles. Thus, forecasting returns with forward rates embeds an implicit restriction on the slope coefficients: γ_i's are constrained by the dual role of removing the persistent component and minimizing the prediction error of excess returns using the cycles.

This interpretation can be tested by allowing the excess returns in (1.23) to load with separate coefficients on $\overline{\gamma}'\mathbf{1}\tau_t$ and $\widetilde{\gamma}'\mathbf{c}_t$. Effectively, we can split the forward factor into two components, and estimate: $\overline{rx}_{t+1} = a_0 + a_1(\overline{\gamma}'\mathbf{1}\tau_t) + a_2(\widetilde{\gamma}'\mathbf{c}_t) + \varepsilon_{t+1}$. Table 1.7 summarizes the estimates. This exercise gives an $\bar{R}^2$ of 30%, similar to 26% obtained with $\gamma'\mathbf{f}_t$. As expected, the predictability comes from the strongly significant $\widetilde{\gamma}'\mathbf{c}_t$ term (Newey-West t-statistic of 5.9). The persistent component $\gamma'\mathbf{1}\tau_t$ is not significantly different from zero. Figure 1.7 superimposes $\gamma'\mathbf{f}_t$ with its cyclical part $\widetilde{\gamma}'\mathbf{c}_t$ (both standardized).

[Figure 1.7 and Table 1.7 here.]

The plot confirms that $\overline{\gamma}'\mathbf{1}\tau_t$ has an almost imperceptible contribution to the total dynamics of the CP factor, $\gamma'\mathbf{f}_t$. The last column in Table 1.7 reports the R^2 values achieved with the single factor $\widehat{cf}_t$, which we can treat as an optimally chosen linear combination of the cycles.[13] This number helps assess the predictability earned by freeing up the coefficients in $\widetilde{\gamma}'\mathbf{c}_t$.

[13] $\widehat{cf}_t$ is constructed as in equation (1.18), but based on yields with maturities corresponding to the forward rates from one to ten years used in Table 1.7.

These results suggest an interpretation of the Cochrane-Piazzesi factor as a constrained linear combination of the cycles. By the presence of the persistent component in forward rates, the factor is restricted in its ability to extract information about premia. Using just forward rates, and with no information about τ_t^{CPI}, this is the best predictability one can achieve.

1.4 Roles of factors in the cross section of yields

We can summarize the yield curve with three factors: (i) the persistent short rate expectations component related with inflation τ_t^{CPI}, (ii) the transitory short rate expectations factor $c_t^{(1)}$, and (iii) the term premium factor $\widehat{cf}_t$:

$$X_t = \left(\tau_t^{CPI}, c_t^{(1)}, \widehat{cf}_t\right)'. \tag{1.27}$$

These factors can always be expressed in terms of a linear combination of yields and τ_t^{CPI}. However, since our goal is to quantify their cross-sectional roles, we rely on the "preprocessed" variables.

1.4.A Quantifying the cross-sectional impact of factors on yields

Level, slope and curvature are known to explain over 99.9% of variation in yields. To obtain a comparable figure for X_t, we regress yields with maturity of one year through 20 years on X_t:

$$y_t^{(n)} = a_n^{\text{LS}} + b_n^{\text{LS}\prime} X_t + \varepsilon_t^{(n)}. \tag{1.28}$$

The fit of the above regression is the best a linear factor model in X_t can reach, therefore we do not impose no-arbitrage restrictions. Relative to three PCs, X_t achieves a slightly lower R^2 of 99.68% on average across maturities. The deterioration is not surprising. The three variables in X_t contain cross-sectional information that is equivalent to lvl_t and $c_t^{(1)}$ (see Section 1.4.B). As such, they cannot do better than the first two PCs in terms of minimizing pricing errors. We could easily improve on this front by including higher order PCs in the state vector. However, the imperfect pricing performance of our setting serves the goal of focussing on economically large effects in the cross-section of yields. Thus, we maintain a low-dimensional form of X_t.

The estimates of (1.28) allow us to assess the role of $\widehat{cf}_t$ in the cross-section. Important results are summarized in Figure 1.8. Panel a plots how yields with different maturities react to a one

standard deviation shock to the elements of X_t . Specifically, each line traces out the regression coefficients b_n^{LS} multiplied by the standard deviation of the corresponding factor.

[Figure 1.8 here.]

The shapes of the loadings are intuitive. The slow moving long-run inflation expectations determine the overall level of interest rates. Indeed, τ_t^{CPI} has the most pronounced effect on the yield curve in terms of magnitude, and propagates almost uniformly throughout maturities. As such, it resembles the usual PCA level factor. The loadings on $c_t^{(1)}$ are downward sloping. Their pattern aligns with the interpretation of $c_t^{(1)}$ as the transitory short rate expectations component whose contribution diminishes as the maturity of the bond increases. Loadings of the premium factor $\widehat{cf}_t$ feature an opposite shape to $c_t^{(1)}$, and rise the with maturity. The two variables $c_t^{(1)}$ and $\widehat{cf}_t$ have approximately equal impact on the yield curve at maturity of ten years. Below that threshold, the impact of transitory short rate expectations dominates that of the premia; above that threshold, the impact of premia dominates that of the transitory short rate expectations.

Panels b through d of Figure 1.8 display the reaction of the yield curve when a factor shifts from its mean to its 10th or 90th percentile value in our sample, ceteris paribus. While movements in τ_t^{CPI} have the largest effect on the cross-section of yields, the impact of the remaining two states is also non-trivial. A hypothetical change in $c_t^{(1)}$ from its 10th to 90th percentile value induces a 360 basis points rise in the two-year yield and a 160 basis point rise in the ten-year yield. An analogous effect of a change in $\widehat{cf}_t$ is 60 and 151 basis points at the two- and ten-year maturity, respectively.

It is informative to analyze the cross-sectional role of our expectation and term premium states relative to that of the level, slope and curvature. Such a comparison is provided in Figure 1.9 which plots the influence of one standard deviation change in each of the variables on the yield curve as a function of maturity. The figure also reports the average absolute impact of each of those shocks in basis points. Panel a compares the effect of the level lvl_t against the persistent inflation expectations, τ_t^{CPI}; panel b plots the effect of the slope slo_t and the transitory rate expectations, $c_t^{(1)}$; panel c juxtaposes the curvature cur_t and the premium factor, $\widehat{cf}_t$. The loadings are obtained by running the OLS regression of a yield on each set of three factors. The results corroborate the statement that the level effect on yields is almost completely determined by the persistent component. On average, one standard deviation change in the PCA level (the persistent component τ_t^{CPI}) moves yields by 250 (232) basis points. An equally interesting pattern pertains to both $c_t^{(1)}$ and $\widehat{cf}_t$. A change in the transitory short rate expectations $c_t^{(1)}$ gives an

average yield response of 76 basis points, which more than doubles the average absolute impact of the slope (29 basis points). Most notably, the role of the return forecasting factor in determining the variation of yields exceeds not only that of the curvature but also the one of the slope. The average absolute impact of $\widehat{cf}_t$ of 54 basis points is higher than 29 basis points induced by the slope and 8 basis points induced by the curvature.

[Figure 1.9 here.]

These results suggest the term premia have a visible influence on the shape of the yield curve. By not including additional factors in X_t, we have deliberately kept the measurement error relatively large. The affine model (1.28) gives an RMSE of 12 basis points on average across maturities used in estimation.[14] This number is large enough to hide higher-order principal components, but clearly not large enough to hide $\widehat{cf}_t$.

1.4.B Link to the level, slope and curvature

Empirical evidence shows that the predictability of bond returns by the level factor is close to zero. Instead, the predictability by the slope gives R^2's of about 15% for long maturities. Moreover, higher-order PCs seem also important for the term premia, even though their effect on the cross-section of yields does not exceed a few basis points.

The principal components rotate short rate expectations and risk premia conveyed by yields into several orthogonal factors by optimizing a statistical criterion. Thus, they can make it hard to separate economically different effects. To demonstrate this point, we explore the link between the PCs and the decomposition we have proposed. Figure 1.5 plots the contribution of τ_t^{CPI}, $c_t^{(1)}$ and $\widehat{cf}_t$ to the explained variance of the PCs. Panel *a* uses yields with maturities from one to ten years, panel *b* extends the maturities up to 20 years.

[Figure 1.5 here.]

The figure shows that the level factor is predominantly related to short rate expectations (τ_t^{CPI} and $c_t^{(1)}$), while the premium component $\widehat{cf}_t$ accounts for a small portion between 3% and 5% of its overall variance. Similarly, the slope of the yield curve combines information about (transitory) short rate expectations and risk premia, with $c_t^{(1)}$ and $\widehat{cf}_t$ explaining roughly two-thirds and one-third of its variance, respectively. This relatively large contribution of $\widehat{cf}_t$ tells why the slope

[14] For comparison, a typical RMSE obtained with three latent factors is about half that number.

carries some degree of predictability for future returns, but this predictability is dampened by an even larger contribution of the transitory short rate expectations $c_t^{(1)}$—reminiscent of an error-in-variables problem. Figure 1.5 also reveals that beyond the level and the slope, our three factors capture only a small part of movements in higher order PCs, $PC3$–$PC5$. A comparison of panels a and b of the figure suggests that the role of $PC3$–$PC5$ in the yield curve varies substantially with yield maturities included to construct the PCs as it does across different data sets (not reported).

To see the connection between the level and the return forecasting factor, we note that:

$$lvl_t = q\mathbf{1}'\mathbf{y}_t, \tag{1.29}$$

where $\mathbf{y}_t = \left(y_t^{(1)}, \ldots, y_t^{(m)}\right)'$, $\mathbf{1}$ is a $m \times 1$ vector of ones, q is a constant and $q\mathbf{1}$ is the eigenvector corresponding with the largest eigenvalue in the singular-value decomposition of the yield covariance matrix. Clearly, lvl_t is proportional to the sum of the persistent component and the average cycle. Therefore, we can project lvl_t onto τ_t^{CPI}, and obtain the average cycle as the cointegrating residual. We denote this residual by c_t^{lvl}:

$$lvl_t = b_0^{lvl} + b_\tau^{lvl}\tau_t^{CPI} + c_t^{lvl}. \tag{1.30}$$

τ_t^{CPI} explains 86% of variation in the level factor, which is consistent with the R^2 of regression (1.11). This exercise leads to several remarks, which we summarize in Table 1.6. Panel A of the table shows the unconditional correlations between c_t^{lvl}, the average cycle across maturities $\overline{c}_t$, and the usual principal components. First, and not surprisingly, c_t^{lvl} and $\overline{c}_t$ capture essentially the same source of variation in the yield curve, and their correlation exceeds 99%. Likewise, the last column of panel A in Table 1.6 shows that the correlation between the two corresponding forecasting factors, $\text{corr}(\widehat{cf}_t^{lvl}, \widehat{cf}_t)$, is 99.9% so the return predictability remains unaffected.

Second, the cyclical element of the level shows a non-negligible correlation with the remaining principal components of yields. For instance, its unconditional correlation with the slope can easily exceed 30%. This suggests that the orthogonalization of the level towards higher-order principal components is achieved only with respect to the most persistent component.

[Table 1.6 here.]

How important are the higher-order PCs become for return predictability? We regress excess returns on the single factor $\widehat{cf}_t$ and the original $PC1$ through $PC5$. The results are stated in

panel B of Table 1.6. The key observation is that in the presence of $\widehat{cf}_t$, the PCs lose most of their economic and statistical significance for maturities from two to ten years. Instead, the single forecasting factor has consistently large coefficients and t-statistics.

Figure 1.6 synthesizes the results by comparing the R^2's obtained with the unconstrained regressions (Section 1.3.A), with the single factor (Section 1.3.C), and those obtained with the single factor and $PC1$ through $PC5$. The plot suggests that, to the first order, $\widehat{cf}_t$ captures the important variation in term premia.[15]

[Figure 1.6 here.]

Given the basic representation of yields in equation (1.1), the level factor should reflect premia unless they are precisely offset by the short rate expectations. Our findings point out that such a cancelation effect is unlikely to take place.

1.5 Macroeconomic fundamentals and $\widehat{cf}$

This section studies the link between the return forecasting factor and macroeconomic fundamentals. We find that $\widehat{cf}_t$ comprises the predictability of many macro-finance variables. Conditional on that factor, the additional predictive power of macroeconomic risk is attached to bonds with short maturities, which we associate with the influence of monetary policy on this segment of the curve.

1.5.A Do macro variables predict returns beyond $\widehat{cf}$?

Including macro-finance variables in predictive regressions together with the CP factor or with yield principal components usually leads to an increase in R^2. Ludvigson and Ng (2009) summarize information in 132 macro-finance series and find that real activity and inflation factors remain highly significant and increase the forecasting power relative to the CP factor. Cooper and Priestley (2009) reach a similar conclusion considering the output gap.

It is natural to ask whether and how these conclusions may change when we take $\widehat{cf}_t$ as our benchmark for predictability. Specifically, we estimate the regression:

[15] As a caveat, the importance of higher order PCs versus the role of the single factor may differ across subsamples, both for economic and statistical reasons. For instance, in unreported results, we find that during the Greenspan's term in office the predictability of bond returns at short maturities (especially two years) is weaker than the predictability of bond returns at long maturities, suggesting that more than one factor may be needed to explain the entire term structure of bond returns. We provide additional discussion of this point in Section 1.5.B.

$$rx_{t+1}^{(n)} = b_0 + b_1 \widehat{cf}_t + b_2' \text{Macro}_t + \varepsilon_{t+1}^{(n)}, \tag{1.31}$$

where Macro_t represents the additional macro-finance information. This regression allows us to assess which macroeconomic variables are reflected in the movements of bond risk premia.

Panel A of Table 1.8 displays estimates of (1.31) with eight macro-finance factors, $\widehat{F}_t$, constructed according to Ludvigson and Ng (2009), and indicates the domains that these factors capture. We use data from 1971:11 through 2007:12. The end of the sample is dictated by the availability of the macro series. Alone, $\widehat{F}_t$ explain more than 20% of variation in bond excess returns. Although we do not report the details of the separate regression of rx on $\widehat{F}_t$, in Table 1.8 we indicate significant factors at the 1%, 5% and 10% level with superscripts H, M, L, respectively. These factors involve financial spreads, stock market returns, inflation, and monetary conditions.

[Table 1.8 here.]

In the presence of $\widehat{cf}_t$, however, most macro variables lose predictive power. Their contribution to R^2, denoted as "ΔR^2" in the table, does not exceed 2%. The only exception is the two-year bond for which inflation and, to a lesser degree, the real activity factor remain significant yielding ΔR^2 of 5%.

We do not report analogous estimates with the CP factor for our sample, and just note that they conform with the conclusions of Ludvigson and Ng (2009). Using the CP factor as a benchmark, changes the role of macroeconomic information in (1.31) in that most $\widehat{F}_t$ variables preserve their significance.

Panel B of Table 1.8 uses output gap to represent macro information in equation (1.31). Following Cooper and Priestley (2009), we obtain gap_t from the unrevised data on industrial production by applying a quadratic time trend.[16] Also here, the estimates suggest that gap_t does not provide additional information beyond that conveyed by $\widehat{cf}_t$.

Out of eight factors considered in panel A, only $\widehat{F}_{2t}$ is statistically significant for intermediate and long maturities. To the extent that $\widehat{F}_{2t}$ is related to different financial spreads, as shown by Ludvigson and Ng (2009), it seems to reflect the variation in funding liquidity. To explore this predictability channel, we construct several liquidity proxies such as spreads on commercial papers, swap rates, Baa corporate bonds, three-month T-bill over Fed's target, and the TED. We also consider the on-the-run liquidity factor recently proposed by Fontaine and Garcia (2010)

[16] We construct gap_t using the industrial production going back to 1948:01 as in Cooper and Priestley (2009).

(henceforth, FG factor).[17] Exact descriptions of the variables are in Appendix A.2. We evaluate the joint predictive role of $\widehat{cf}_t$ and each of those variables within the following regression:

$$rx_{t+1}^{(n)} = b_0 + b_1 \widehat{cf}_t + b_2 \text{liq}_t + \varepsilon_{t+1}^{(n)}, \tag{1.32}$$

where liq_t denotes the respective liquidity measure. Due to data availability, the sample is 1987:04 through 2007:12. Panel C in Table 1.8 presents the results. The FG factor and the Moodys Baa spread turn out to be the only variables that, albeit weakly, continue to contribute to the predictability achieved with $\widehat{cf}_t$.

At this juncture, it is worth recalling two properties of $\widehat{cf}_t$ revealed by our analysis up to now: (*i*) its predictive power increases with bond maturity, and (*ii*) the factor has a non-trivial effect on the cross-section of yields. In combination with the conclusions of the current section, macroeconomic risk in term premia appear to have have interesting properties across bond maturities. Their contribution seems particularly prominent at the short maturity. Specifically, using macroeconomic information beside $\widehat{cf}_t$ could improve investors' forecast of the return on the two-year bond, but not on bonds with longer maturities. Next section looks into this matter in more detail.

1.5.B What is special about the return of a two-year bond?

Two characteristics of the two-year bond return make it worthy of further scrutiny. While over the 1971–2009 period $\widehat{cf}_t$ explains 53% of variation in the ten-year bond return, its predictive power for the two-year bond is visibly lower at 38%. Interestingly, the opposite holds true for macro fundamentals, which compensate the deterioration in the forecasting power of $\widehat{cf}_t$ precisely at the short maturity range. We link the latter finding with monetary policy, and the role it plays at the short end of the curve. To this end, we re-examine the regression (1.31) for the two-year bond considering two subsamples: (*i*) the inflationary period 1971:11–1987:12, and (*ii*) the post-inflation period, 1988:01–2007:12. Depending on the sample, we find different results. In the first period, the inflation factor, $\widehat{F}_{4t}$, is the only one that adds extra predictive power. Quite differently, in the post-inflation period it is the real factor, $\widehat{F}_{1t}$, that remains significant. This pattern roughly coincides with the two domains—nominal versus real—that have been driving monetary policy actions in the respective samples.

It is convenient to rewrite the excess return on a two-year bond as:

[17] We thank Jean-Sébastien Fontaine for providing the data on their liquidity factor.

$$rx_{t+1}^{(2)} = f_t^{(2)} - y_{t+1}^{(1)}. \tag{1.33}$$

$f_t^{(2)}$ represents investors' risk-neutral expectation about the evolution of the one-year yield into next year, and $y_{t+1}^{(1)}$ is its true realization. We can always write the excess return as the sum of expected and unexpected return, $rx_{t+1}^{(2)} = E_t(rx_{t+1}^{(2)}) + U_{t+1}$. From equation (1.33), the unexpected return U_{t+1} is (inversely) related to the forecast error investors make about the path of $y_t^{(1)}$, i.e. $U_{t+1} = E_t(y_{t+1}^{(1)}) - y_{t+1}^{(1)}$.

We ask whether macroeconomic fundamentals help predict U_{t+1}, thus contributing to the predictability of realized excess returns. Have investors incorporated all relevant macroeconomic information into their predictions of $y_{t+1}^{(1)}$? Yield curve surveys come in handy in answering this question. Limited by the data availability, we focus on the post-inflation period, for which we obtain median prediction of $y_{t+1}^{(1)}$ one year ahead, $E_t^s(y_{t+1}^{(1)})$, from Blue Chip Financial Forecasts (BCFF). Let us consider the regression:

$$y_{t+1}^{(1)} - E_t^s(y_{t+1}^{(1)}) = b_0 + b_1 UNEMPL_t + b_2 \widehat{cf}_t + \varepsilon_{t+1}, \tag{1.34}$$

where $y_{t+1}^{(1)} - E_t^s(y_{t+1}^{(1)}) = -U_{t+1}^s$ is the forecast error implied by the survey expectations, and $UNEMPL_t$ denotes the unemployment rate. Given the dual mandate of the Fed to target the full employment and price stability, $UNEMPL$ is well-suited to represent a major macro risk in the post-inflation period. We also include $\widehat{cf}_t$ to account for the fact that surveys may be an imperfect proxy for the expectation of $y_{t+1}^{(1)}$.

If investors used all available information to forecast yields, the coefficient on unemployment in regression (1.34) should be insignificant. Panel B in Table 1.9 suggests the contrary. Not only is the $UNEMPL$ highly significant (t-statistic of -5.9), but also it accounts for most of the explained 33% of variation in $y_{t+1}^{(1)} - E_t^s(y_{t+1}^{(1)})$. A more detailed inspection of the forecast error (not plotted) shows that investors have largely failed to predict the turning points between monetary policy easing and tightening regimes. These turning points roughly coincide with two peaks of unemployment in our sample, thus explaining its predictive content in regression (1.34).[18] As such, unemployment appears as a predictor of realized bond returns.

[Table 1.9 here.]

[18] In the last 20 years, the rule of thumb has been that the Fed would not start tightening unless the unemployment has peaked and reliably gone down. This belief has been presented both by practitioners and the Fed officials. The evidence we provide does not necessarily imply that investors have been processing macro information inefficiently. It is well-known that it is difficult to forecast the exact timing of peaks in any cyclical macro series in real time.

With this narrative evidence, we point to unexpected returns as a possible channel through which fundamentals enter the predictive regression for realized bond returns. Clearly, with an increasing maturity of the bond, and as the direct impact of monetary policy on yields tapers off, we expect this channel to loose its appeal. This intuition seems to be supported by our results (see Table 1.8).

1.6 Robustness

In this section, we analyze the robustness of our results. In the first step, we test the predictive performance of the cycles out of sample. Then, we show that the predictability results are not driven by the data sets we use or the way we construct the zero curve.

1.6.A Out-of-sample predictability of bond returns

Suppose an investor perceives the process generating the slow variation in yields as being driven by the long-run inflation expectations, and estimates the persistent factor τ_t^{CPI} using core CPI. In doing so, they exploit inflation information that is available only up to time t, and update the estimates of τ_t^{CPI} with the gain parameter of 0.9868.

We consider three out of sample periods starting in 1978:01, 1985:01 and 1995:01, ending in 2009:12. For each of the samples, we obtain the initial estimates based on the period from 1971:11 until 1977:01, until 1984:01 and until 1994:01, respectively. With information up to this point, say t_0, we obtain cycles as in equation (1.13), and run regression (1.14) predicting excess returns realized up t_0 using cycles up to t_0 less 12 months. At the estimated parameters, we then predict excess returns 12 months ahead, i.e. realized at t_0 plus 12 months. We extend the sample month-by-month, and repeat these steps until we reach the maximum sample length. The performance of cycles is compared to that of forward rates and the slope.

Our out-of-sample evaluation involves three measures (see Appendix A.9 for implementation details). We start with the encompassing test (ENC-NEW) proposed by Clark and McCracken (2001). By results of Section 1.3.D, we treat cycles as an unrestricted model and forward rates as a restricted one. The null hypothesis of the ENC-NEW test is that the restricted model (forwards) encompasses all the predictability in bond excess returns, and it cannot be further improved by the unrestricted model (cycles). Clark and McCracken (2005) show that the ENC-NEW test statistic has a non-standard distribution under the null, therefore we obtain the critical values by bootstrapping.

The second measure is the ratio of mean squared errors implied by the unrestricted versus restricted model, $\text{MSE}_{\text{cyc}}/\text{MSE}_{\text{fwd}}$. A number less than one indicates that the unrestricted model is able to generate lower prediction errors.

Finally, the third measure is the out-of-sample R^2 proposed by Campbell and Thompson (2008), R^2_{OOS}. R^2_{OOS} compares the forecasting performance of a given predictor toward a "naive" forecast obtained with the historical average return. The statistic is analogous to the in-sample R^2: Its positive value indicates that the predictive model has a lower mean-squared prediction error than the "naive" forecast.

Throughout, for forward rate regressions, we use forward rates with maturities of one, two, five, seven, ten and 20 years as predictors. For cycle regressions (except the ENC test), we use the short maturity cycle $c_t^{(1)}$ and the average cycle $\bar{c}_t$ as predictors in bivariate regressions.[19] For the sake of comparability with the forward rate regressions, in the ENC test we simply employ six cycles with the same maturities as the forward rates. For slope regressions, we forecast excess return on the n-year bond with the corresponding spot forward spread, $f_t^{(n)} - y_t^{(1)}$. The slope regressions provide a useful benchmark out of sample because, in contrast to the forward rate regressions, they do not require an estimation of a large number of coefficients.

The panels of Table 1.10 report the results for 1971–2009, 1985–2009, and 1995–2009, respectively. The ENC-NEW test rejects the null hypothesis for all maturities at the 95% confidence level: The cycles' model significantly improves the predictive performance over forwards. The MSE ratio, $\text{MSE}_{\text{cyc}}/\text{MSE}_{\text{fwd}}$, is reliably below one for all maturities. In the recent sample, the $\text{MSE}_{\text{cyc}}/\text{MSE}_{\text{fwd}}$ ratio is substantially lower than in the period 1971–2009. Indeed, while in the last two decades the performance of forward rates deteriorates compared to the full sample, the performance of the cycles remains relatively stable. With one exception, R^2_{OOS} values obtained with cycles are large and positive for all maturities across all sample periods. In summary, the out-of-sample statistics support the previous in-sample evidence, indicating the relevance of the economic mechanism that the cycles capture.

[Table 1.10 here.]

Additionally, in Appendix A.7.D we show that the out-of-sample results are only weakly influenced when varying the learning parameter between 0.975 (fast updating) and 0.995 (very

[19] The results remain almost identical if we first construct $\widehat{cf}$ as in Section 1.3.C, and then use it for predicting returns out of sample.

slow updating). Based on the literature which we have summarized above, this range for the gain parameter can be viewed as covering the extremes.

1.6.B Other data sets

One may be concerned that the return predictability we document is contingent upon the CMT rates, and the way we construct the zero curve. To show that our results are robust to these choices, we perform the predictive exercise on other two commonly used data sets constructed by Fama and Bliss (FB) and Gürkaynak, Sack, and Wright (2006, GSW). We remain conservative on two fronts. First, we focus on the range of maturities from one to five years, as dictated by the FB data. Second, to assess the sensitivity of our results to the recent crisis, we consider two samples: (*i*) excluding the crisis 1971–2006, and (*ii*) including the crisis 1971–2009. Note that the data sets we consider differ not only in the way of constructing the zero curve, but also in the choice of the underlying yields. For instance, CMT yields are based on the on-the-run securities while GSW yields are off-the-run. We are therefore able to assess if our conclusions are driven by the liquidity premium pertaining to the on-the-run curve.

Table 1.11 displays the predictive R^2's across the three data sets. As a summary statistic, we regress the average excess return (across maturities), $\overline{rx}_{t+1} = \frac{1}{4}\sum_{i=2}^{5} rx_{t+1}^{(i)}$, on each of the variables indicated in the first column of the table. Rows (1) and (2) in each panel consider cycles as regressors, rows (3) and (4)—yields and forward rates, rows (5) and (6)—spreads of cycles and yields. The columns denoted as "sample" give the adjusted R^2 values for the regressions, and the columns denoted as "bootstrap" provide the 5%, 50% and 95% bootstrapped percentile values for the R^2.

[Table 1.11 here.]

The forecasting ability of the cycles is confirmed across all data sets. Even though we use a restricted number of maturities, the R^2's obtained with the cycles are in the 50% range.[20] Using yields and forward rates, or spreads leads to clearly inferior predictability, diminishing the R^2's at least by half. The gap between cycles and other predictors becomes even more apparent when we include the crisis years. While the recent turmoil leads to a weakened performance across

[20] Unreported results show that the one to five year yield maturity range that we use here can be restrictive. For example, in the Greenspan's subperiod we find that the inclusion of cycles with longer maturities improves the forecasting performance.

all regressors, with forward rates explaining just about 17% of variation in $\overline{rx}_{t+1}$, the predictive power of the cycles still remains confidently above 45%.

1.7 Conclusions

The essential observation of this paper is concerned with the role of frequencies in the yield curve and how they encode different economic forces at work. In a first step, we split these effects into (i) a smooth and slow adjustment related to the changing long-run mean of inflation, and (ii) transitory fluctuations—cycles—around the smooth component reflecting current macro-finance conditions. The cycles across different maturities combine the term structure of transitory short rate expectations with the term structure of risk premia. Using their cross-sectional composition, in the second step, we distill these two elements into separate factors. Those steps leave us with three observable variables: the persistent and transitory short rate expectations, and the term premium factor, $\widehat{cf}$. These factors explain 99.7% of variation in yields across maturities, and summarize key economic frequencies in the yield curve, which we respectively term as the generational frequency, the business cycle frequency and the risk premium frequency.

The term premium factor $\widehat{cf}$ has strong predictive properties for future bond excess returns. We justify this fact in several ways. First, the interpretation of cycles as "risk premium plus transitory short rate expectations" emerges naturally from substituting a Taylor rule into the basic yield curve equation. Second, we argue that cycles present stationary deviations from the long-run relationship between yields and the persistent component of short rate expectations.

Our decomposition facilitates a number of findings. First, we show that the predictability of bond excess returns using one factor, $\widehat{cf}$, is significantly higher than documented so far in the literature. The return forecasting factor is visible in the cross-section of yields, and its average impact on the curve exceeds the one of both slope and curvature in the usual PCA framework. Second, we propose an alternative interpretation of the level effect in the yield curve: We show that the level type of shock, i.e. a shock that is uniform across maturities, is driven by the persistent inflation expectations component. We point out that the traditional level ($PC1$) contains nontrivial information about the term premia. However, when trying to predict excess returns, this information remains unexploited because it is overwhelmed by the persistent variation that the level embeds. Third, and related, once we account for the predictive content in the level, the slope and higher-order PCs tend to lose significance for forecasting excess bond returns. Finally, conditioning on $\widehat{cf}$, we are able to revisit the additional role of macroeconomic risks

in term premia. We show that, to the first order, $\widehat{cf}$ subsumes the key part of predictability contained in a broad panel of macroeconomic indicators.

We subject these conclusions to several robustness checks. We find that the predictive power of the cycles is not affected by the choice of the data set, the procedure used to construct the zero curve, and the inclusion of the monetary experiment or the recent financial crisis. We also show that our forecasting factor provides stable and positive out-of-sample performance. Taken together, these results indicate that the yield decomposition we propose captures a highly relevant characteristic of the bond market data.

Table 1.1: Modified Taylor rule (OLS)

The table reports the parameter estimates for the modified (**panel A**) and restricted (**panel B**) version of the Taylor rule for three sample periods. τ_t^{CPI} is computed as a discounted moving average of the last ten years of core CPI data. CPI_t^c is the cyclical component of annual inflation, $CPI_t^c = CPI_t - \tau_t^{CPI}$, and $UNEMPL_t$ denotes unemployment. The restriction in panel B is that CPI_t^c and τ_t^{CPI} share the same coefficient. The 1971–2009 sample includes the Volcker period. We split it into two parts: before and after the disinflation, 1971:11–1984:12 and 1985:01–2009:12. The short rate is represented by the monthly average of the effective Fed funds rate. All t-statistics (in parentheses) are obtained using Newey-West adjustment with 15 lags.

Panel A. Unrestricted rule

$r_t = \gamma_0 + \gamma_c CPI_t^c + \gamma_y UNEMPL_t + \gamma_\tau \tau_t^{CPI} + \varepsilon_t$

Coefficient	1971-2009	1985-2009	1971-1984
γ_c	0.53	0.92	0.44
	(4.38)	(7.46)	(4.15)
γ_y	-1.41	-1.71	-1.47
	(-5.44)	(-15.66)	(-3.57)
γ_τ	2.23	2.16	2.59
	(11.80)	(21.43)	(5.97)
$\bar{R}^2$	0.79	0.91	0.61

Panel B. Restricted rule

$r_t = \gamma_0 + \gamma_\pi (CPI_t^c + \tau_t^{CPI}) + \gamma_y UNEMPL_t + \varepsilon_t$

Coefficient	1971-2009	1985-2009	1971-1984
γ_π	1.07	2.20	0.76
	(6.19)	(10.00)	(3.25)
γ_y	-0.20	-1.26	0.07
	(-0.51)	(-6.31)	(0.16)
γ_τ	–	–	–
$\bar{R}^2$	0.56	0.76	0.30

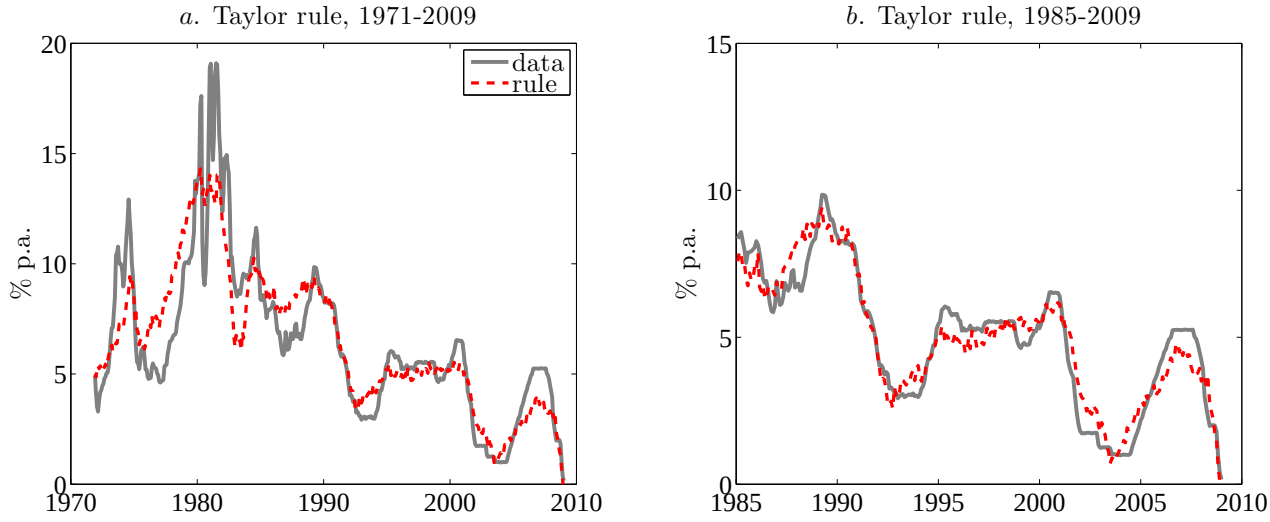


Figure 1.1: Fit of the modified Taylor rule (OLS)

The figure plots observed and fitted Fed funds rate for two sample periods: 1971–2009 (panel *a*) and 1985-2009 (panel *b*). The fit to the Fed funds rate is obtained by estimating the Taylor rule specification given as $r_t = \gamma_0 + \gamma_c CPI_t^c + \gamma_y UNEMPL_t + \gamma_\tau \tau_t^{CPI} + \varepsilon_t$, and corresponds to panel A in Table 1.1.

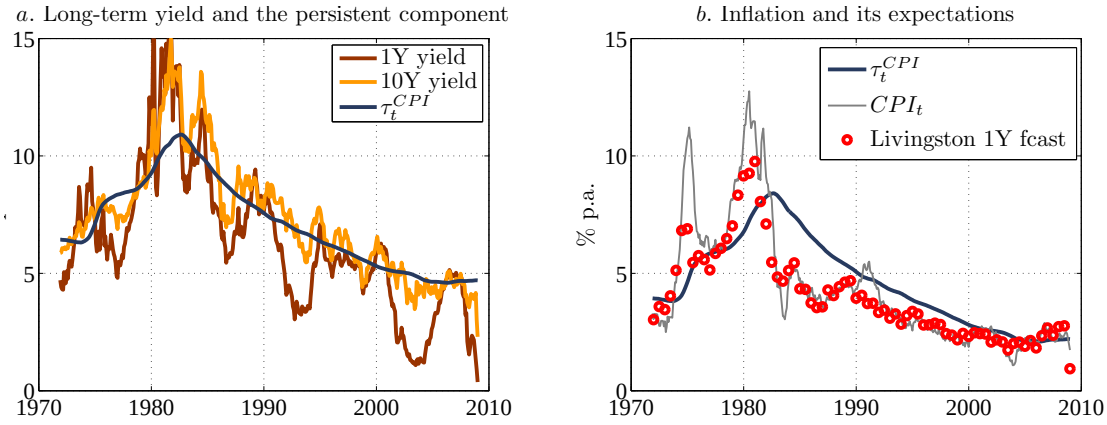


Figure 1.2: The persistent factor, τ_t^{CPI}

Panel *a* superimposes the one- and ten-year yield with τ_t^{CPI}. τ_t^{CPI} is constructed as the discounted moving average of the core CPI in equation (1.8), with sums truncated at $N = 120$ months and the discount factor $v = 0.9868$. τ_t^{CPI} is fitted to yields so that all variables match in terms of magnitudes. Panel *b* plots the one-year ahead median inflation forecasts from the Livingston survey and realized core CPI inflation.

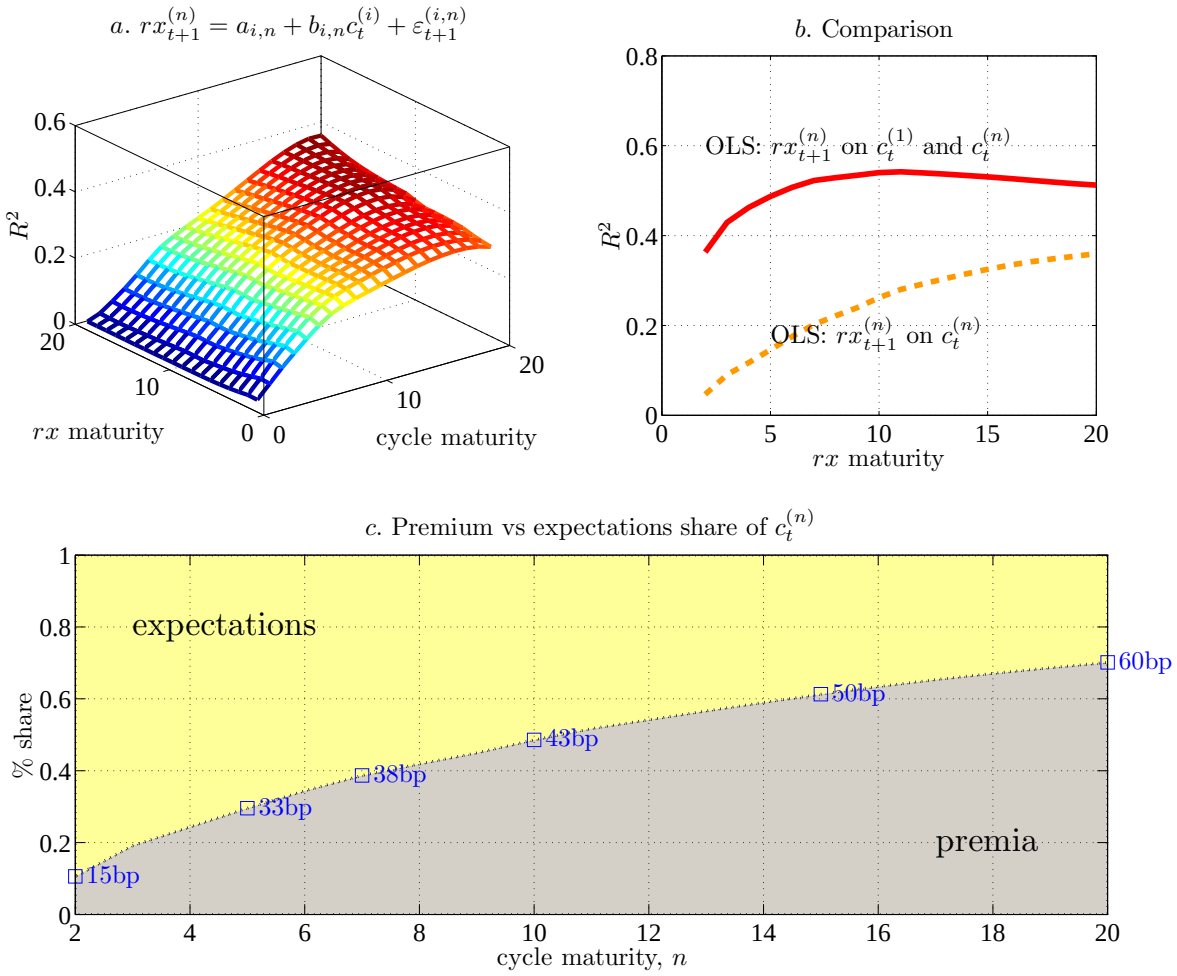


Figure 1.3: The anatomy of the cycle

Panel a plots the R^2's from a univariate predictive regression of $rx_{t+1}^{(n)}$ on yield cycles $c_t^{(i)}$ with different maturities, $i = 1, \ldots, 20$ years. Panel b compares the $\bar{R}^2$'s obtained by regressing $rx_{t+1}^{(n)}$ on $c_t^{(n)}$ (i.e. the diagonal of panel a) versus the $\bar{R}^2$'s obtained by regressing $rx_{t+1}^{(n)}$ on $c_t^{(n)}$ and $c_t^{(1)}$. Panel c decomposes the amount of variation in $c_t^{(n)}$ associated with the transitory short rate expectations and the premia. The decomposition into $R_p^{2,(n)}$ and $R_{ex}^{2,(n)}$ follows equation (1.17). The squares show the term premium share of cycles' variation in basis points for maturities two, five, seven, ten, 15 and 20 years. The numbers are obtained as: $R_p^{2,(n)} \times \text{std}(c_t^{(n)})$, where $\text{std}(c_t^{(n)})$ is the sample standard deviation of the n-maturity cycle.

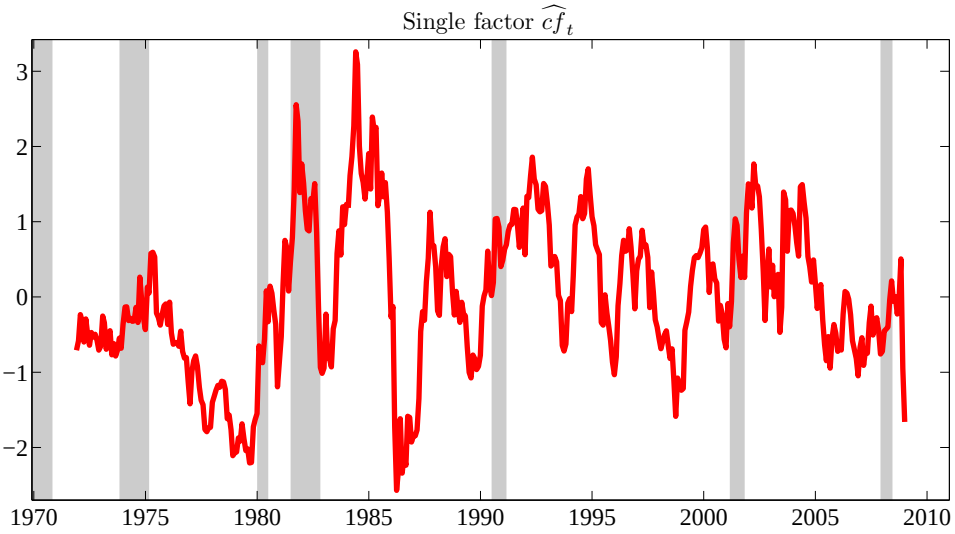


Figure 1.4: Single factor

The figure displays the return forecasting factor $\widehat{cf}_t$ formed with equation (1.19). Shaded areas mark the NBER recessions. The series has been standardized.

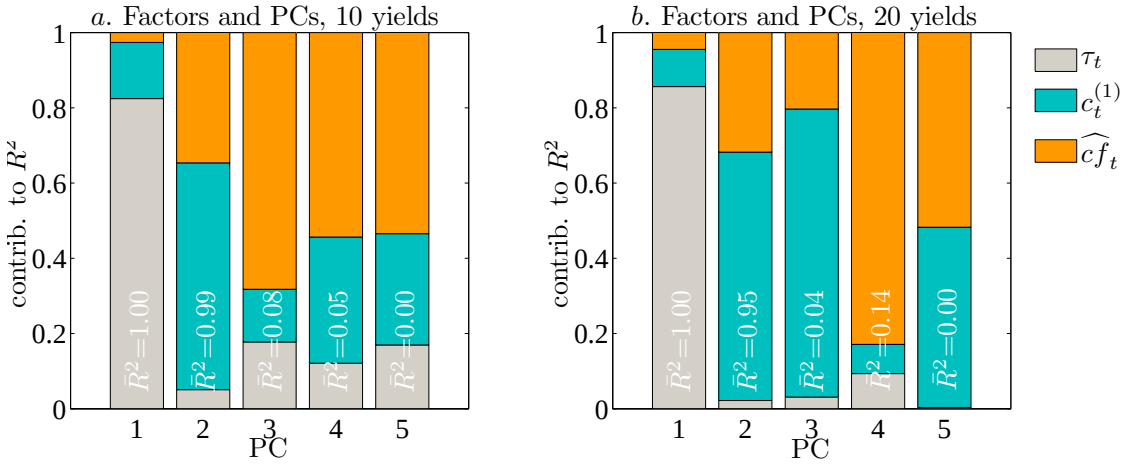


Figure 1.5: Contributions of τ_t^{CPI}, $c_t^{(1)}$ and $\widehat{cf}_t$ to explained variance of PCs

The figure plots the contributions of τ_t^{CPI}, $c_t^{(1)}$ and $\widehat{cf}_t$ to the explained variance of the respective principal components. The total explained variance (R^2) is reported in each bar. The contribution of each factor is computed using Shapley decomposition. In panel *a*, principal components are obtained from ten yields with maturities between one and ten years. Panel *b* reports the same results but obtained using yields with maturity up to 20 years.

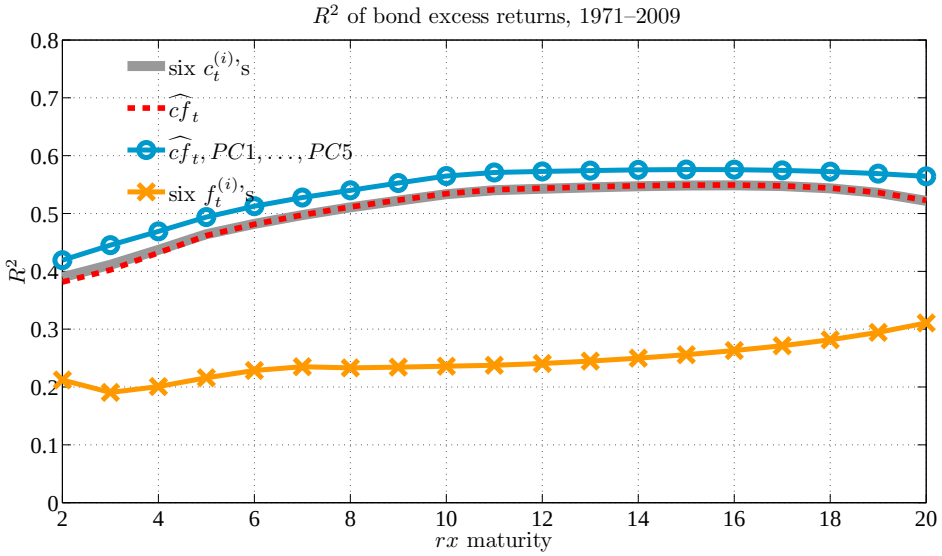


Figure 1.6: Comparing the R^2's

The figure juxtaposes the adjusted R^2's of different predictive regressions. Lines denoted "six $c_t^{(i)}$'s" correspond to the unrestricted regression of excess returns on six cycles in equation (1.14) (Table 1.4). Lines marked as "$\widehat{cf}_t$" correspond to the restricted regression using the single factor, as constructed in equation (1.19) (Table 1.5). Finally, lines labeled "$\widehat{cf}_t, PC1_t, \ldots, PC5_t$" correspond to regressing excess returns on the single factor and five PCs of yields (see Table 1.6).

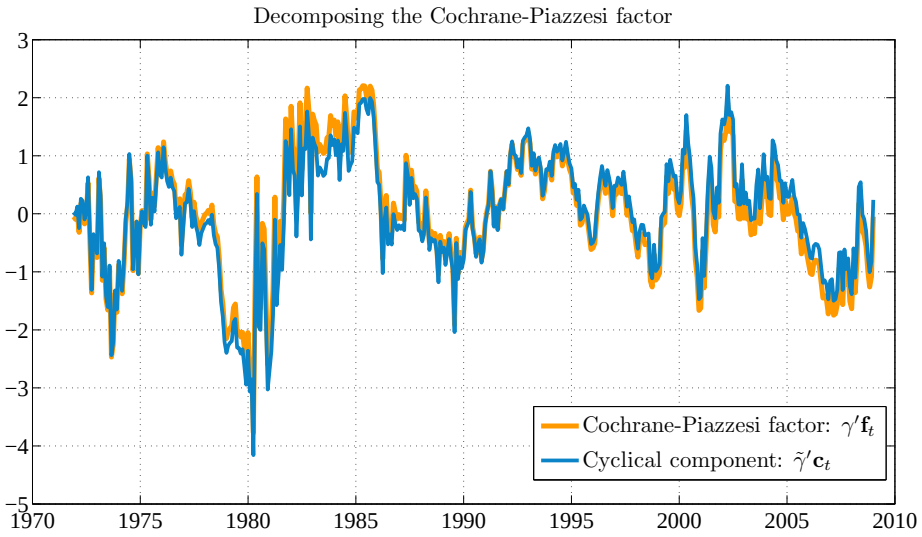


Figure 1.7: Decomposing the Cochrane-Piazzesi factor

The figure superimposes the single forecasting factor $\gamma'\mathbf{f}_t$ as constructed by Cochrane and Piazzesi (2005) with its cyclical component $\widetilde{\gamma}'\mathbf{c}_t$. The decomposition is stated in equation (1.23): $\gamma'\mathbf{f}_t = \overline{\gamma}'\mathbf{1}\tau_t + \widetilde{\gamma}'\mathbf{c}_t$. For comparison, both variables are standardized. We use ten forward rates with maturities one to ten years to construct the CP factor.

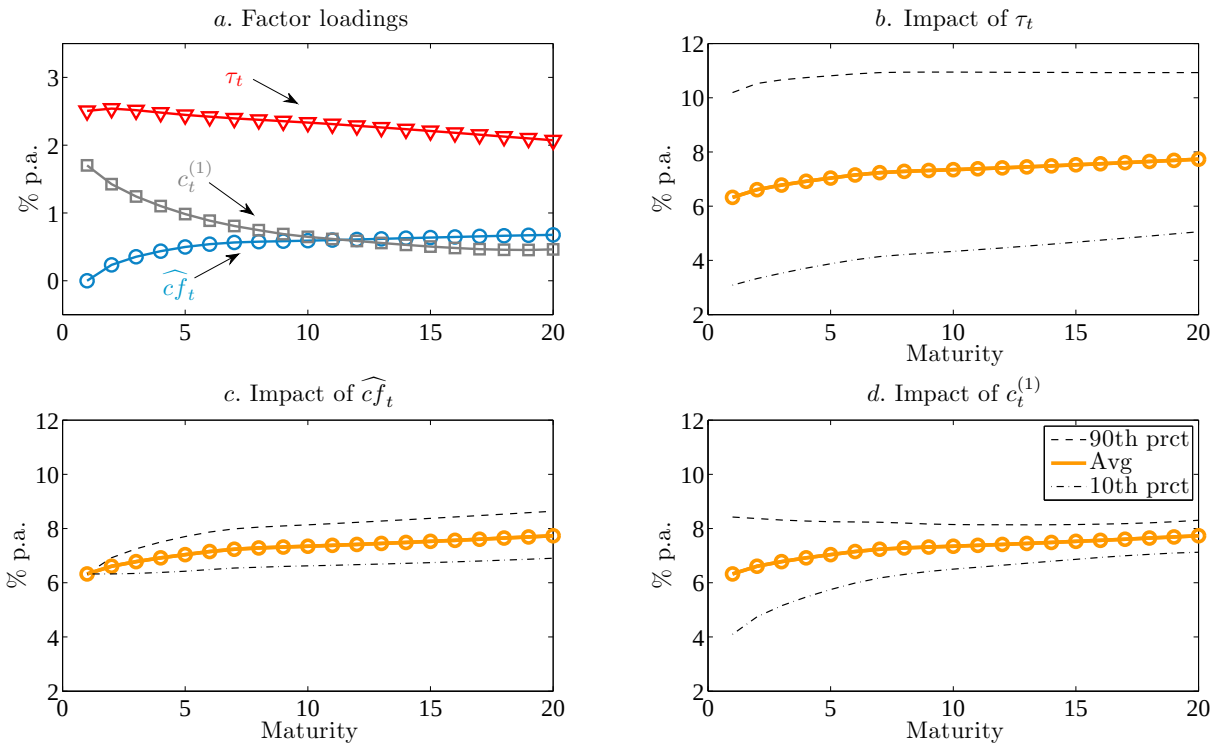


Figure 1.8: Cross-sectional impact of factors on yields

The figure discusses the implications of the observable factors for the cross section of yields introduced in Section 1.4. Panel *a* displays the cross-sectional impact of each factor $X_t = (\tau_t^{CPI}, \widehat{cf}_t, c_t^{(1)})$. To make the impacts comparable, loadings are multiplied by the standard deviation of the respective factor. The loadings are obtained from the regression of yields on factors in equation (1.28). Panels *b* through *d* show the reaction of the yield curve to factor perturbations. The solid line is generated by setting all variables to their unconditional means. The circles indicate maturities used in estimation. The dashed lines are obtained by setting a given state variable to its 10th and 90th percentile, respectively, and holding the remaining factors at their unconditional average. The sample period is 1971–2009.

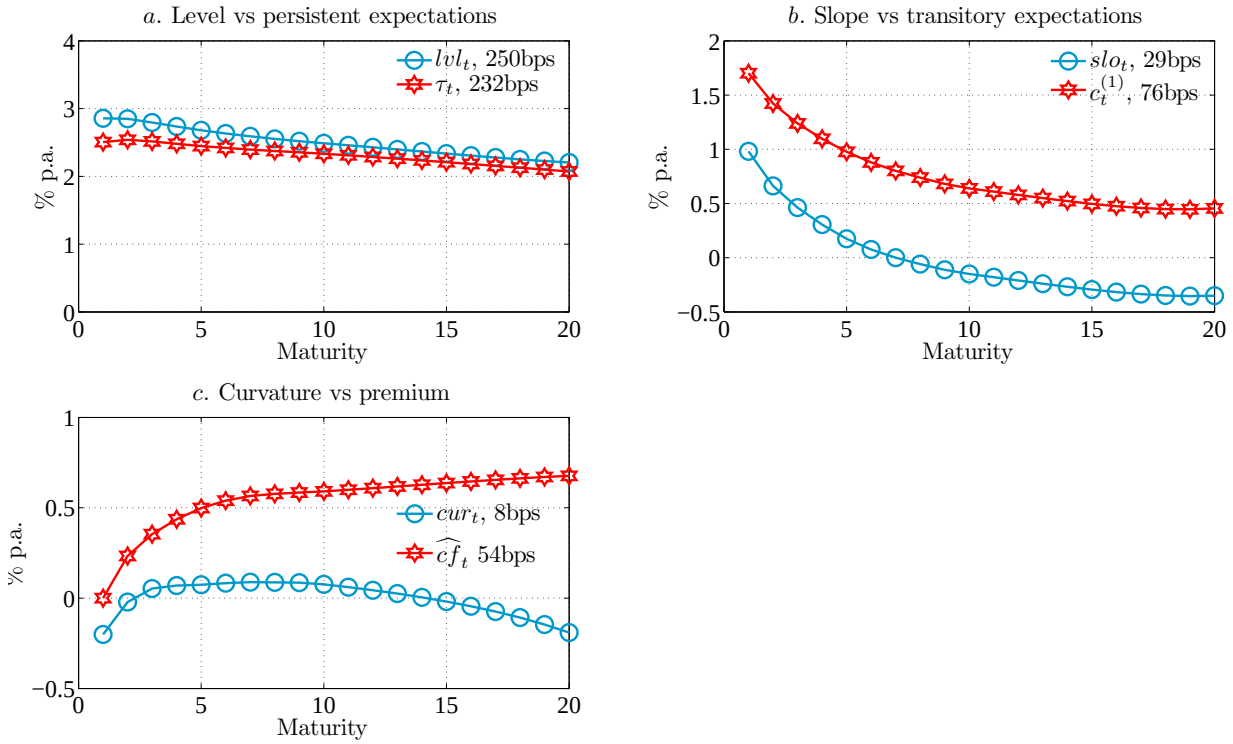


Figure 1.9: Comparing the cross-sectional impact of factors and PCs

The figure shows the cross-sectional impact of the three PCs: lvl_t, slo_t, cur_t, and compares them with the persistent and transitory expectations and the term premium factors: $\tau_t^{CPI}, c_t^{(1)}, \widehat{cf}_t$. Loadings are estimated with the OLS regressions of yields on each set of factors. The legend in each plot reports the average absolute impact of one standard deviation change in the factor on yields across different maturities. The sample period is 1971–2009.

Table 1.2: Estimates of the vector error correction model

The table reports the estimated coefficients from the error correction model on monthly frequency:

$$\Delta y_t^{(n)} = a_c c_{t-\Delta t}^{(n)} + a_y \Delta y_{t-\Delta t}^{(n)} + a_\tau \Delta \tau_{t-\Delta t} + a_0 + \varepsilon_t, \quad \Delta t = 1 \text{ month}$$

Reported t-statistics use Newey-West adjustment with 12 lags. For ease of comparison, all variables are standardized. $\Delta \overline{y}_t$ in the last column denotes the average yield change across maturities.

	Dependent variable						
Regressor	$\Delta y_t^{(1)}$	$\Delta y_t^{(2)}$	$\Delta y_t^{(5)}$	$\Delta y_t^{(7)}$	$\Delta y_t^{(10)}$	$\Delta y_t^{(20)}$	$\Delta \overline{y}_t$
$c_{t-\Delta t}^{(n)}$	-0.19	-0.20	-0.21	-0.20	-0.20	-0.20	-0.20
	(-2.37)	(-2.84)	(-3.54)	(-3.78)	(-3.94)	(-4.06)	(-3.64)
$\Delta y_{t-\Delta t}^{(n)}$	0.22	0.22	0.18	0.14	0.14	0.12	0.19
	(3.02)	(4.41)	(4.22)	(3.15)	(3.28)	(2.71)	(3.99)
$\Delta \tau_{t-\Delta t}$	0.07	0.07	0.09	0.08	0.09	0.06	0.09
	(1.16)	(1.36)	(1.66)	(1.62)	(1.75)	(1.22)	(1.65)
$\bar{R}^2$	0.06	0.06	0.06	0.05	0.05	0.04	0.06

Table 1.3: Bond excess returns: summary statistics

Panel A reports summary statistics for bond excess returns. **Panel B** reports the mean and standard deviation of duration-standardized excess returns to remove the effect of duration. AR(1) denotes the first order autocorrelation coefficient. **Panel C** reports the correlation between excess returns of different maturities. Bond excess returns are computed at an annual horizon as $rx_{t+1}^{(n)} = p_{t+1}^{(n-1)} - p_t^{(n)} - y_t^{(1)}$ and multiplied by 100.

Panel A. Bond excess returns

	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
Mean	0.69	1.61	1.97	1.99	2.57	3.05
Stdev	1.99	6.29	8.60	11.69	16.99	22.59
AR(1)	0.94	0.93	0.93	0.93	0.93	0.92

Panel B. Duration standardized excess returns

	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
Mean	0.34	0.32	0.28	0.20	0.17	0.15
Stdev	1.00	1.26	1.23	1.17	1.13	1.13

Panel C. Correlation of excess returns

	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
$rx^{(2)}$	1.00	–	–	–	–	–
$rx^{(5)}$	0.95	1.00	–	–	–	–
$rx^{(7)}$	0.91	0.99	1.00	–	–	–
$rx^{(10)}$	0.86	0.96	0.99	1.00	–	–
$rx^{(15)}$	0.81	0.93	0.96	0.99	1.00	–
$rx^{(20)}$	0.76	0.87	0.91	0.94	0.97	1.00

Table 1.4: First look at predictive regressions of bond returns

The table reports the results of predictive regressions in equation (1.14). In the first row, we provide adjusted R^2 values. To assess the small sample (**SS**) properties of $\bar{R}^2$, the next three rows give its 5%, 50% and 95% percentile values obtained with the block bootstrap (see Appendix A.4). The $\chi^2(6)$ tests if the coefficients (excluding the constant) are jointly equal to zero. We report the Hansen-Hodrick (**HH**) and the Newey-West (**NW**) correction, using 12 and 15 lags, respectively. "**LS**" means that the statistics were estimated using the full sample. The row "$\chi^2(6)$ (SS 5%)" states the lower 5% bound on the values of the χ^2-test (using NW adjustment) obtained with the bootstrap. We also provide conservative standard errors obtained using the reverse regression delta method (**rev.reg.**) of Wei and Wright (2010) and the corresponding p-values. The last five rows summarize the corresponding results for the forward rate regressions. Cycles $\mathbf{c}_t$ and forward rates $\mathbf{f}_t$ are of maturities one, two, five, seven, ten, and 20 years. Sample is 1971–2009.
The asymptotic 1%, 5%, and 10% critical values for $\chi^2(6)$ are 16.81, 12.59, and 10.64, respectively.

Statistic	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
Cycle regressions: $rx_{t+1}^{(n)} = \delta_0 + \delta' \mathbf{c}_t + \varepsilon_{t+1}^{(n)}$						
$\bar{R}^2$	0.42	0.49	0.52	0.55	0.56	0.57
$\bar{R}^2$ (SS,5%)	0.31	0.37	0.40	0.44	0.45	0.44
$\bar{R}^2$ (SS,50%)	0.47	0.53	0.56	0.59	0.59	0.59
$\bar{R}^2$ (SS,95%)	0.61	0.65	0.67	0.68	0.68	0.69
$\chi^2(6)$ (LS, HH)	46.98	117.38	149.18	182.40	181.22	149.69
$\chi^2(6)$ (LS, NW)	61.59	131.12	150.20	172.52	166.63	125.39
$\chi^2(6)$ (SS,5%)	48.01	86.10	95.26	116.47	116.08	87.13
$\chi^2(6)$ (rev. reg.)	13.19	24.21	28.56	35.80	36.66	30.02
pval	0.04	0.00	0.00	0.00	0.00	0.00
Forward-rate regressions: $rx_{t+1}^{(n)} = d_0 + d' \mathbf{f}_t + \varepsilon_{t+1}^{(n)}$						
$\bar{R}^2$	0.21	0.22	0.24	0.24	0.26	0.31
$\chi^2(6)$ (LS, HH)	23.82	26.07	23.13	22.77	23.20	20.77
$\chi^2(6)$ (LS, NW)	25.38	28.76	28.13	28.59	28.84	26.95
$\chi^2(6)$ (rev. reg.)	9.14	12.73	13.15	13.48	13.63	14.22
pval	0.17	0.05	0.04	0.04	0.03	0.03

Table 1.5: Predicting returns with the single forecasting factor

Panel A reports the estimates of equation (1.18). Rows denoted as "LS" give the full sample t-statistics and adjusted R^2's. Rows denoted as "SS" summarize the small sample distributions of the statistics obtained with the block bootstrap. **Panel B** shows the predictability of individual bond returns with the single factor. Again, the full sample (LS) and small sample (SS) distributions are provided. Row "$\Delta\bar{R}^2$" gives the difference in $\bar{R}^2$ values between the corresponding unconstrained predictive regressions using six cycles in Table 1.4 and the regressions using the single factor, $\widehat{cf}_t$. HH denotes Hansen-Hodrick adjustment in standard errors, NW denotes the Newey-West adjustment. We use 12 and 15 lags, respectively. Bootstrapped t-statistics use the NW adjustment with 15 lags to ensure a positive definite covariance matrix in all bootstrap samples. To facilitate comparisons, in panel B all left- and right-hand variables have been standardized.

Panel A. Constructing the single factor: $\bar{c}_t = \gamma_1 c_t^{(1)} + \bar{\varepsilon}_t$, where $\bar{c}_t = \frac{1}{m-1}\sum_{i=2}^{m} c_t^{(i)}$

	$\hat{\gamma}_1$	tstat (HH, NW)	R^2
LS	0.42	(7.58, 8.74)	0.62
SS (5%, 50%, 95%)		[7.10, 9.79, 13.71]	[0.47, 0.61, 0.72]

Panel B. Single factor predictive regression: $rx_{t+1}^{(n)} = \beta_0 + \beta_1 \widehat{cf}_t + \varepsilon_{t+1}^{(n)}$, where $\widehat{cf}_t = \bar{c}_t - \hat{\gamma}_1 c_t^{(1)}$

Statistic	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
β_1	0.62	0.68	0.71	0.74	0.74	0.72
tstat (LS, HH)	5.91	8.83	9.55	10.37	10.16	9.14
tstat (LS, NW)	6.75	9.64	10.12	10.72	10.38	9.10
tstat (SS,5%)	4.17	5.85	6.47	7.00	7.13	6.65
tstat (SS,50%)	7.26	9.74	10.13	10.65	10.45	9.28
tstat (SS,95%)	12.21	14.66	14.28	14.52	13.88	12.31
$\bar{R}^2$ (LS)	0.39	0.47	0.50	0.54	0.55	0.52
ΔR^2 (LS)	0.03	0.02	0.02	0.01	0.01	0.05
$\bar{R}^2$ (SS, 5%)	0.20	0.29	0.33	0.37	0.39	0.37
$\bar{R}^2$ (SS, 50%)	0.37	0.46	0.49	0.52	0.54	0.51
$\bar{R}^2$ (SS, 95%)	0.53	0.59	0.61	0.64	0.64	0.61

Table 1.6: The link between the level and the return forecasting factor

Panel A reports the unconditional correlation of the cycle obtained from the level factor (c_t^{lvl}) with the PCs and the average cycle ($\overline{c}_t$). c_t^{lvl} is obtained from the decomposition (1.30). Last column in panel A states the correlation of $\widehat{cf}_t^{lvl}$ and $\widehat{cf}_t$. **Panel B** reports the results for predictive regressions including $\widehat{cf}_t$ and five principal components $\mathbf{PC}_t = (PC1_t, \ldots, PC5_t)'$ of yields. "$\Delta\bar{R}^2$" denotes the increase in $\bar{R}^2$ by including five principal components in the predictive regression on top of $\widehat{cf}_t$. In panel B, t-statistics are in parentheses and are computed using the Newey-West adjustment with 15 lags. All variables are standardized.

Panel A. Correlations

(c_t^{lvl}, lvl_t)	$(c_t^{lvl}, PC2_t)$	$(c_t^{lvl}, PC3_t)$	$(c_t^{lvl}, PC4_t)$	$(c_t^{lvl}, PC5_t)$	$(c_t^{lvl}, \bar{c}_t)$	$(\widehat{cf}_t^{lvl}, \widehat{cf}_t)$
0.38	-0.35	0.08	-0.28	0.00	1.00	1.00

Panel B. Predictive regressions: $rx_{t+1}^{(n)} = b_0 + b_1\widehat{cf}_t + \mathbf{b}_2'\mathbf{PC}_t + \varepsilon_{t+1}^{(n)}$

	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
$\widehat{cf}_t$	**0.62**	**0.72**	**0.73**	**0.77**	**0.75**	**0.66**
	(4.72)	(6.60)	(6.94)	(7.60)	(7.40)	(6.42)
$PC1$ (level)	0.07	-0.03	-0.04	-0.05	-0.05	-0.04
	(0.53)	(-0.26)	(-0.32)	(-0.41)	(-0.48)	(-0.38)
$PC2$ (slope)	-0.11	-0.11	-0.08	-0.06	-0.02	0.03
	(-0.96)	(-0.99)	(-0.67)	(-0.55)	(-0.15)	(0.26)
$PC3$ (curve)	-0.10	**-0.13**	**-0.14**	**-0.13**	-0.06	0.07
	(-1.25)	(-2.16)	(-2.53)	(-2.46)	(-1.16)	(1.18)
$PC4$	-0.12	-0.06	-0.06	-0.00	-0.03	-0.16
	(-1.16)	(-0.61)	(-0.65)	(-0.03)	(-0.34)	(-1.83)
$PC5$	0.06	0.06	0.06	**0.11**	**0.15**	**0.13**
	(0.87)	(1.06)	(1.16)	(2.41)	(3.59)	(3.04)
$\bar{R}^2$	0.42	0.49	0.53	0.56	0.58	0.56
$\Delta\bar{R}^2$	0.04	0.03	0.03	0.03	0.03	0.04

Table 1.7: Decomposing the forward-rate predictive regressions

We decompose the Cochrane-Piazzesi factor into a persistent and a cyclical component, and predict the average return (across maturities) $\overline{rx}_{t+1}$ using the two components as separate regressors (see Section 1.3.D). We report the coefficient estimates and t-statistics with Hansen-Hodrick (HH) and Newey-West correction (NW) using 12 and 15 lags, respectively. Column "$\bar{R}^2$" reports the adjusted R^2 from this regression. For comparison, column "$\bar{R}^2$ $(\gamma'\mathbf{f}_t)$" gives the $\bar{R}^2$ when Cochrane-Piazzesi factor is used as a predictor, and column "$\bar{R}^2$ $(\widehat{cf}_t)$"—when $\widehat{cf}_t$ is used. We construct $\gamma'\mathbf{f}_t$ from ten forward rates with maturities one to ten years. The same maturities are included when forming the $\widehat{cf}_t$. Accordingly, $\overline{rx}$ is the average of returns with maturities from two to ten years. All variables are standardized.

$\overline{rx}_{t+1} = a_0 + a_1(\overline{\gamma}'\mathbf{1}\tau_t) + a_2(\widetilde{\gamma}'\mathbf{c}_t) + \varepsilon_{t+1}$

a_1	t-stat (HH, NW)	a_2	t-stat (HH, NW)	$\bar{R}^2$	$\bar{R}^2$ $(\gamma'\mathbf{f}_t)$	$\bar{R}^2$ $(\widehat{cf}_t)$
-0.0446	(-0.29,-0.33)	0.55	(5.15,5.89)	0.30	0.26	0.50

Table 1.8: Marginal predictability of bonds excess returns by macro and liquidity factors

Panel A reports predictive regressions of bond excess returns on the single return forecasting factor $\widehat{cf}_t$ and eight macro factors proposed by Ludvigson and Ng (2009), $\widehat{F}_{1t}, \ldots, \widehat{F}_{8t}$. "$\Delta\bar{R}^2$" denotes the **gain** in adjusted R^2 from adding all eight macro factors to the predictive regression with $\widehat{cf}_t$. "$\bar{R}^2$ ($\widehat{F}_t$ only)" reports the adjusted R^2 values from regressing the excess returns on $\widehat{F}_{1t}, \ldots, \widehat{F}_{8t}$. Macro factors are constructed from 132 macroeconomic and financial series. The sample period is 1971:11–2007:12. Superscripts H, M, L at t-statistics indicate variables that are significant in the macro-only regression of rx on $\widehat{F}_t$ at 1%, 5% and 10%, respectively. **Panel B** reports the predictive regression of rx on $\widehat{cf}_t$ and output gap ("gap_t") proposed by Cooper and Priestley (2009). The sample period is 1971:11–2007:12. **Panel C** shows the predictive regressions of rx on $\widehat{cf}_t$ and a given liquidity or credit measure. Commercial paper spread is the difference between the yield on three-month commercial paper and the yield of three-month T-bill. Swap spread is the difference between ten-year swap rate and the corresponding CMT yield. T-bill 3M spread is the difference between the three month T-bill rate and the Fed funds target. FG liquidity factor, proposed by Fontaine and Garcia (2010), tracks the variation in funding liquidity. All variables are described in detail in Appendix A.2. The sample period is 1987:04–2007:12. In parentheses, t-statistics use the Newey-West adjustment with 15 lags. All variables are standardized. For ease of comparison, in Panel C we report the ratio $\frac{b_2}{b_1}$ of liquidity measures relative to $\widehat{cf}_t$.

Panel A. Macro factors: $rx_{t+1}^{(n)} = b_0 + b_1\widehat{cf}_t + \mathbf{b_2'}\widehat{\mathbf{F}}_\mathbf{t} + \varepsilon_{t+1}^{(n)}$, sample 1971-2007

Regressor	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
$\widehat{cf}_t$	0.52	0.60	0.63	0.67	0.69	0.68
	4.89	6.53	6.72	7.18	7.16	6.37
$\widehat{F}_{1t}$ (real)	0.17	0.08	0.04	0.01	-0.01	0.00
	$(1.66)^M$	(0.86)	(0.45)	(0.15)	(-0.15)	(-0.03)
$\widehat{F}_{2t}$ (financial spreads)	0.05	0.07	0.09	0.09	0.09	0.06
	$(1.03)^M$	$(1.53)^M$	$(2.00)^H$	$(2.02)^H$	$(1.91)^H$	$(1.14)^M$
$\widehat{F}_{3t}$ (inflation)	-0.03	-0.01	0.00	0.01	0.01	0.01
	(-1.58)	(-0.44)	(0.17)	(0.31)	(0.51)	(0.47)
$\widehat{F}_{4t}$ (inflation)	-0.15	-0.06	-0.02	0.01	0.05	0.06
	$(-2.12)^H$	$(-0.84)^M$	$(-0.23)^M$	$(0.17)^L$	(0.64)	(0.77)
$\widehat{F}_{5t}$	0.07	0.02	0.01	0.00	-0.01	0.00
	(1.17)	(0.35)	(0.19)	(-0.03)	(-0.18)	(-0.01)
$\widehat{F}_{6t}$ (monetary)	-0.09	-0.10	-0.10	-0.10	-0.10	-0.09
	$(-1.00)^H$	$(-1.06)^H$	$(-1.05)^H$	$(-1.17)^H$	$(-1.14)^H$	$(-0.92)^H$
$\widehat{F}_{7t}$ (bank reserves)	-0.04	-0.08	-0.08	-0.08	-0.09	-0.09
	$(-0.74)^M$	$(-1.24)^H$	$(-1.26)^H$	$(-1.45)^H$	$(-1.56)^H$	$(-1.47)^H$
$\widehat{F}_{8t}$ (stock market)	0.01	0.02	0.02	0.01	0.01	0.03
	(0.23)	$(0.72)^M$	$(0.66)^M$	$(0.43)^M$	$(0.52)^H$	$(1.18)^H$
$\bar{R}^2$	0.45	0.49	0.52	0.55	0.57	0.54
R^2 ($\widehat{F}_t$ only)	0.25	0.22	0.22	0.22	0.22	0.20
$\Delta\bar{R}^2 = \bar{R}^2 - \bar{R}^2(\widehat{cf}_t)$	0.05	0.02	0.01	0.02	0.02	0.02

Panel B. Output gap: $rx_{t+1}^{(n)} = b_0 + b_1\widehat{cf}_t + b_2\mathbf{gap}_t + \varepsilon_{t+1}^{(n)}$, sample 1971-2007

Regressor	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
gap_t	-0.14	-0.02	-0.01	0.00	0.00	0.02
	(-1.20)	(-0.25)	(-0.13)	(-0.05)	(0.04)	(0.28)
$\Delta\bar{R}^2 = \bar{R}^2 - \bar{R}^2(\widehat{cf}_t)$	0.01	0.00	0.00	0.00	0.00	0.00

Continued on the next page

Continued from the previous page

Panel C. Liquidity factors: $rx_{t+1}^{(n)} = b_0 + b_1\widehat{cf}_t + b_2\mathbf{liq}_t + \varepsilon_{t+1}^{(n)}$, sample 1987-2007

Regressor	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
ComPaper spread, $\frac{b2}{b1}$	0.19	0.09	0.12	0.11	0.09	0.08
	(0.91)	(0.64)	(1.06)	(1.25)	(1.09)	(1.06)
$\Delta\bar{R}^2 = \bar{R}^2 - \bar{R}^2(\widehat{cf}_t)$	0.01	0.00	0.01	0.01	0.01	0.00
TED spread, $\frac{b2}{b1}$	0.17	0.07	0.10	0.09	0.06	0.06
	(0.82)	(0.54)	(1.01)	(1.15)	(0.94)	(0.90)
$\Delta\bar{R}^2 = \bar{R}^2 - \bar{R}^2(\widehat{cf}_t)$	0.01	0.00	0.01	0.01	0.00	0.00
Swap spread, $\frac{b2}{b1}$	0.36	0.12	0.07	-0.03	-0.11	-0.10
	(1.78)	(0.93)	(0.79)	(-0.43)	(-1.77)	(-1.69)
$\Delta\bar{R}^2 = \bar{R}^2 - \bar{R}^2(\widehat{cf}_t)$	0.06	0.01	0.00	0.00	0.01	0.01
T-bill3M spread, $\frac{b2}{b1}$	-0.38	-0.18	-0.17	-0.12	-0.06	-0.03
	(-1.73)	(-1.35)	(-1.68)	(-1.51)	(-0.83)	(-0.43)
$\Delta\bar{R}^2 = \bar{R}^2 - \bar{R}^2(\widehat{cf}_t)$	0.06	0.02	0.02	0.01	0.00	0.00
FG liquidity factor, $\frac{b2}{b1}$	0.46	0.16	0.13	0.08	0.00	-0.07
	(2.30)	(1.32)	(1.37)	(1.14)	(-0.03)	(-0.92)
$\Delta\bar{R}^2 = \bar{R}^2 - \bar{R}^2(\widehat{cf}_t)$	0.10	0.02	0.01	0.00	0.00	0.00
Moodys Baa spread, $\frac{b2}{b1}$	-0.24	-0.24	-0.15	-0.10	-0.06	-0.01
	(-1.92)	(-2.83)	(-2.08)	(-1.44)	(-0.87)	(-0.18)
$\Delta\bar{R}^2 = \bar{R}^2 - \bar{R}^2(\widehat{cf}_t)$	0.02	0.04	0.02	0.01	0.00	0.00

Table 1.9: Macro risks and predictability: the case of the two-year bond

Panel A reports the predictive regression of one-year yield one year ahead $y_{t+1}^{(1)}$ on median survey forecast of one-year yield four quarters ahead $E_t^s y_{t+1}^{(1)}$. The forecast is obtained from the Blue Chip Financial Forecasts. **Panel B** reports the regression of prediction errors $y_{t+1}^{(1)} - E_t^s y_{t+1}^{(1)}$ on $\widehat{cf}_t$ and unemployment. The sample period is 1988:01–2007:12. In parentheses, t-statistics use the Newey-West adjustment with 15 lags. All variables in panel B are standardized.

Panel A. $y_{t+1}^{(1)} = b_0 + b_1 E_t^s y_{t+1}^{(1)} + \varepsilon_{t+1}$

Regressor	coef	t-stat
$E_t^s y_{t+1}^{(1)}$	0.89	5.68
$\bar{R}^2 = 0.48$		

Panel B. $y_{t+1}^{(1)} - E_t^s y_{t+1}^{(1)} = b_0 + b_1 UNEMPL_t + b_2\widehat{cf}_t + \varepsilon_{t+1}$

Regressor	coef	t-stat
$UNEMPL_t$	-0.46	-5.88
$\widehat{cf}_t$	-0.27	-2.72
$\bar{R}^2 = 0.33$		

Table 1.10: Out-of-sample tests

The table reports the results of out-of-sample tests for the period 1978–2009 (**panel A**), 1985–2009 (**panel B**), and 1995–2009 (**panel C**). Row (1) in each panel contains the ENC-NEW test. The null hypothesis is that the predictive regression with forward rates (restricted model) encompasses all predictability in bond excess returns. The null is tested against the alternative that cycles (unrestricted model) improve the predictability achieved by the forward rates. For forwards and cycles we use maturities of one, two, five, seven, ten and 20 years. Row (2) reports bootstrapped critical values (CV) for the ENC-NEW statistic at the 95% confidence level. Row (3) shows the ratio of mean squared errors for the unrestricted and restricted models, $\text{MSE}_{\text{cyc}}/\text{MSE}_{\text{fwd}}$. Rows (4), (5) and (6) report the out-of-sample R^2, R^2_{OOS}, defined in equation (A.37), for cycles, forwards and the yield curve slope, respectively. For forwards we use six maturities as above, for cycles we use $c_t^{(1)}$ and $\bar{c}_t$. The slope for predicting the bond return with maturity n is constructed as $f_t^{(n)} - y_t^{(1)}$. Implementation details for the out-of-sample tests are collected in Appendix A.9.

Test	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
Panel A. Out-of-sample period: 1978–2009						
(1) ENC-NEW	131.41	140.12	150.39	168.35	167.68	149.99
(2) Bootstrap 95% CV	74.77	65.22	64.04	61.39	61.72	63.27
(3) $\text{MSE}_{\text{cyc}}/\text{MSE}_{\text{fwd}}$	0.65	0.56	0.54	0.49	0.50	0.63
(4) R^2_{OOS} cyc	0.18	0.29	0.34	0.40	0.39	0.31
(5) R^2_{OOS} fwd	-0.25	-0.26	-0.22	-0.24	-0.23	-0.10
(6) R^2_{OOS} slope	0.10	0.09	0.08	0.08	0.10	0.10
Panel B. Out-of-sample period: 1985–2009						
(1) ENC-NEW	108.16	111.04	121.87	137.93	138.71	127.53
(2) Bootstrap 95% CV	42.36	41.10	44.50	43.01	45.24	46.74
(3) $\text{MSE}_{\text{cyc}}/\text{MSE}_{\text{fwd}}$	0.59	0.52	0.51	0.49	0.51	0.59
(4) R^2_{OOS} cyc	0.12	0.34	0.40	0.44	0.44	0.41
(5) R^2_{OOS} fwd	-0.48	-0.26	-0.18	-0.14	-0.09	-0.00
(6) R^2_{OOS} slope	0.07	0.07	0.06	0.09	0.10	0.11
Panel C. Out-of-sample period: 1995–2009						
(1) ENC-NEW	53.74	53.64	61.99	74.83	80.02	75.91
(2) Bootstrap 95% CV	20.35	22.01	24.12	24.57	29.71	39.26
(3) $\text{MSE}_{\text{cyc}}/\text{MSE}_{\text{fwd}}$	0.48	0.47	0.41	0.40	0.37	0.31
(4) R^2_{OOS} cyc	-0.06	0.10	0.18	0.20	0.27	0.37
(5) R^2_{OOS} fwd	-1.20	-0.92	-1.00	-1.00	-0.99	-1.06
(6) R^2_{OOS} slope	-0.20	-0.06	-0.08	-0.04	-0.04	-0.05

Table 1.11: Comparing predictive R^2 in different data sets

The table compares the predictive adjusted R^2's for three different zero curves obtained from: Fama-Bliss (FB), Gürkaynak, Sack, and Wright (2006, GSW), and Treasury constant maturity (CMT) rates. The dependent variable is:

$$\overline{rx}_{t+1} = \frac{1}{4}\sum_{i=2}^{5} rx_{t+1}^{(i)}, \tag{1.35}$$

and regressors are indicated in the first column. In both panels, row (1) uses two cycles with maturity one and five years, row (2): five cycles with maturities from one through five years, row (3): two yields with maturity one and five years, row (4): five forward rates with maturity one through five years, row (5): spread between five- and one-year cycle, row (6): spread between five- and one-year yield. The column "sample" provides adjusted R^2's for each regression; the column "bootstrap" gives the 5%, 50% and 95% percentile values for the adjusted R^2's obtained with the block bootstrap (Appendix A.4).

	CMT		GSW		FB	
Regressor	Sample	Bootstrap	Sample	Bootstrap	Sample	Bootstrap
Panel A. Pre-crisis: 1971–2006						
(1) $c^{(1)}, c^{(5)}$	0.53	[0.39, 0.53, 0.66]	0.53	[0.40, 0.54, 0.66]	0.51	[0.38, 0.51, 0.64]
(2) $c^{(1)}, ..., c^{(5)}$	0.54	[0.42, 0.55, 0.68]	0.53	[0.42, 0.55, 0.67]	0.56	[0.44, 0.57, 0.69]
(3) $y^{(1)}, y^{(5)}$	0.22	[0.10, 0.25, 0.44]	0.19	[0.08, 0.23, 0.42]	0.19	[0.08, 0.22, 0.42]
(4) $f^{(1)}, ..., f^{(1)}$	0.27	[0.18, 0.32, 0.49]	0.21	[0.12, 0.27, 0.46]	0.30	[0.21, 0.34, 0.49]
(5) $c^{(5)} - c^{(1)}$	0.13	[0.02, 0.13, 0.29]	0.11	[0.01, 0.12, 0.27]	0.11	[0.01, 0.12, 0.27]
(6) $y^{(5)} - y^{(1)}$	0.13	[0.02, 0.14, 0.30]	0.11	[0.01, 0.12, 0.29]	0.11	[0.01, 0.12, 0.28]
Panel B. Post crisis: 1971–2009						
(1) $c^{(1)}, c^{(5)}$	0.47	[0.32, 0.49, 0.62]	0.46	[0.31, 0.49, 0.61]	0.44	[0.30, 0.47, 0.60]
(2) $c^{(1)}, ..., c^{(5)}$	0.47	[0.34, 0.51, 0.65]	0.47	[0.34, 0.50, 0.63]	0.48	[0.35, 0.53, 0.65]
(3) $y^{(1)}, y^{(5)}$	0.15	[0.05, 0.19, 0.37]	0.13	[0.04, 0.17, 0.35]	0.13	[0.04, 0.16, 0.35]
(4) $f^{(1)}, ..., f^{(1)}$	0.17	[0.10, 0.25, 0.42]	0.14	[0.08, 0.21, 0.38]	0.21	[0.13, 0.27, 0.43]
(5) $c^{(5)} - c^{(1)}$	0.11	[0.00, 0.10, 0.25]	0.09	[0.00, 0.09, 0.23]	0.09	[0.00, 0.09, 0.23]
(6) $y^{(5)} - y^{(1)}$	0.11	[0.00, 0.11, 0.26]	0.09	[0.00, 0.10, 0.25]	0.09	[0.00, 0.09, 0.24]

Appendix A

A.1 Cointegration

In Section 1.2.C, we invoke cointegration to argue that cycles should predict bond returns. This Appendix provides unit root tests for yields, τ_t^{CPI} and residuals from the cointegrating regression (1.10).

Table A.1 reports values of the augmented Dickey-Fuller (ADF) test. We consider changes in respective variables up to lag 12 as indicated in the first column. Tests in panel A are specified with a constant since all series have nonzero mean. Tests in panel B are specified without a constant since the cointegration residuals are zero mean by construction. Each panel provides the corresponding critical values. Additionally, we also apply the Phillips-Perron test and find that it conforms very closely with the ADF test. Therefore, we omit these results for brevity. The tests indicate that: (i) we cannot reject the hypothesis that both yields and τ_t have a unit root, (ii) that cointegraton residuals (cycles) are stationary.

A.2 Data

This section describes the construction of data series and compares bond excess returns obtained from different data sets: Gürkaynak, Sack, and Wright (2006, GSW), Fama-Bliss (FB) and constant maturity Treasury rates (CMT).

Interest rate data:

– *CMT rates.* We use constant maturity Treasury rates (CMT) compiled by the US Treasury, and available from the H.15 Fed's statistical release. The maturities comprise one, two, three, five, seven, ten and 20 years. Our sample period is November 1971 through December 2009. The beginning of our sample coincides with the end of the Bretton Woods system in August 1971. This is also when the GSW data for long-term yields become available. Data on 20-year CMT yield are not

Table A.1: Unit root test

Panel A reports values of the ADF test for τ_t^{CPI} and yields with different maturities. τ_t^{CPI} is specified in equation (1.8). In the last column, $\overline{y}_t$ is the average of yields across maturities: $\overline{y}_t = \frac{1}{20}\sum_{i=1}^{20} y_t^{(i)}$. For all variables the test contains a constant since yields and τ_t are both nonzero mean. **Panel B** reports the values of the ADF test for the cointegrating residuals $c_t^{(n)}$ from the regression of $y_t^{(n)}$ on τ_t^{CPI} (the regression includes a constant). We specify the test without a constant since $c_t^{(n)}$ is zero mean by construction. $\overline{c}_t$ in the last column is obtained as the residual from a regression of $\overline{y}_t$ on τ_t^{CPI}. The null hypothesis states that a variable has a unit root. Corresponding critical values are reported separately in each panel.

Panel A. ADF test for τ_t^{CPI} and yields

# lags	τ_t^{CPI}	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(5)}$	$y_t^{(7)}$	$y_t^{(10)}$	$y_t^{(20)}$	$\overline{y}_t$
1	-2.75	-1.90	-1.68	-1.34	-1.09	-0.95	-1.14	-1.09
3	-1.15	-1.42	-1.25	-1.02	-0.90	-0.80	-0.90	-0.84
6	-1.07	-1.21	-1.11	-1.03	-0.95	-0.90	-1.19	-0.94
12	-0.87	-1.63	-1.54	-1.38	-1.27	-1.15	-1.31	-1.22
Critical values: -3.46 (1%), -2.87 (5%), -2.59 (10%)								

Panel B. ADF test for cointegrating residual

# lags	$c_t^{(1)}$	$c_t^{(2)}$	$c_t^{(5)}$	$c_t^{(7)}$	$c_t^{(10)}$	$c_t^{(20)}$	$\overline{c}_t$
1	-4.05	-4.22	-4.44	-4.28	-4.29	-4.30	-4.35
3	-3.38	-3.58	-3.88	-3.93	-3.97	-3.81	-3.84
6	-3.12	-3.39	-3.97	-4.10	-4.27	-4.58	-4.14
12	-4.05	-4.41	-4.97	-5.09	-5.21	-5.10	-5.08
Critical values: -2.58 (1%), -1.96 (5%), -1.63 (10%)							

available for the period from January 1987 through September 1993. We fill this gap by computing the monthly yield returns of the 30-year CMT yield and using them to write the 20-year CMT yield forward. To compute the zero curve, we treat CMT rates as par yields and apply the piecewise cubic Hermite polynomial.

– *Short maturity rate.* The six-month T-bill rate is from the H.15 tables. We use secondary market quotes, and convert them from the discount to the continuously compounded basis.

– *Zero curve.* For comparison, we also use the GSW and Fama-Bliss zero yields. GSW data set is compiled by the Fed. The GSW data are available at `http://www.federalreserve.gov/econresdata/researchdata.htm`. Fama-Bliss data are obtained from the CRSP database.

Macroeconomic variables:

– *Inflation.* CPI for all urban consumers less food and energy (core CPI) is from Bureau of Labor Statistics, downloaded from the FRED database. We define core CPI inflation as the year-on-year simple growth rate in the core CPI index. We construct the cyclical component of inflation CPI_t^c as

the difference between the core CPI inflation and permanent component τ_t^{CPI} computed according to equation (1.8).

– *Unemployment.* UNEMPL is the year-on-year log growth in the unemployment rate provided by the Bureau of Labor Statistics. The series is downloaded from the FRED database.

Financial variables:

– *Commercial paper spread.* Commercial paper spread is defined as the difference between the yield on a three-month commercial paper and the yield on a three-month T-bill.

– *Swap spread.* Swap spread is the difference between ten-year swap rate and the corresponding CMT yield.

– *Moody's Baa spread.* Moody's Baa spread is the difference between the Moody's Baa corporate bond yield and the 30-year CMT yield. To compute the yield, Moody's includes bonds with remaining maturities as close as possible to 30 years.

– *TED spread.* The TED spread is the difference between the three-month LIBOR and the yield on three-month Treasury bill.

– *T-bill3M spread.* T-bill3M spread is the difference between the three-month T-bill and the Fed funds target rate.

– *Fed funds rate.* The Federal funds denotes the monthly effective Fed funds rate. Monthly Fed funds rates are obtained as the average of daily values.

All financial data series are obtained from the FRED database, the only exception are the swap and LIBOR rates which are downloaded from Datastream.

Survey data:

– *Blue Chip Financial Forecasts.* Blue Chip Financial Forecasts (BCFF) survey contains monthly forecasts of yields, inflation and GDP growth given by approximately 45 leading financial institutions. The BCFF is published on the first day of each month, but the survey itself is conducted over a two-day period, usually between the 23rd and 27th of each month. The exception is the survey for the January issue which generally takes place between the 17th and 20th of December. The precise dates as to when the survey was conducted are not published. The BCFF provides forecasts of constant maturity yields across several maturities: three and six months, one, two, five, ten, and 30 years. The forecasts are quarterly averages of interest rates for the current quarter, the next quarter out to five quarters ahead.

– *Livingston survey.* Livingston survey was started in 1946, it covers the forecasts of economists from banks, government and academia. The survey contains semi-annual forecasts of key macro and financial variables such as inflation, industrial production, GDP, unemployment, housing starts,

corporate profits and T-bills. It is conducted in June and December each year. The survey contains forecast out to ten years ahead for some variables. However, the inflation forecasts ten years ahead start only in 1990.

– *Survey of professional forecasters.* Conducted quarterly; respondents provide estimates of the one- and ten-year inflation, among other variables. One-year inflation forecasts start in 1981:Q3, and the ten-year forecasts begin in 1991:Q4.

A.2.A Comparison of excess returns from different data sets

Realized bond excess returns are commonly defined on zero coupon bonds. Since the computation of returns can be sensitive to the interpolation method, we compare returns obtained from CMTs to those from the GSW and FB data. Table A.2 presents the regressions of one-year holding period CMT excess returns on their GSW and FB counterparts with matching maturities. Figure A.1 additionally graphs selected maturities. Excess returns line up very closely across alternative data sets. The R^2's from regressions of CMT excess returns on GSW and FB consistently exceed 99%, except for the ten-year bond for which the R^2 drops to 98% due to one data point in the early part of the sample (1975). Beta coefficients are not economically different from one. We conclude that any factor that aims to explain important features of excess bond returns shall perform similarly well irrespective of the data set used. Therefore, our key results are not driven by the choice of the CMT data.

Table A.2: Comparison of one-year holding period excess returns: CMT, GSW and FB data

The table reports β's and R^2's from regressions of excess returns constructed from CMT data on GSW (**panel A**) and FB (**panel B**) counterparts. We consider a monthly sample 1971:11–2009:12 with maturities from two to ten (five) years for GSW (FB) data. Excess returns are defined over a one-year holding period.

	$rx^{(2)}$	$rx^{(3)}$	$rx^{(4)}$	$rx^{(5)}$	$rx^{(6)}$	$rx^{(7)}$	$rx^{(8)}$	$rx^{(9)}$	$rx^{(10)}$
Panel A. Regressions of rx from CMT on GSW									
β	1.04	1.03	1.04	1.05	1.06	1.05	1.04	1.04	1.04
R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98
Panel B. Regressions of rx from CMT on FB									
β	1.04	1.01	1.02	1.05	—	—	—	—	—
R^2	0.99	0.99	0.99	0.99	—	—	—	—	—

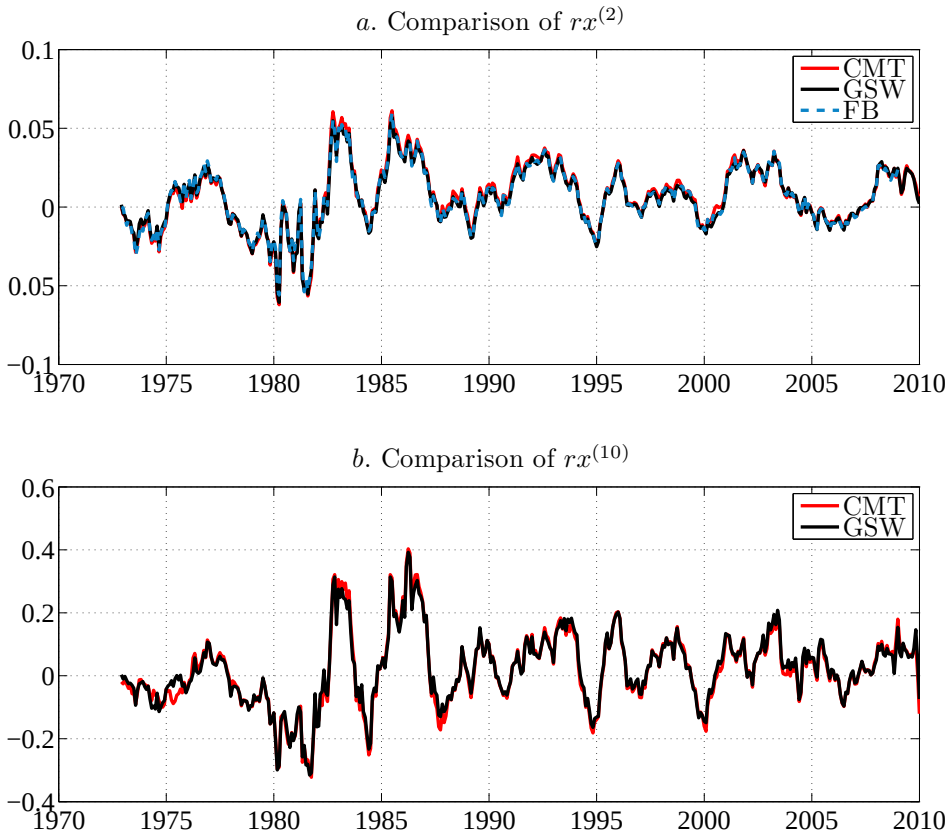


Figure A.1: Comparison of realized excess returns across data sets

The figure plots one-year holding period returns on zero bonds constructed from three data sets: CMT, GSW and FB over the period 1971:11–2009:12. Upper panel provides a comparison for the excess returns on a two-year bond, the bottom panel compares the excess returns on the ten-year bond.

A.3 Basic expression for the long-term yield

It is straightforward to express an n-period yield as the expected sum of future short rates plus the term premium. For completeness, we briefly provide the argument. The price of an n-period nominal bond P_t^n satisfies:

$$P_t^{(n)} = E_t\left(M_{t+1}P_{t+1}^{(n-1)}\right), \tag{A.1}$$

where M_{t+1} is the nominal stochastic discount factor. Let lowercase letters $\left(m_t, p_t^{(n)}\right)$ denote natural logarithms of the corresponding variables. Under conditional joint lognormality of M_{t+1} and the bond

price, from (A.1) we obtain the recursion:

$$p_t^{(n)} = E_t\left(p_{t+1}^{(n-1)} + m_{t+1}\right) + \frac{1}{2}Var_t\left(p_{t+1}^{(n-1)} + m_{t+1}\right),$$

where r_t is the short rate: $r_t = y_t^{(1)}$. By recursive substitution, we can express $p_t^{(n)}$ as:

$$\begin{aligned} p_t^{(n)} = &-E_t\left(r_t + r_{t+1} + ... + r_{t+n-1}\right) + E_t\left[\frac{1}{2}Var_t\left(p_{t+1}^{(n-1)}\right) + Cov_t\left(p_{t+1}^{(n-1)}, m_{t+1}\right)\right. \\ &\left. + \frac{1}{2}Var_{t+1}\left(p_{t+2}^{(n-2)}\right) + Cov_{t+1}\left(p_{t+2}^{(n-2)}, m_{t+2}\right) + ... + \frac{1}{2}Var_{t+n-2}\left(p_{t+n-1}^{(1)}\right) + Cov_{t+n-2}\left(p_{t+n-1}^{(1)}, m_{t+n-1}\right)\right]. \end{aligned}$$

Let $rx_{t+1}^{(n)} = \ln \frac{P_{t+1}^{(n-1)}}{P_t^{(n)}} - r_t$ and $y_t^{(n)} = -\frac{1}{n} p_t^{(n)}$. For an n-maturity yield, since $E_t\left(rx_{t+1}^{(n)}\right) = -Cov_t\left(m_{t+1}, p_{t+1}^{(n-1)}\right) - \frac{1}{2} Var_t\left(p_{t+1}^{(n-1)}\right)$, we obtain:

$$y_t^{(n)} = \frac{1}{n} E_t\left(\sum_{i=0}^{n-1} r_{t+i}\right) + \underbrace{\frac{1}{n} E_t\left(\sum_{i=0}^{n-2} rx_{t+i+1}^{(n-i)}\right)}_{:= rpy_t^{(n)}}. \tag{A.2}$$

A.4 Small sample standard errors

We use the block bootstrap (e.g., Künsch, 1989) to assess the small sample properties of the test statistics and to account for the for the uncertainty about $c_t^{(n)}$. This appendix provides the details of the bootstrap procedure for regressions reported in Table 1.5, which use the single factor to forecast individual bond returns. Small sample inference in other regressions is analogous.

The estimation consists of the following steps:

Step 1. Project yields on the persistent component τ_t^{CPI} to obtain the cycles, $c_t^{(n)}$:

$$y_t^{(n)} = b_0^{(n)} + b_\tau^{(n)} \tau_t^{CPI} + c_t^{(n)}, \quad n = 1, \ldots, m. \tag{A.3}$$

Step 2. Construct the single forecasting factor, $\widehat{cf}_t$ by regressing:

$$\bar{c}_t = \gamma_1 c_t^{(1)} + \overline{\varepsilon}_{t+1}, \tag{A.4}$$

$$\overline{c}_t = \frac{1}{m-1} \sum_{i=2}^{m} c_t \tag{A.5}$$

$$\widehat{cf}_t = \overline{c}_t - \hat{\gamma}_1 c_t^{(1)}. \tag{A.6}$$

Step 3. Forecast individual returns with $\widehat{cf}_t$:

$$rx_{t+1}^{(i)} = \beta_0^{(i)} + \beta_1^{(i)} \widehat{cf}_t + \varepsilon_{t+1}^{(i)}. \tag{A.7}$$

Let Z be a $T \times p$ data matrix with the t-th row: $Z_t = \left(\mathbf{y}_t', \tau_t^{CPI}, rx_{t+1}^{(i)}\right)'$, and $\mathbf{y}_t = \left(y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(m)}\right)'$. We split Z into blocks of size $bs \times p$, where $bs = \sqrt{T}$ ($bs = 21$ for the 1971–2009 sample). Specifically, we create $(T - bs + 1)$ overlapping blocks consisting of observations: $(1, \ldots, bs)$, $(2, \ldots, bs+1)$, $\ldots$, $(T - bs + 1, \ldots, T)$. In each bootstrap iteration, we select T/bs blocks with replacement, out of which we reconstruct the sample in the order the blocks were chosen. We perform steps 1 through 3 on the newly created sample, store the coefficients, t-statistics and adjusted R^2 values. For the statistics of interest, we approximate the empirical distribution using 1000 bootstrap repetitions, and obtain its 5% and 95% percentile values.

A.5 Constructing the single factor

This appendix introduces alternative approaches to constructing the single factor discussed in Section 1.3.C.

A.5.A Exploiting information about future returns

We noted that the baseline construction of the return forecasting factor in equation (1.18) uses only time-t variables, i.e. unlike the CP factor, it does not involve future returns. Now, we ask how the results change when, as an alternative approach, we estimate a regression that does use information about future returns:

$$\overline{rx}_{t+1} = \gamma_0 + \gamma_1 c_t^{(1)} + \gamma_2 \overline{c}_t + \overline{\varepsilon}_{t+1}, \tag{A.8}$$

where $\overline{rx}_{t+1} = \frac{1}{m-1}\sum_{i=2}^{m} rx_{t+1}^{(i)}$ and $\overline{c}_t = \frac{1}{m-1}\sum_{i=2}^{m} c_t^{(i)}$. We form the single forecasting factor as the fitted value from this regression, and label it $\widehat{cf}_t^{TS}$:

$$\widehat{cf}_t^{TS} = \hat{\gamma}_0 + \hat{\gamma}_1 c_t^{(1)} + \hat{\gamma}_2 \overline{c}_t. \tag{A.9}$$

The superscript "TS" shall remind us that we provide forward looking time series information to construct the forecasting factor.

A.5.B One-step NLS estimation

We form a single factor as a linear combination of c_t's:

$$\widehat{cf}_t^{NLS} = \lambda' \mathbf{c}_t, \tag{A.10}$$

and estimate the restricted system:

$$\mathbf{rx}_{t+1} = A \begin{pmatrix} 1 \\ \lambda' \mathbf{c}_t \end{pmatrix} + \varepsilon_{t+1}, \tag{A.11}$$

where $\mathbf{rx}_{t+1}$ is a $(m-1) \times 1$ vector of individual returns with maturities from two to m years, $\mathbf{rx}_{t+1} = \left(rx_{t+1}^{(2)}, rx_{t+1}^{(3)}, ..., rx_{t+1}^{(m)}\right)'$, $\mathbf{c}_t$ is a vector of cycles, and A is a matrix parameters:

$$A = \begin{pmatrix} \alpha_0^{(2)} & \alpha_1^{(2)} \\ \alpha_0^{(3)} & \alpha_1^{(3)} \\ \vdots & \vdots \\ \alpha_0^{(m)} & \alpha_1^{(m)} \end{pmatrix}. \tag{A.12}$$

We perform non-linear least squares (NLS) estimation, by minimizing the sum of squared errors:

$$(\hat{A}, \hat{\lambda}) = \min_{A,\lambda} \sum_{t=1}^{T} \left(\mathbf{rx}_{t+1} - A \begin{pmatrix} 1 \\ \lambda' \mathbf{c}_t \end{pmatrix} \right)' \left(\mathbf{rx}_{t+1} - A \begin{pmatrix} 1 \\ \lambda' \mathbf{c}_t \end{pmatrix} \right). \tag{A.13}$$

For identification, we set $\alpha_1^{(7)} = 1$. This choice is without loss of generality. The loss function (A.13) is minimized iteratively until its values are not changing between subsequent iterations. In application, being interested in the dynamics of the single factor $\widehat{cf}_t^{NLS}$, we additionally standardize excess returns cycles prior to estimation.

A.5.C Common factor by eigenvalue decomposition

Alternatively, in constructing the single factor we can exploit the regression (1.16) of an individual excess return on $c_t^{(1)}$ and the cycle of the corresponding maturity:

$$rx_{t+1}^{(n)} = \alpha_0^{(n)} + \alpha_1^{(n)} c_t^{(1)} + \alpha_2^{(n)} c_t^{(n)} + \varepsilon_{t+1}^{(n)}. \tag{A.14}$$

We form a vector $\mathbf{erx}_t$ of expected excess returns obtained from this model:

$$\mathbf{erx}_t = E_t \left(rx_{t+1}^{(2)}, rx_{t+1}^{(3)}, ..., rx_{t+1}^{(m)} \right)'. \tag{A.15}$$

The single factor is obtained as the first principal component of the covariance matrix of $\mathbf{erx}_t$:

$$\widehat{cf}_t^{PC} = U'_{(:,1)} \mathbf{erx}_t, \tag{A.16}$$

where $Cov\left(\mathbf{erx}_t\right) = ULU'$, and $U_{(:,1)}$ denotes the eigenvector associated with the largest eigenvalue in L. Using returns from two to 20 years, the first principal component explains 94% of common variation in $\mathbf{erx}_t$.

A.5.D Comparing the results

We compare the single factor obtained with the different procedures. To distinguish between approaches, we use the notation: $\widehat{cf}_t^{TS}$ for the construction involving future returns, $\widehat{cf}_t^{NLS}$ for the one-step NLS estimation, $\widehat{cf}_t^{PC}$ for the factor obtained with the eigenvalue decomposition of expected returns, and $\widehat{cf}_t$ for the simple approach introduced in the body of the paper in Section 1.3.C.

First, panel A of Table A.3 presents the correlations among the four measures. Clearly, while the methods differ, they all identify virtually the same dynamics of the single factor. The correlation between the constructed factors reaches 98% or more.

Second, the way we obtain the single factor is inconsequential for the predictability we report. As a summary, panel B of Table A.3 displays the adjusted R^2 values obtained by regressing individual excess returns on $\widehat{cf}_t^{NLS}$, $\widehat{cf}_t^{PC}$ and $\widehat{cf}_t$, respectively. The difference between the three measures is negligible.

Table A.3: Comparing alternative constructions of the single factor

The table reports correlations between alternative approaches to constructing the single forecasting factor (**panel A**), as well as $\bar{R}^2$ values for predictability of individual excess returns (**panel B**). $\widehat{cf}_t$ is used in the body of the paper, and defined in equation (1.19); $\widehat{cf}_t^{TS}$ differs from the baseline specification in that it involves information about future returns as discussed in Section A.5.A; $\widehat{cf}_t^{PC}$ is obtained from the eigenvalue decomposition of expected excess returns in Section A.5.C of this Appendix; $\widehat{cf}_t^{NLS}$ is obtained in a one-step estimation in Section A.5.B.

Panel A. Correlations

	$\widehat{cf}_t$	$\widehat{cf}_t^{TS}$	$\widehat{cf}_t^{PC}$	$\widehat{cf}_t^{NLS}$
$\widehat{cf}_t$	1	0.999	0.998	0.979
$\widehat{cf}_t^{TS}$	·	1	0.999	0.979
$\widehat{cf}_t^{PC}$	·	·	1	0.986
$\widehat{cf}_t^{NLS}$	·	·	·	1

Panel B. $\bar{R}^2$ from predictive regressions

	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
$\widehat{cf}_t$	0.38	0.46	0.50	0.53	0.55	0.52
$\widehat{cf}_t^{TS}$	0.38	0.46	0.50	0.53	0.55	0.52
$\widehat{cf}_t^{PC}$	0.39	0.47	0.51	0.54	0.55	0.52
$\widehat{cf}_t^{NLS}$	0.40	0.48	0.52	0.55	0.56	0.53

A.6 Predictability of bond excess returns at different horizons

In the body of the paper, we constrain our analysis to bond excess returns for the one-year holding period. In this appendix, we summarize the results of predictive regressions for bond excess returns at shorter horizons (h): one, three, six and nine months. Table A.4 reports the results. The construction of $\widehat{cf}_t$ is described in Section 1.3.C. $\widehat{cf}_t$ is highly significant across all horizons and the R^2 increases with the investment horizon. The results suggest that the single factor is a robust predictor across horizons.

A.7 Long-run inflation expectations: the persistent component

Our τ_t^{CPI} variable can be interpreted as an endpoint of inflation expectations, i.e. the local long-run mean to which current inflation expectations converge. In this appendix, we show how τ_t^{CPI} can be embedded

Table A.4: Predictability of bond excess returns across horizons

The table reports the results from predictive regression for bond excess returns at different investment horizons, $rx_{t+h/12}, h = 1, 3, 6, 9$ months. The single factor $\widehat{cf}_t$ is constructed from the yield cycles using the τ_t^{CPI} as a proxy for the persistent component of yields. In parentheses, t-statistics use the Newey-West adjustment with 15 lags. All variables are standardized.

	$rx^{(2)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$	$rx^{(15)}$	$rx^{(20)}$
			a. $h = 1$ month			
$\widehat{cf}_t$	0.00	0.01	0.01	0.01	0.01	0.02
	(3.11)	(4.12)	(4.77)	(5.02)	(5.08)	(4.77)
$\bar{R}^2$	0.02	0.03	0.04	0.04	0.04	0.04
			b. $h = 3$ months			
$\widehat{cf}_t$	0.01	0.02	0.02	0.03	0.04	0.06
	(4.44)	(5.41)	(5.94)	(6.10)	(6.14)	(5.85)
$\bar{R}^2$	0.07	0.10	0.12	0.14	0.14	0.13
			c. $h = 6$ months			
$\widehat{cf}_t$	0.01	0.03	0.04	0.06	0.08	0.11
	(5.60)	(6.68)	(7.22)	(7.54)	(7.77)	(7.45)
$\bar{R}^2$	0.17	0.23	0.25	0.27	0.28	0.26
			d. $h = 9$ months			
$\widehat{cf}_t$	0.01	0.04	0.06	0.09	0.13	0.17
	(6.85)	(8.14)	(8.56)	(8.97)	(8.88)	(7.99)
$\bar{R}^2$	0.29	0.35	0.39	0.42	0.42	0.39

within a simple model of the term structure of inflation expectations. To obtain the gain parameter v that is consistent with inflation forecasts we estimate the model using survey data for CPI.

A.7.A Model of the term structure of inflation expectations

The model of the term structure of inflation expectations follows Kozicki and Tinsley (2006). Let the realized inflation CPI_t follow an AR(p) process, which we can write in a companion form as:

$$CPI_{t+1} = e_1' z_{t+1} \tag{A.17}$$

$$z_{t+1} = Cz_t + (I - C)\mathbf{1}\mu_\infty^{(t)} + e_1\varepsilon_{t+1}, \tag{A.18}$$

where $z_t = (CPI_t, CPI_{t-1}, ..., CPI_{t-p+1})'$, $e_1 = (1, 0, ..., 0)'$ with dimension $(p \times 1)$, $\mathbf{1}$ is a $(p \times 1)$ vector of ones, and companion matrix C is of the form

$$C = \begin{pmatrix} c_1 & c_2 & \dots & c_{p-1} & c_p \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ & & \ddots & & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}. \tag{A.19}$$

Then,

$$CPI_{t+1} = e_1' C z_t + e_1'(I - C)\mathbf{1}\mu_\infty^{(t)} + e_1' e_1 \varepsilon_{t+1}. \tag{A.20}$$

In the above specification, inflation converges to a time varying, rather than constant, long-run mean:

$$\mu_\infty^{(t)} = \lim_{k\to\infty} E_t\left(CPI_{t+k}\right), \tag{A.21}$$

which itself follows a random walk:

$$\mu_\infty^{(t+1)} = \mu_\infty^{(t)} + v_{t+1}. \tag{A.22}$$

The expected inflation j-months ahead is given as:

$$E_t\left(CPI_{t+j}\right) = e_1' C^j z_t + e_1'(I - C^j)\mathbf{1}\mu_\infty^{(t)}. \tag{A.23}$$

Thus, survey expectations can be expressed as:

$$s_{t,k} = \frac{1}{k}\sum_{j=1}^{k} E_t^s\left(CPI_{t+j}\right), \tag{A.24}$$

where the survey is specified as the average inflation over k periods, and E^s denotes the survey expectations. We treat the survey data as expected inflation plus a normally distributed measurement noise:

$$s_{t,k} = \frac{1}{k}\sum_{j=1}^{k} E_t\left(CPI_{t+j}\right) + \eta_{t,k}. \tag{A.25}$$

It is convenient to cast the model in a filtering framework with the state equation given as:

$$\mu_\infty^{(t+1)} = \mu_\infty^{(t)} + v_{t+1}, \quad v_{t+1} \sim N(0, Q), \tag{A.26}$$

and the measurement equation:

$$m_t = A z_t + H\mu_\infty^{(t)} + w_t, \qquad w_t \sim N(0, R) \tag{A.27}$$

where $m_t = (CPI_{t+1}, s_{t,k_1}, s_{t,k_2}, \ldots, s_{t,k_n})$ and $w_t = (\varepsilon_{t+1}, \eta_{t,k_1}, \eta_{t,k_2}, \ldots \eta_{t,k_n})'$, where k_i is the forecast horizon of a given survey i. We assume that the covariance matrix R is diagonal, and involves only two distinct parameters: (i) the variance of the realized inflation shock, and (ii) the variance of the measurement error for $s_{k,t}$, which is assumed identical across different surveys. From equations (A.23) and (A.24), A and H matrices in (A.27) have the form:

$$A = \begin{pmatrix} e_1' C \\ e_1' \frac{1}{k_1} \sum_{j=1}^{k_1} C^j \\ e_1' \frac{1}{k_2} \sum_{j=1}^{k_2} C^j \\ \cdots \\ e_1' \frac{1}{k_n} \sum_{j=1}^{k_n} C^j \end{pmatrix}, \quad H = \begin{pmatrix} e_1' (I - C) \mathbf{1} \\ e_1' \left(I - \frac{1}{k_1} \sum_{j=1}^{k_1} C^j \right) \mathbf{1} \\ e_1' \left(I - \frac{1}{k_2} \sum_{j=1}^{k_2} C^j \right) \mathbf{1} \\ \cdots \\ e_1' \left(I - \frac{1}{k_n} \sum_{j=1}^{k_n} C^j \right) \mathbf{1} \end{pmatrix}. \tag{A.28}$$

We consider two versions of the endpoint process, $\mu_\infty^{(t)}$. In the first version, we treat $\mu_\infty^{(t)}$ as a random walk as in equation (A.22). We estimate the model by maximum likelihood combined with the standard Kalman filtering of the latent state, $\mu_\infty^{(t)}$.

In the second version, we obtain the endpoint as the discounted moving average of past inflation, as we do in the body of the paper, i.e.

$$\mu_\infty^{(t)} := \tau_t^{CPI}(v, N). \tag{A.29}$$

In the expression above, we explicitly stress the dependence of τ_t^{CPI} on the parameters. This case allows us to infer the gain parameter v that is consistent with the available survey data. We estimate the model with maximum likelihood. Since v and N are not separately identified (see also Figure A.3 below), we fix the window size at $N = 120$ months, and estimate the v parameter for this window size. In the subsequent section, we provide an extensive sensitivity analysis of the predictive results for bond returns to both v and N parameters.

A.7.B Data

We combine two inflation surveys compiled by the Philadelphia Fed that provide a long history of data and cover different forecast horizons:

- Livingston survey: Conducted bi-annually in June and December; respondents provide forecasts of the CPI level six and twelve months ahead. Following Kozicki and Tinsley (2006) and Carlson (1977) we convert the surveys into eight- and 14-month forecasts to account for the real time information set of investors. We use data starting from 1955:06.

- Survey of Professional Forecasters: Conducted quarterly; respondents provide estimates of the one- and ten-year inflation. One-year forecasts start in 1981:Q3, and the ten-year forecasts begin in 1991:Q4.

We use the median survey response. We match the data with the realized CPI (all items) because it underlies the surveys. The estimation covers the period from 1957:12–2010:12. In the adaptive learning version of the model, we use data from 1948:01 to obtain the first estimate of τ_t^{CPI}.

A.7.C Estimation results

We estimate the model assuming an AR(12) structure for inflation. In the adaptive learning version of the model, we estimate the gain parameter at $v = 0.9868$. The BHHH standard error of 0.0025 suggest that v is highly significant. Its value implies that when forming their long-run inflation expectations, each month agents attach the weight of about 1.3% to the current inflation. Other parameters are not reported for brevity. Panel a of Figure A.2 displays the realized inflation and the estimates of its long-run expectations considering the two specifications. The series labeled as "random walk" is the filtered $\mu_\infty^{(t)}$ state. The series marked as "adaptive" shows discounted moving average of inflation τ_t^{CPI} constructed at the estimated parameter $v = 0.9686$, and assuming $N = 120$ months. Panel b of Figure A.2 plots the survey data used in estimation. Comparing the filtered series in panel a, we note that while both estimates trace each other closely, the random walk specification points to a faster downward adjustment in inflation expectation during the disinflation period compared to the adaptive learning proxy. This finding is intuitive in that in the first part of the sample until early 1990s, the long-horizon survey information is unavailable. Thus, the filtered inflation endpoint $\mu_\infty^{(t)}$ is tilted towards the realized inflation, and short horizon survey forecasts.

A.7.D Sensitivity of predictive results to τ_t^{CPI}

We analyze the sensitivity of our predictive results towards the specification of the persistent component τ_t^{CPI}. One concern is that these results could be highly dependent on the weighting scheme (v parameter) or the length of the moving average window (N) used to construct τ_t. For this reason, in Figure A.3 we plot in-sample and out-of-sample R^2's varying v between 0.975 and 0.995 and N between 100 and 150 months. While the predictability is stable across a wide range of parameter combinations, it weakens for values of v approaching one and for long window sizes. The deterioration in this region of the parameter space is intuitive: When combined with a long moving window, v close to one oversmooths the CPI data and leads to a less local estimate of the persistent component.[1]

In Figure A.4, we use a simple moving average of past core CPI to isolate how the in- and out-of-sample predictive results depend on the window size, N. We consider N between 10 and 150 months. The predictive results are relatively stable for windows between 40 and 100 months, and taper off at the extremes. A very short moving window tilts τ_t^{CPI} to current realized inflation, a very long window, in turn, oversmooths the data. Both cases provide a poor measurement of the current long-run inflation mean, thus the predictability of bond returns weakens.

Beside the core CPI, Figure A.4 considers two alternative variables used in the literature to construct proxies of the local mean reversion in interest rates: (i) the effective fed funds rate, and (ii) the one-year yield. For usual and economically plausible window sizes, neither alternative delivers predictability of bond returns at the level documented with the CPI. The question that underlies the difference in predictability is how each variable captures the persistent movement in interest rates. Using the moving average of the

[1] As $v \to 1$, the discounted moving average converges to a simple moving average with the corresponding window.

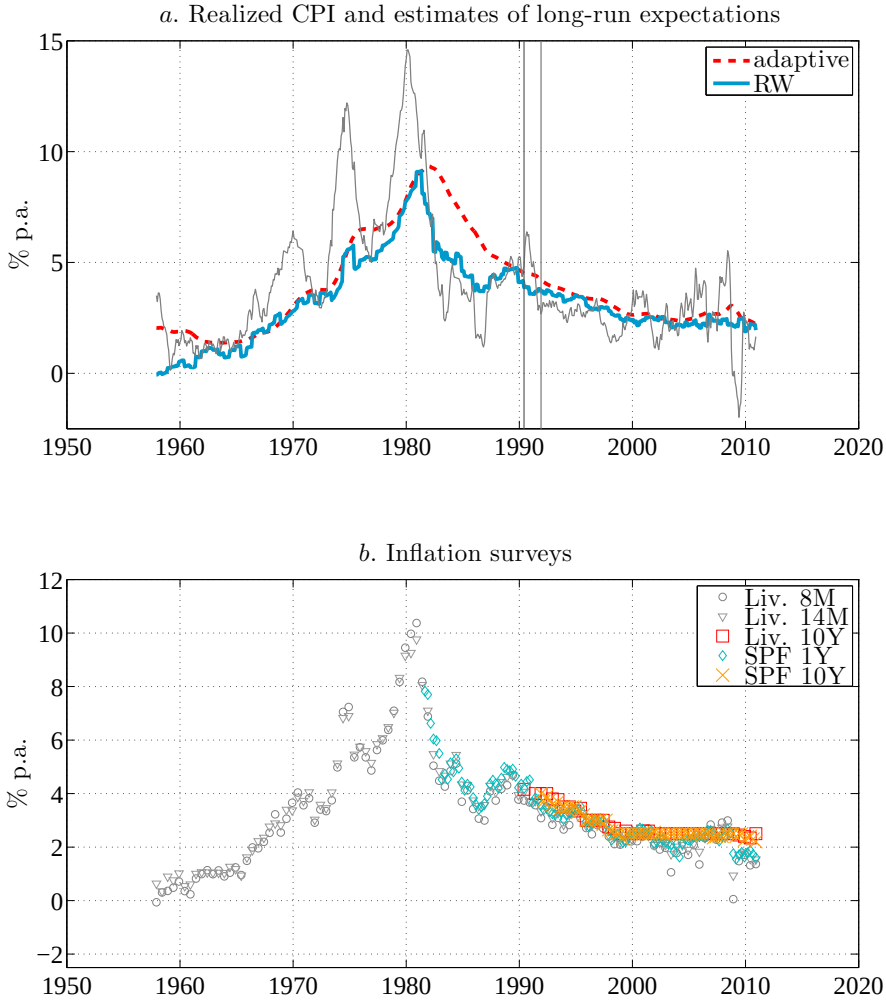


Figure A.2: Long-horizon inflation expectations

Panel *a* shows the realized CPI and the estimated long-run inflation expectations specified as a random walk and discounted moving average of past CPI data (adaptive). Vertical lines mark the dates on which 10-year inflation forecasts from Livingston and SPF surveys become available, respectively. Panel *b* shows the CPI surveys used in estimation.

short rate one is faced with a tradeoff between smoothing the business cycle frequency in the short rate and contemporaneously measuring the generational inflation factor. Apart from the statistical fit, the advantage of τ_t^{CPI} lies in its direct link to an economic quantity, rather than to bond prices themselves. The benefit of an economic interpretation is also revealed in the estimates of a simple Taylor rule that we entertain in

the Introduction to this paper. Indeed, considering $r_t = \gamma_0 + \gamma_c CPI_t^c + \gamma_y UNEMPL_t + \gamma_\tau \tau_t^i + \varepsilon_t$ with different proxies for $\tau_t, i = \{CPI, FFR\}$ shows that it is the CPI that provides highly stable coefficients across different subsamples (not reported).

A.8 Predictability within a macro-finance model

This appendix shows that our decomposition of the yield curve can be easily embedded within a macro-finance model. The model corroborates many of the results we have presented in the body of the paper. It turns out that τ_t is not only important for uncovering the predictability of bond returns but also helps to understand the monetary policy. We provide details on the modified Taylor rule used in the Introduction, and integrate it into a dynamic term structure model. This Taylor rule fills with economic variables the equation (1.3) that we have used to convey the intuition for our decomposition.

A.8.A Incorporating τ_t into a Taylor rule

We specify a Taylor rule in terms of inflation described by two components CPI_t^c and τ_t^{CPI}, unemployment $UNEMPL_t$, and a monetary policy shock f_t:

$$r_t = \gamma_0 + \gamma_c CPI_t^c + \gamma_y UNEMPL_t + \gamma_\tau \tau_t^{CPI} + f_t. \tag{A.30}$$

Below, we discuss the choice of these variables.

Our key assumption concerns how market participants process inflation data. Specifically, investors and the Fed alike perceive separate roles for two components of realized inflation:

$$CPI_t = \mathcal{T}_t + CPI_t^c, \tag{A.31}$$

where $\mathcal{T}_t$ is the long-run mean of inflation, and CPI_t^c denotes its cyclical variation. We approximate $\mathcal{T}_t$ using equation (1.8), denoted τ_t^{CPI}, and obtain CPI_t^c simply as a difference between CPI_t and τ_t^{CPI}.

The decomposition (A.31) is economically motivated and can be mapped to existing statistical models such as the shifting-endpoint autoregressive model of Kozicki and Tinsley (2001a). The decomposition has also an intuitive appeal: One can think of transient inflation CPI_t^c as controlled by the monetary policy actions. In contrast, representing market's conditional long-run inflation forecast, τ_t^{CPI}, is largely determined by the central bank's credibility and investors' perceptions of the inflation target. Monetary policy makers react not only to the higher-frequency swings in inflation and unemployment but also watch the long-run means of persistent macro variables.[2] Therefore, we let τ_t^{CPI} enter the short rate independently from CPI_t^c. Indeed, τ_t^{CPI} is what connects the monetary policy and long term interest rates.

[2]This fact is revealed by the FOMC transcripts, in which both surveys and the contemporaneous behavior of long-term yields provide important gauge of long-horizon expectations.

Taylor rules are usually specified without the distinction between the two components in (A.31), thus precluding that different coefficients may apply to the long-run and transient inflation shocks. We empirically show that removing this restriction helps explain the monetary policy in the last four decades, and improves the statistical fit of a macro-finance model. The situation after the rapid disinflation in the 1980s demonstrates the relevance of this point. Core CPI inflation fell from about 14% in 1980 to less than 4% in 1983 and has remained low since then. However, the steep decline in inflation was not followed by a similar drop in the short rate as the traditional Taylor rule would suggest. Rather, the short rate followed a slow decline in line with the persistent component of inflation.

In that employment is one of the explicit monetary policy objectives and given the difficulties in measuring the output gap in real time, in equation (A.30) we include the unemployment rate as a key real indicator. Mankiw (2001) emphasizes two reasons why the Fed may want to respond to unemployment: (i) its stability may be a goal in itself, (ii) it is a leading indicator for future inflation.[3]

Finally, to complete the Taylor rule, we add a latent monetary policy shock denoted by f_t which summarizes other factors (e.g. financial conditions) that can influence the monetary policy.[4]

As a preliminary check for the specification (A.30), we run an OLS regression of the Fed funds rate on $(CPI_t^c,\, UNEMPL_t, \tau_t^{CPI})$ for three samples (i) including the Volcker episode (1971–2009), (ii) the period after disinflation (1985–2009), (iii) the period before disinflation (1971–1984). In the introductory example, Table 1.1 reports the results and Figure 1.1 plots the fit.

To appreciate the importance of disentangling two inflation components, panels A and B in Table 1.1 juxtapose equation (A.30) with the restricted rule using CPI_t as a measure of inflation, i.e.:

$$r_t = \gamma_0 + \gamma_\pi(CPI_t^c + \tau_t^{CPI}) + \gamma_y UNEMPL_t + \varepsilon_t. \tag{A.32}$$

The unrestricted Taylor rule (A.30) explains 79%, 91%, and 61% of variation in the short rate in the two samples, respectively. This fit is remarkably high given that it uses only macroeconomic quantities. The restricted Taylor rule (A.32) gives lower R^2's of 56%, 76%, and 30%, respectively. We can quantify the effect of the restriction by looking at the difference between τ_t^{CPI} and CPI_t^c coefficients. We note that the coefficient on τ_t^{CPI} is higher than the one on the transitory component of inflation CPI_t^c. Also, the estimated coefficients in the unrestricted rule are more stable across the two periods. Finally, the restricted version underestimates the role of unemployment in determining the monetary policy actions.

[3] Mankiw (2001) proposes a simple formula for setting the Fed funds rate: Fed funds $= 8.5 + 1.4\,(\text{core inflation} - \text{unemployment})$.

[4] Hatzius, Hooper, Mishkin, Schoenholtz, and Watson (2010) offer a thorough discussion of financial conditions, and their link to growth and monetary policy.

A.8.B Model setup

All state variables discussed above enter the short rate expectations in the basic yield equation (1.3). To capture the variation in term premia, we introduce one additional state variable, s_t. We collect all factors in the state vector $\mathcal{M}_t = \left(CPI_t^c, UNEMPL_t, f_t, s_t, \tau_t^{CPI}\right)'$ that follows a VAR(1) dynamics:

$$\mathcal{M}_{t+\Delta t} = \mu_M + \Phi_M \mathcal{M}_t + S_M \varepsilon_{t+\Delta t}, \quad \varepsilon_t \sim N(0, I_5), \quad \Delta t = \frac{1}{12}. \tag{A.33}$$

A.8.C Model estimation

We estimate the model on the sample 1971–2009, considering zero coupon yields with maturities six months, one, two, three, five, seven and ten years at monthly frequency. The zero coupon yields are bootstrapped from the CMT data. Details on the construction of zero curve are provided in Appendix A.2.

We estimate the model by the standard Kalman filter, by providing measurements for yields and for three macro factors appearing in the short rate: cyclical core CPI for CPI_t^c, unemployment rate for $UNEMPL_t$ and discounted moving average of core CPI defined in equation (1.8) for τ_t^{CPI}. We assume identical variance of the measurement error for yield measurements, and different variance of measurement error for each of the macro measurements.

Due to the presence of latent factors, parameters μ_M, Φ_M, S_M are not identified. Therefore, we impose both the economic and identification restrictions as follows:

$$\Phi_M = \begin{pmatrix} \phi_{\pi\pi} & \phi_{\pi y} & 0 & 0 & 0 \\ \phi_{y\pi} & \phi_{yy} & 0 & 0 & 0 \\ 0 & 0 & \phi_{ff} & 0 & 0 \\ 0 & 0 & 0 & \phi_{ss} & 0 \\ 0 & 0 & 0 & 0 & \phi_{\mu\mu} \end{pmatrix}, \quad \mu_M = \begin{pmatrix} 0 \\ \mu_y \\ 0 \\ 0 \\ \mu_\pi \end{pmatrix}, \quad S_M = \begin{pmatrix} \sigma_{\pi\pi} & 0 & 0 & 0 & 0 \\ 0 & \sigma_{yy} & 0 & 0 & 0 \\ 0 & 0 & \sigma_{ff} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\mu\mu} \end{pmatrix}.$$

The market prices of risk have the usual affine form $\lambda_t^M = \Lambda_0^M + \Lambda_1^M \mathcal{M}_t$, with restricted Λ_0^M and Λ_1^M:

$$\Lambda_0^M = \begin{pmatrix} \lambda_{0,\pi} \\ \lambda_{0,y} \\ \lambda_{0,f} \\ 0 \\ 0 \end{pmatrix}, \quad \Lambda_1^M = \begin{pmatrix} 0 & 0 & \lambda_{f\pi} & \lambda_{s\pi} & 0 \\ 0 & 0 & \lambda_{fy} & \lambda_{sy} & 0 \\ 0 & 0 & \lambda_{ff} & \lambda_{sf} & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Under these restrictions, factors s_t and f_t drive the variation in bond premia over time. In this way, we allow the model to reveal the premia structure that is analogous to the construction of $\widehat{cf}_t$. Bond pricing equation have the well-known affine form, therefore we omit the details.

Figure A.5 plots filtered factors. The dynamics of CPI_t^c, $UNEMPL_t$ and τ_t^{CPI} closely follow the observable quantities. Notably, latent factor s_t has stationary and cyclical dynamics similar to the cycles $c_t^{(n)}$ (s_t has a half-life of approximately one year). Despite having only two latent factors, the model is able to fit yields reasonably well across maturities. We summarize this fit in Figure A.6, and for brevity do not report the parameter estimates.

A.8.D Predictability of bond excess returns with filtered states

Our estimation does not exploit any extra information about factors in expected returns. Therefore, predictive regressions on filtered factors provide an additional test on the degree of predictability present in the yield curve. We run two regressions of realized excess return on the filtered states:

$$rx_{t+1}^{(n)} = b_0 + b_1 s_t + \varepsilon_{t+1}^{(n)} \tag{A.34}$$

$$rx_{t+1}^{(n)} = b_0 + b_1 s_t + b_2 f_t + \varepsilon_{t+1}^{(n)}. \tag{A.35}$$

Factor f_t is by construction related to the short-maturity yield, while s_t is designed to capture the cyclical variation at the longer end of the curve. In this context, f_t corresponds to the cycle $c_t^{(1)}$, and s_t aggregates the information from the cycles with longer maturity, $c_t^{(n)}, n \geq 2$. Building on the intuition of $\widehat{cf}_t$, one would expect that f_t can improve the predictability by removing the transient short rate expectations part from s_t.

Regression results confirm that a large part of predictability in bond premia is carried by a single factor. s_t explains up to 37% of the variation in future bond excess returns, and the R^2 increases with maturity (panel A Table A.5). The loadings are determined up to a rotation of the latent factor. The monetary shock f_t is virtually unrelated to future returns, giving zero R^2's (panel B). However, the presence of both factors in regression (A.35) significantly increases the R^2 (panel C). The largest increase in R^2 occurs at the short maturities where the monetary policy plays an important role. These results confirm our intuition for the role of f_t in predictive regressions: it eliminates the expectations part from s_t. The level of predictability achieved by s_t and f_t is close to that of the single predictor $\widehat{cf}_t$ reported in Table 1.5 (Panel A.II.).

Results from this simple macro-finance model lend support to our yield curve decomposition, and more generally to the interpretation of bond return predictability we propose. The form of the Taylor rule turns out particularly important for the distinction between the short rate expectations and term premium component in yields.

Table A.5: Bond premia predictability by filtered states s_t and f_t from the macro-finance model

Panel A of the table reports the results for predictive regressions of bond excess returns on the term premia factor s_t. **Panel B** reports the results for predictive regressions of bond excess returns on monetary policy shock f_t. **Panel C** reports the results for predictive regressions of bond excess returns on f_t and s_t. Factors s_t and f_t are filtered from the no-arbitrage macro-finance model given by (A.30)–(A.33). The sample period is 1971–2009. In parentheses, t-statistics use the Newey-West adjustment with 15 lags. All variables are standardized.

Panel A. $rx_{t+1}^{(n)} = b_0 + b_1 s_t + \varepsilon_{t+1}^{(n)}$					
	$rx^{(2)}$	$rx^{(3)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$
s_t	-0.50	-0.50	-0.55	-0.59	-0.62
	(-4.74)	(-4.64)	(-5.37)	(-5.79)	(-6.18)
R^2	0.24	0.25	0.31	0.35	0.38
Panel B. $rx_{t+1}^{(n)} = b_0 + b_1 f_t + \varepsilon_{t+1}^{(n)}$					
	$rx^{(2)}$	$rx^{(3)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$
f_t	0.05	0.04	-0.01	-0.04	-0.06
	(0.42)	(0.34)	(-0.08)	(-0.33)	(-0.53)
R^2	0.00	0.00	0.00	0.00	0.00
Panel C. $rx_{t+1}^{(n)} = b_0 + b_1 f_t + b_2 s_t + \varepsilon_{t+1}^{(n)}$					
	$rx^{(2)}$	$rx^{(3)}$	$rx^{(5)}$	$rx^{(7)}$	$rx^{(10)}$
f_t	0.48	0.47	0.44	0.43	0.42
	(5.15)	(5.16)	(5.05)	(5.08)	(4.83)
s_t	-0.77	-0.77	-0.80	-0.84	-0.86
	(-6.83)	(-6.77)	(-7.38)	(-7.76)	(-7.84)
$\bar{R}^2$	0.40	0.40	0.44	0.48	0.50

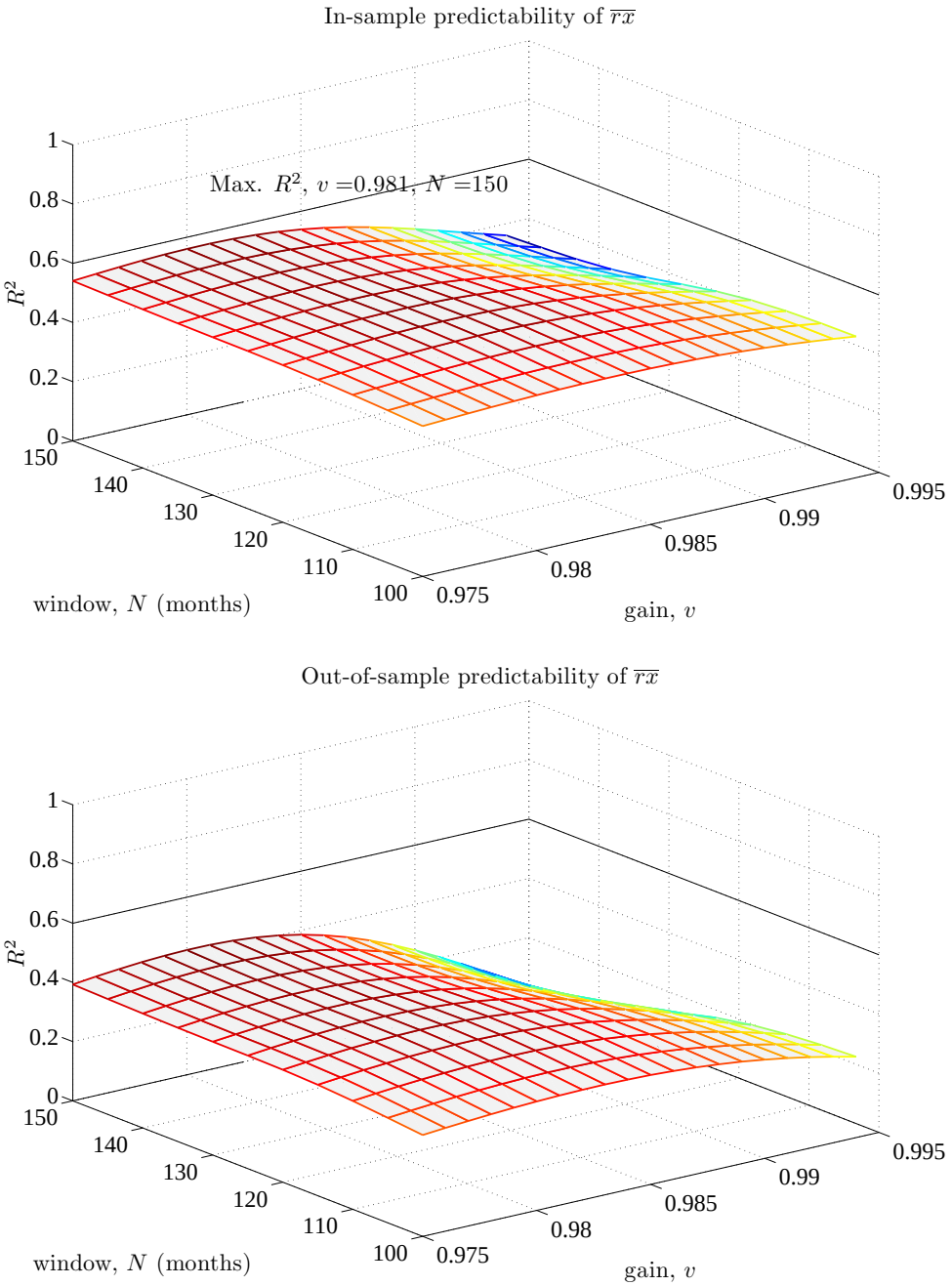


Figure A.3: Sensitivity of the predictability evidence to v and N parameters

The figure studies the sensitivities of predictive $\bar{R}^2$'s to the values of N and v parameters used to construct τ_t^{CPI}. We predict the average bond return across maturities by regressing it on $c_t^{(1)}$ and $\bar{c}_t$ as in the body of the paper. We consider the gain parameter v between 0.975 and 0.995, and the window size between 100 and 150 months. Panel a and b correspond to in-sample and out-of-sample results, respectively.

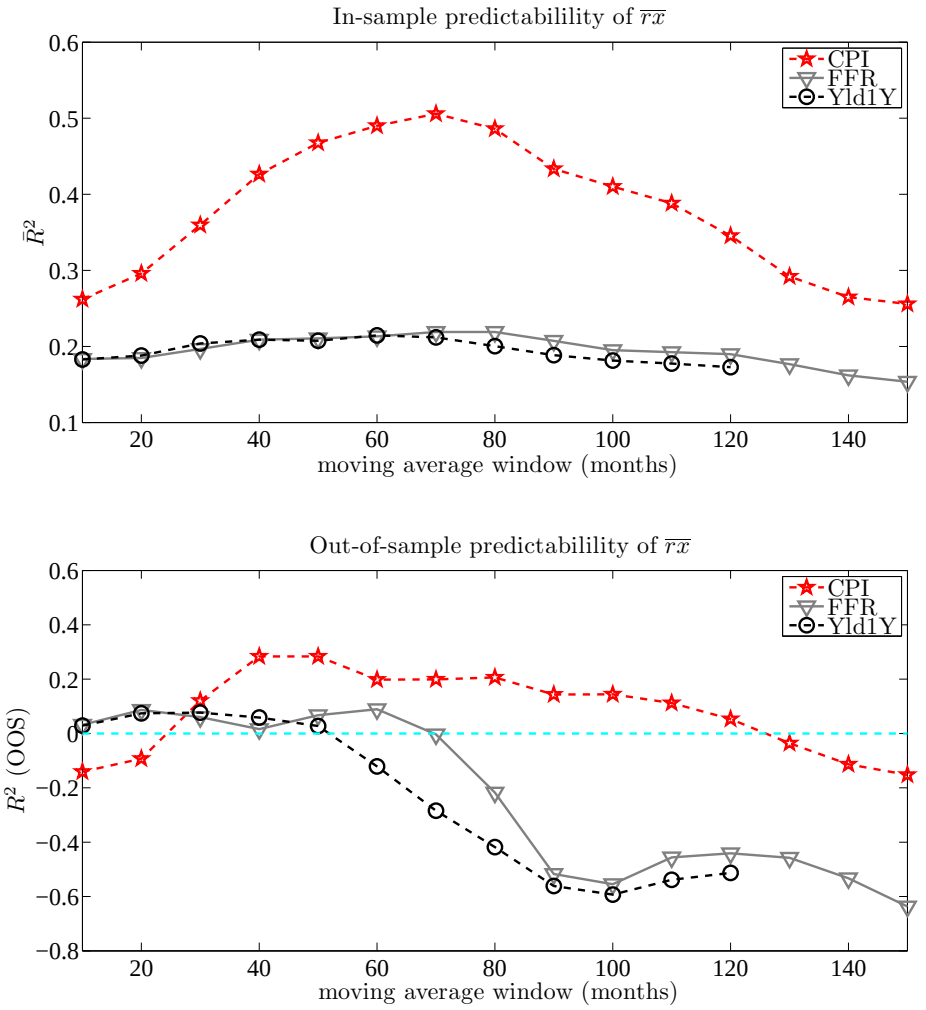


Figure A.4: Sensitivity of predictability evidence to the window size

The figure depicts the sensitivity of the predictive results, in- and out-of-sample, to the length of the moving average window. All results are based on a simple (i.e. undiscounted) moving average. The window varies from ten to 150 months. We consider predictability of $\overline{rx}$, and use three different proxies for the persistent component using moving average of: (*i*) past core CPI, (*ii*) past fed funds rate and (*iii*) past one-year rate.

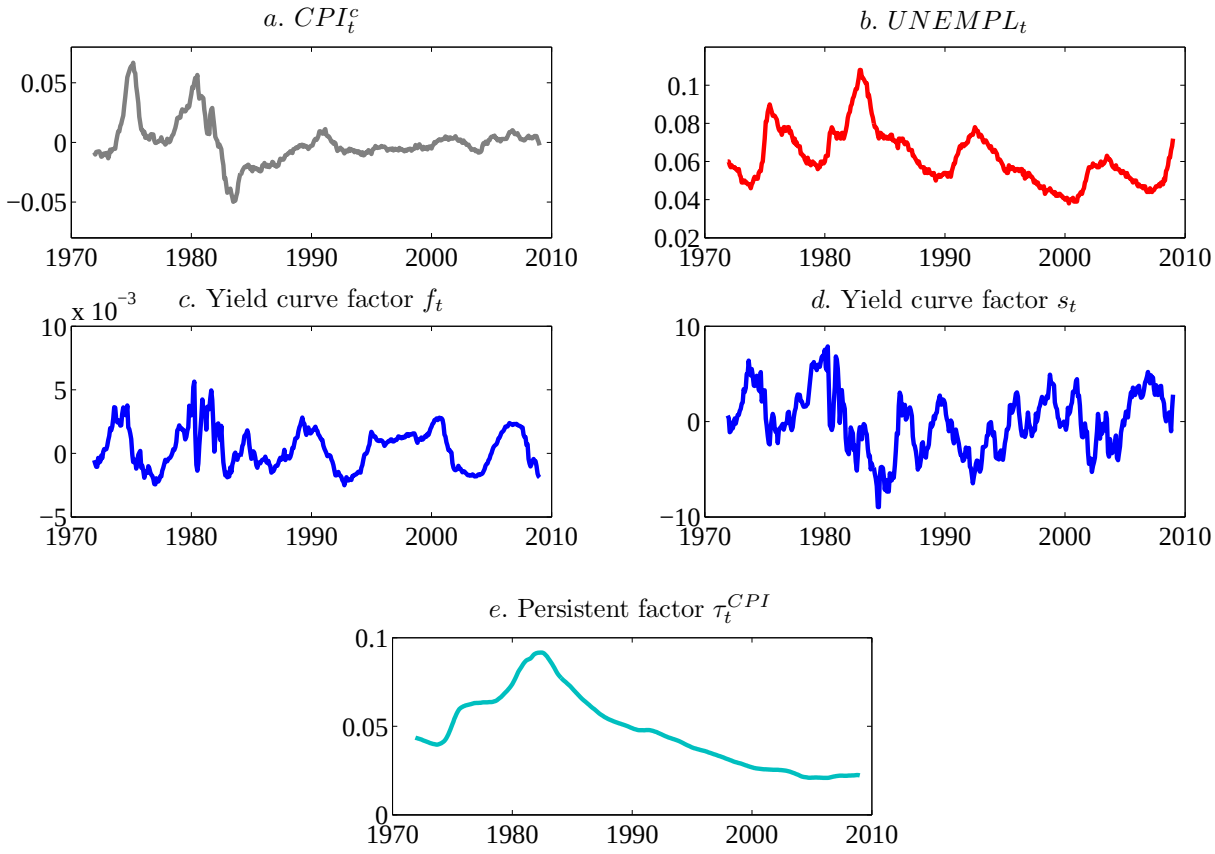


Figure A.5: Macro-finance model: filtered yield curve factors

The figure plots filtered yield curve factors from the macro finance model. The sample period is 1971–2009. The model is estimated with the maximum likelihood and the Kalman filter.

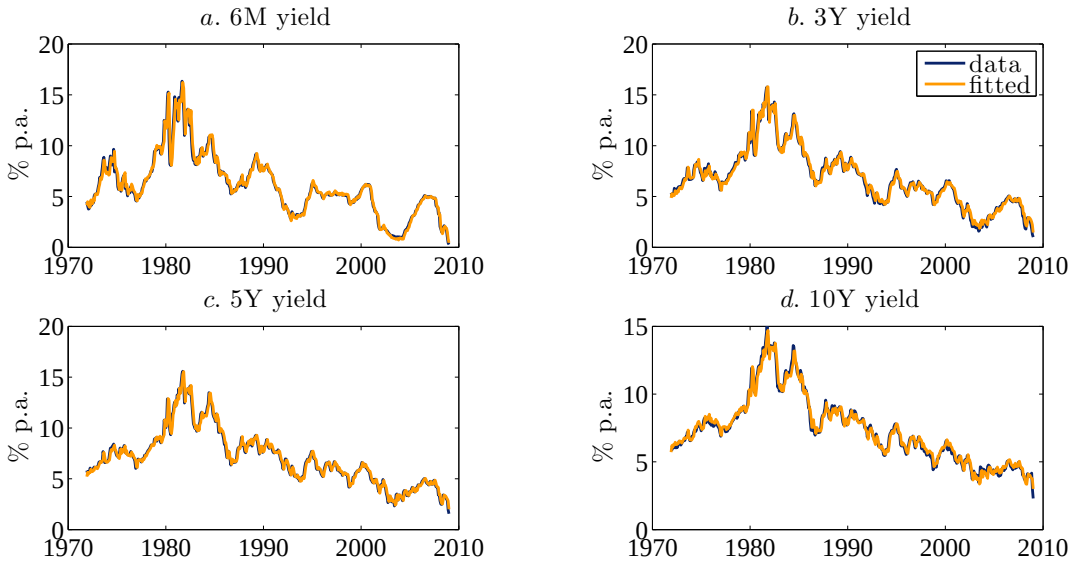


Figure A.6: Macro-finance model: fit to yields

The figure plots observed and fitted yields for maturities six months, three, five and ten years. The sample period is 1971:11–2009:09.

A.9 Out-of-sample tests

Below we describe the implementation of the bootstrap procedure to obtain the critical values for the ENC-NEW test. The test statistic for maturity n is given by:

$$\text{ENC-NEW}^{(n)} = (T-h+1)\,\frac{\sum_{t=1}^{T}\left(u_{t+12}^{2,(n)} - u_{t+12}^{(n)}\varepsilon_{t+12}^{(n)}\right)}{\sum_{t=1}^{T}\varepsilon_{t+12}^{2,(n)}}, \tag{A.36}$$

where T is the number of observations in the sample, $\varepsilon_t^{(n)}$ and $u_t^{(n)}$ denote the prediction error from the unrestricted and restricted model, respectively, and h measures the forecast horizon, in our case $h = 12$ months. Note that the time step in (A.36) is expressed in months.

Our implementation of bootstrap follows Clark and McCracken (2005) and Goyal and Welch (2008). To describe the dynamics of yields and to obtain shocks to the state variables generating them, we assume that the yield curve is described by four principal components following a VAR(1). Persistent component τ_t is assumed to follow an AR(12) process. We account for the overlap in bond excess returns by implementing an MA(12) structure of errors in the predictive regression. Imposing the null of predictability by the linear combination of forward rates, we estimate the predictive regression, the VAR(1) for yield factors and VAR(12) for τ_t by OLS on the full sample. We store the estimated parameters and use the residuals as shocks to state variables for the resampling. We sample with replacement from residuals and apply the estimated model parameters to construct the bootstrapped yield curve, the persistent component and bond excess returns. To start each series, we pick a random date and take the corresponding number of previous observations to obtain the initial bootstrap observation. In our case, the maximum lag equals 12, hence we effectively sample from $T-12$ observations. We construct 1000 bootstrapped series, run the out-of-sample prediction exercise and compute the ENC-NEW statistic for each of the constructed series. We repeat this scheme for different maturities. The critical value is the 95-th percentile of the bootstrapped ENC-NEW statistics.

The out-of-sample R^2 proposed by Campbell and Thompson (2008) is defined as:

$$R_{\text{OOS}}^{2,(n)} = 1 - \frac{\sum_{t=1}^{T-12}\left(rx_{t+12}^{(n)} - \widehat{rx}_{cyc,t+12}^{(n)}\right)^2}{\sum_{t=1}^{T-12}\left(rx_{t+12}^{(n)} - \overline{rx}_{t+12}^{(n)}\right)^2}, \tag{A.37}$$

where the time step t and sample size T is expressed in months. $\widehat{rx}_{cyc,t+12}^{(n)}$ is the forecast of annual excess return based on time t cycles, where the parameters of the predictive model are estimated using cycles up to time $t-12$ and returns realized up to time t. $\overline{rx}_{t+12}^{(n)}$ is the return forecast using historical average excess return estimated through time t.

Chapter 2

Understanding the term structure of yield curve volatility

We study the structure, economic content and pricing implications of the fluctuating covariance matrix of interest rates.[1] Using almost two decades of tick-by-tick bond transaction data, we obtain a novel look into the development of interest rates and their volatilities, and show how these two pieces complement each other. On this basis, we design a no-arbitrage term structure model able to simultaneously accommodate the dynamics of both curves.

We contribute to the existing yield curve literature along two dimensions. First, on the methodological front, we are the first to embed information from the realized yield covolatilities within the estimation of a term structure model. That step allows to uncover a multivariate nature of factors in the yield volatility curve. Second, we use the model-based decomposition to argue that volatility states carry economic information that cannot be read from yields observed at infrequent intervals. Most notably, factors at the short and long end of the volatility curve encode the duration structure of macroeconomic uncertainties and are informative about different aspects of market-wide liquidity.

The model has three standard yield curve factors. Additionally, to generate an adequate variation in the covariance matrix of yields, we introduce a dynamic dependence between those

[1]This chapter is based on the paper under the same title written in collaboration with Pavol Povala. We thank Torben Andersen, Luca Benzoni, John Cochrane, Jerome Detemple, Fabio Trojani, Pietro Veronesi, Liuren Wu, Haoxiang Zhu and seminar and conference participants at the Federal Reserve Bank of Chicago, University of Chicago, Baruch College, European Finance Association Meeting (2010), 3rd Annual SoFiE Meeting, EC^2, European Winter Finance Summit (Skinance), 4th Financial Risk International Forum, TADC London Business School, SFI NCCR PhD Workshop in Gerzensee for their comments.

factors. The stochastic covariance process we adopt gives rise to a three-variate model of yield volatilities, and thus aligns well with the empirical properties of the realized yield covariation that we document in a model-free analysis. The specification discriminates between the sources of persistence and shocks in yield versus volatility factors. With sufficient flexibility along both dimensions, no additional parameter constraints turn out to be required to explain the data.

While few would contend that higher-dimensional settings are key to explain the interest rate risk (e.g. Joslin, 2007; Kim, 2007b; Andersen and Benzoni, 2010), the most comprehensive models typically do not exceed dimension four and allow at most two volatility states. Going beyond this scope appears empirically desirable, yet it also raises two important concerns: (*i*) high parametrization, and (*ii*) the inability to identify bond volatilities from yields alone.[2] The structure we consider here adds flexibility for modeling volatility, but with 13 parameters in the physical dynamics it remains tractable by the standards of affine term structure models (ATSMs). As for the second concern, we do agree that identification of volatilities from the low-frequency yield data could frustrate any model. However, by constructing a high-quality proxy for the stochastic covariance matrix of yields, and exploiting it in estimation, we are not troubled by this point.

The model achieves a good performance in explaining the yield covariance matrix across a range of maturities without sacrifices in fitting yields. Its statistical record authorizes our investigation of the model-implied states, which so far has not been ventured in the literature. Several findings are worth highlighting.

The identified volatility states correspond closely with the roles of the respective yield curve factors. In both, we disentangle short- versus long-end components that play distinct roles along the curve. Those shorter-term factors, governing yields and volatilities at maturity of two to three years, are more transient and erratic. The longer-term ones, instead, that are responsible for maturities from beyond five years, exhibit more persistent and smoother dynamics. Additionally, we identify a covolatility state, which captures the comovement between the long and the intermediate region of the curve. The presence of several volatility variables extends the

[2]Joslin (2006, 2007) discusses a model with two conditionally Gaussian and two CIR states, one of which is dedicated to volatilities. The volatility state is identified from interest rate derivatives. By Dai and Singleton (2000) such a model is classified as $A_2(4)$. Outside the pure latent factor domain, macro-finance delivers some examples of models with more than two variables displaying stochastic volatility. We discuss them in the literature review below. An important group of papers studies high dimensional HJM settings (e.g. Trolle and Schwartz, 2009; Han, 2007). Yet, by taking the current yield curve as given, these are not of direct comparison to the equilibrium motivated ATSMs.

evidence pertaining to the preferred affine models, $A_1(n)$,[3] which have dominantly focused on the single factor short-rate volatility.

While we do not exclude volatility states from entering the cross-section of yields, we find their cross-sectional importance to be quantitatively negligible compared to that of the interest rate states. In fact, only the long-end volatility factor can move the yield curve by a visible amount. But its impact does not exceed seven basis points on average, and is limited to maturities above five years. There is little hope that the information about volatilities can be extracted from the cross-section of yields alone. We show that the backing-out approach, i.e. inverting the term structure equation from a subset of yields to recover states, fails at identifying even one, let alone three volatility factors. For that reason, our estimation relies on filtering and extra spot covolatility measurements. This latter choice sets our evidence apart from the literature relying on interest rate derivatives that, in general, are not delta-neutral, and involve an additional layer of modeling assumptions. The use of realized covolatility, instead, gives us a clean view of the volatility factors, and allows a precise assessment of their interactions with the yield curve.

We find that the model-implied factors contain economic information, even though they are not directly linked to specific macro quantities by design. Given the notion that prices should reflect economic prospects more than past events, we use survey-based forecasts about key macro variables instead of their realized numbers. The combination of expectation and uncertainty proxies turns out to be highly informative about the filtered dynamics, being able to explain up to above 90% and 50% of variation in the yield and volatility factors, respectively. Importantly, macro variables related to short- versus long-run states form disjoint sets. Stemming from long-duration bonds, the longer-end volatility shows a pronounced response to the persistent real activity measures such as expectations of GDP growth, or uncertainty about unemployment. The short-term volatility, in turn, is linked to the uncertainty and expectations about the monetary policy, and uncertainties surrounding inflation and industrial production, both of which give a more short-lived description of the economic conditions in our sample. Prospects of the housing sector emerge as the only variable with a jointly significant effect on short- and long-end volatility components. Finally, the covolatility state is associated with uncertainty proxies on monetary policy and the real economy, showing that those variables influence the correlation between the intermediate and long region of the curve.

[3]The naming convention used here follows the convenient taxonomy of ATSMs introduced by Dai and Singleton (2000). $A_m(n)$ denotes an n-factor model, in which m factors feature stochastic volatility.

Aggregating the intraday dynamics in on-the-run yields, the filtered volatility factors turn out to be informative about the variation in the market-wide liquidity. We find an intriguing pattern of correlations between the segments of the volatility curve and different liquidity measures proposed by the recent literature. The long-duration volatility component captures almost half of the variation in the common liquidity factor that underlies the on/off-the-run premium across maturities. That element of liquidity reflects the tightness of financial conditions, and typically preempts an action of the monetary authority (Fontaine and Garcia, 2010). Quite differently, the short-run volatility dominates in explaining the transitory episodes of liquidity dry-ups that do not necessarily trigger a monetary policy reaction (e.g. GM/Ford downgrade), but increase the cross-sectional price deviations in the US Treasury market (Hu, Pan, and Wang, 2010).

By linking the time series (volatility) with the cross-sectional (liquidity) dimension, these results provide a new perspective on the unspanned factors in the yield curve. Although our filtered volatility states are not revealed by the cross section of yields, they aggregate information that is present in the on-the-run Treasury market, but can be extracted only at higher frequencies. Likewise, liquidity factors contain information that cannot be identified using a smooth on-the-run Treasury curve, but are revealed in small cross-sectional price movements; as such they are by construction unspanned. Our results suggest that these two dimensions mirror a common underlying state of the economy.

Related literature

Recent research into interest rate volatility has evolved in at least three loosely related directions. Below, we provide a review of their different leading themes: (*i*) unspanned factors, (*ii*) realized volatility and jumps, (*iii*) non-Gaussian models. Our work touches upon these three strands. In addition, our paper is also related to the fast growing literature on funding liquidity.

Unspanned volatility. Several papers document a weak relation between the bond volatility, realized as well as derivative-based, and the spot yield curve factors. Collin-Dufresne and Goldstein (2002, CDG), Heidari and Wu (2003), and Li and Zhao (2006), among others, all conclude that movements in yields explain at best only about a half of the variation in interest rate derivatives.

Not surprisingly, the hypothesis of such unspanned risk in the bond market has provided for both active research and a controversy in the affine term structure literature. CDG (2002) formalize its intuition for the standard three-factor ATSMs under the name of unspanned stochastic volatility.

By providing testable predictions, the USV theory has triggered increased interest in the ATSMs' ability to explain the yield volatility dynamics. Here the evidence is mixed. In support of the unspanning hypothesis, Collin-Dufresne, Goldstein, and Jones (2009, CDGJ) report that, over the 1988–2005 sample, variance series generated by the standard $A_1(3)$ model are essentially unrelated to the model-free conditional volatility measures. Jacobs and Karoui (2009) find a correlation up to 75% using the same model estimated on Treasury yields over the 1970–2003 period. In the more recent sample 1991–2003, however, this correlation breaks down and becomes slightly negative at the long end of the curve.

While imposing the USV restriction does improve the ability of low-dimensional ATSMs to fit the time-series of volatilities, it also comes at a cost of higher cross-sectional pricing errors. As a consequence, several papers reject the USV in favor of an unconstrained model (Bikbov and Chernov, 2009; Joslin, 2007; Thompson, 2008). This evidence mostly concerns ATSMs with four factors[4] and a univariate volatility structure. What lies at the heart of the unspanning hypothesis, however, is an economic effect more than a failure of a specific model restriction. In line with this intuition, Kim (2007b) highlights a major demand for term structure models, not necessarily USV ones, with sufficient flexibility to match jointly yields and their volatility dynamics. Our analysis focuses on designing, implementing, and deriving the implications of such a model. For the 1992–2007 sample period, the model-generated term structure of conditional volatilities consistently tracks the observed series with an R^2 exceeding 96%.

Realized volatility. In combination with recent advances in high-frequency finance, the USV debate has encouraged a new model-free look into the statistical properties of bond volatility. Andersen and Benzoni (2010, AB) test empirically the linear spanning restriction of ATSMs using measures of realized volatility over the 1991–2000 period. They confirm that systematic volatility factors are largely independent from the cross section of yields, and call for essential extensions of the popular models on the volatility front. While our model is cast within the general affine framework, we attain this goal by combining two ingredients: the rich form of covolatility states plus their identification from the realized data.

The availability of high-frequency observations from spot and futures fixed income markets has revived interest in the impact of economic news releases on bond return volatility and jumps. As such, this literature has remained mostly empirical. Focusing on the discontinuities in bond returns, Wright and Zhou (2009) show for instance that the mean jump size extracted from the

[4]Bikbov and Chernov (2009) test the USV in the $A_1(3)$ model studied by CDG (2002). Thompson (2008) extends the evidence to the $A_1(4)$ USV model. $A_2(4)$ is considered in Joslin (2007).

30-year interest rate futures has a significant predictive power for expected excess bond returns. Indeed, relative to other liquid asset markets, bond prices tend to provide the most clear and pronounced reaction to economic news (Andersen, Bollerslev, Diebold, and Vega, 2007; Jones, Lamont, and Lumsdaine, 1998). These studies suggest that a rich economic content is present in bond volatilities. We provide a model-based decomposition of the volatility curve, and find that its components have different reactions to measures of economic conditions.

From latent to macro-motivated models. Several recent papers mark an important development by going beyond the standard Gaussian macro finance setup. Examples include Campbell, Sunderam, and Viceira (2010, CSV), Adrian and Wu (2009), Hautsch and Ou (2008), Bekker and Bouwman (2009) or Haubrich, Pennacchi, and Ritchken (2008, HPR). Some features unify the economics of these models. In particular, volatility is multivariate, and reflects different sources of risk. Both, Adrian and Wu (2009) and CSV highlight the importance of stochastic covariation between the real pricing kernel and (expected) inflation in determining excess bond returns. Likewise, Hautsch and Ou (2008) find ex post that the extracted persistent volatility factors are important for explaining bond excess returns. HPR introduce a similar effect through the relationship between inflation and the real interest rate.

In an extension of the latent factor approach, these models attach economic labels to different yield volatility components. To the extent that the volatility itself remains unobservable or is extracted from an auxiliary model, the identification and interpretation of its components relies on specific model assumptions. Indeed, explaining the volatility curve per se is not in direct focus of those models. We, in contrast, start completely latent, and having explained yields and volatilities, try to understand the impact of economic quantities on the states forming both curves.

Funding liquidity. The 2007–2009 financial crisis has stressed the importance of funding liquidity. Relating to recent advances in measuring liquidity from Treasury data, we point to a relationship between volatility and liquidity factors extracted from bond prices. Fontaine and Garcia (2010) use on- and off-the-run Treasuries to obtain a factor that tracks the value of funding liquidity. More recently, Hu, Pan, and Wang (2010) propose a noise illiquidity measure obtained as an average yield pricing error of Treasury bonds. We show that yield volatility factors at short and long end of the curve feature distinct links to those two concepts of market-wide liquidity.

2.1 Empirical facts about the term structure of yield volatilities

This section describes our data set. We discuss the properties of yield covolatilities and their relation with the yield curve. Evidence collected here sets the stage for our model design in Section 2.2.

2.1.A Data

We are the first to analyze and model yield volatility with the help of high-frequency Treasury bond data spanning two long expansions, one recession and three monetary cycles in the US economy. We obtain 16 years' worth of high-frequency price data of US Treasury securities covering the period from January 1992 through December 2007. We construct the sample by splicing historical observations from two inter-dealer broker (IDB) platforms: GovPX (1992:01–2000:12) and BrokerTec (2001:01–2007:12). The merged data set covers the majority of transactions in the US Treasury secondary market with a market share of 60% and 61% for GovPX and BrokerTec, respectively (Mizrach and Neely, 2006). As such, our dataset provides a comprehensive description of the contemporary yield curve environment. Beside the unavailability of high frequency Treasury bond data prior to 1991 when the GovPX started operating, other reasons speak against considering longer samples. Most importantly, there is empirical evidence that the conduct of monetary policy changed significantly during the eighties (e.g., Ang, Boivin, Dong, and Loo-Kung, 2009). The market functioning has also shifted dramatically with the advent of computers (automated trading), interest rate derivative instruments, and the swap market. Capturing such institutional features is not the object of our analysis.

GovPX comprises Treasury bills and bonds of maturities: three, six and 12 months, and two, three, five, seven, ten and 30 years. BrokerTec, instead, contains only Treasury bonds with maturities: two, three, five, ten and 30 years. In the GovPX period, we identify on-the-run securities and use their mid-quotes for further analysis. Unlike GovPX, which is a voice-assisted brokerage system, BrokerTec is a fully electronic trading platform attracting vast liquidity and thus allowing us to consider traded prices of the on-the-run securities. In total, we work with around 37.7 million on-the-run Treasury bond quotes/transactions. Appendix B.1.A reports the average number of quotes and trades per day that underlie our subsequent analysis.

The US Treasury market is open around the clock, but the trading volumes and volatility are concentrated during the New York trading hours. Roughly 95% of trading occurs between 7:30AM and 5:00PM EST (see also Fleming, 1997). This interval covers all major macroeconomic

and monetary policy announcements, which are commonly scheduled either for 9:00AM EST or 2:15PM EST. We consider this time span as a trading day. Especially around US bank holidays, there are trading days with a very low level of trading activity. In such cases, we follow the approach of Andersen and Benzoni (2010) and delete days with no trading for more than three hours.

We sample bond prices at ten-minute intervals taking the last available price for each sampling point. We choose this sampling frequency so that it strikes the balance between the non-synchronicity in trading and the efficiency of the realized volatility estimators (Zhang, Mykland, and Ait-Sahalia, 2005). The microstructure noise does not appear to be an issue in our data, as indicated by the volatility signature plots and very low autocorrelation of equally spaced yield changes (see Appendix B.1.B).

While the raw data set contains coupon bonds, it is crucial for our analysis to have precise and timely estimates of zero coupon yields. Using equally-spaced high-frequency price data, we construct the zero coupon yield curve for every sampling point. To this end, we apply smoothing splines with roughness penalty as described in Fisher, Nychka, and Zervos (1994). We purposely avoid using the Nelson-Siegel type of method because it tends to wash out some valuable information. The technical details on our zero coupon yield curve estimation are collected in Appendix B.1.C.

The liquidity in the secondary bond market is concentrated in two-, three-, five- and ten-year securities (see also Fleming and Mizrach, 2008, Table 1). We assume that the dynamics of this most liquid segment spans the information content of the whole curve. Since any bootstrapping method is precise for maturities close to the observed yields, for subsequent covolatility analysis we select yields which are closest to the observed coupon bond maturities.

2.1.B Realized yield covariances

Our analysis of interest rate risks focuses on nominal bonds. The high-frequency zero curve serves as an input for the calculation of the realized covariance matrix of yields. We consider zero yields with two, three, five, seven and ten-year maturities. Let y_t be the vector of yields with different maturities observed at time t. Time is measured in daily units. The realized covariance matrix is constructed by summing up outer products of a vector of ten-minute yield changes, and aggregating them over the interval of one day $[t, t+1]$:

$$RCov(t, t+1; N) = \sum_{i=1,\ldots,N} \left(y_{t+\frac{i}{N}} - y_{t+\frac{i-1}{N}} \right) \left(y_{t+\frac{i}{N}} - y_{t+\frac{i-1}{N}} \right)'. \tag{2.1}$$

$N = 58$ is the number of equally spaced bond prices (yields) per day t implied by the ten-minute sampling, and i denotes the i-th change during the day. The weekly or monthly realized covariances follow by aggregating the daily measure over the corresponding time interval. To obtain annualized numbers, we multiply $RCov$ by 250 for daily, 52 for weekly or 12 for monthly frequency, respectively. Based on Jacod (1994) and Barndorff-Nielsen and Shephard (2004), for frequent sampling the quantity (2.1) converges to the underlying quadratic covariation of yields. In Section 2.7, we positively assess the robustness of this estimator, and compare it to the alternatives proposed in the literature.

We aim to ensure that our volatility measures reflect views of active market participants rather than institutional effects. This motivates the following two choices: First, our construction of $RCov$ dynamics relies exclusively on the within-day observations, excluding the volatility patterns outside the US trading hours. Even though between- and within-day volatilities track each other fairly closely, we observe several instances of substantial differences and abrupt spikes in the between-day volatility pattern, which we cannot relate to any major news on the US market. To account for the total magnitude of volatility, we add to the within-day number the squared overnight yield change from close (5:00PM) to open (7:30PM). We then compute the unconditional average of the total and within-day realized yield covariation, respectively, and each day scale the within-day $RCov$ dynamics by the total-to-within ratio.

The second choice lies in focusing on the intermediate and long maturities (two to ten years), i.e. very liquid and frequently traded bonds. The short end of the curve (maturities of one year and below) is deliberately excluded from the realized covariance matrix computations for several reasons. Over our sample period, this segment of the curve exhibited a continuing decline in trading and quoting activity, and was completely suspended in March 2001 (see Appendix B.1.A).[5] A lower liquidity at the short end of the on-the-run curve is documented by Fleming (2003). Moreover, relative to the latter part, the dynamics of the short segment is complicated by its interactions with the LIBOR market and monetary policy operations. Such distortions, while interesting in their own right, are not directly relevant to the analysis we perform.

[5]The decline in the trading of short maturity bonds is not particular to GovPX or BrokerTec data. A similar development took place in the interest rate futures market.

2.1.C Information in second moments of yields

Table 2.1 reports summary statistics for weekly yields (panel *a*) and realized volatilities (panel *b*). Figure 2.1 plots average curves, both unconditional and contingent upon the monetary policy cycle. A monotonically increasing term structure of average yields is accompanied by a humped term structure of volatilities, with the hump occurring at the three-year maturity. In our sample, monetary easing not only increases the slope of the first curve, but also lifts the level in the latter and the magnitude of the hump. While both exhibit non-normalities, the statistical properties of the two objects are very different. Not surprisingly, the non-normality becomes more pronounced in the term structure of volatilities.

[Figure 2.1 and Table 2.1 here.]

Compared to the smooth evolution of yields, the volatility curve experiences periods of elevated and abruptly changing dynamics apparent in Figure 2.2. This autonomy of the volatility process encourages a more detailed look into its behavior. In the remainder of this section we examine the second moments of yields along three dimensions: (*i*) changes in their dynamic properties, (*ii*) the number of underlying factors, and (*iii*) their potential link to the yield levels.

[Figure 2.2 here.]

Level, slope and curvature viewed dynamically. Much of the intuition about factors driving the zero curve has been obtained from the principal component analysis (PCA) of the unconditional covariances of yields (Litterman and Scheinkman, 1991). Such analysis aggregates decades of yield curve information into a single set of numbers. In contrast, Eq. (2.1) combined with the high-frequency data provides a proxy for the unobservable conditional covariance matrix, and allows its dynamic decomposition. This step tells us that the unconditional PCA, usually applied to motivate three-factor models, washes away some valuable information about factors driving yields. While dynamically we do find three main factors, their relative importance fluctuates over time. The nature of factors can change with instances of slope moves taking the lead over level moves, and curvature moves—over the slope. The portion of yield variation explained by the level factor, typically exceeding 90%, can at times drop to just above 50%.[6]

[6]Details of the dynamic decomposition are omitted to conserve space and are available upon request. Additionally, we perform a formal log-likelihood test (see e.g. Fengler, Härdle, and Villa (2001) for details of the testing procedure). We find a strong rejection of: (*i*) a constant covariance matrix hypothesis (constant eigenvalues

We decompose the unconditional covariance matrix of four yields with maturities two, three, five and ten years. Factor loadings from the unconditional PCA serve to construct the level, slope and curvature tick-by-tick, and to estimate their realized correlations as plotted in Figure 2.3.[7] All correlations display a persistent pattern over time. For instance, the correlation between the slope and the level factor oscillates between ±50%, and is generally lower during periods of monetary easings. Intuitively, interest rate cuts tend to increase the slope of the curve on concerns of looming inflation. Instead, periods of monetary stability make the short and long end of the curve more independent. However, the superposition of these correlations against the Fed regimes in Figure 2.3 also shows that interactions between factors are more complex than just a monetary policy response. For modeling, this picture translates into the requirement of time-varying dependence between state variables determining yields.

[Figure 2.3 here.]

Factors in volatilities. We find that yield volatilities do not move on a single determinant. Similar to the cross-section of yields, at least three factors are also needed to explain the dynamics of the realized volatility curve. The first three principal components explain 90.2%, 6.1% and 2.3% of its variation (panel *b*). Importantly, this observation comes from analyzing middle to long yields only, and thus is not driven by idiosyncratic volatility at the very short end of the curve.

[Figure 2.4 here.]

From the modeling perspective, this result raises a natural question: Are factors driving volatilities related to those those typically found in yields?

Link between interest rates and volatilities. Much of the theoretical and empirical evidence points to a link between the level of interest rates and their volatility. The affine or quadratic models, for instance, imply that the same subset of factors determines both yields and their

and eigenvectors), and (*ii*) a common principal component hypothesis (constant eigenvectors but time-varying eigenvalues). Both hypotheses are tested against the alternative under which the covariance matrices do not have a constant factor structure across subsamples. In addition to performing the test over the whole sample period, we check the stability of monthly conditional covariances year-by-year. The test consistently rejects both hypotheses. While the rejection of the first one is not surprising and could easily arise from changing yield volatilities, the latter one is more important: It confirms that the space spanned by eigenvectors, giving rise to the level, slope and curvature interpretation, is in fact not stable across periods.

[7]The results persist if we follow a different portfolio construction that is immunized to all but one type of yield curve movements, and rebalanced at the beginning of each month. The dynamics of factor correlations are very close to those obtained with the unconditional PCA loadings.

volatilities. As one simple example, a single-factor CIR model suggests that the volatility is high whenever the short rate is high—a prediction that remained valid through the early 1980s (Chan, Karolyi, Longstaff, and Sanders, 1992).

More recently, however, the USV literature has argued that the yield-volatility relation is weak. We add to this evidence by showing more complex interactions than those implied by the linear regressions used in the USV tests. Figure 2.5 scatter-plots weekly realized volatilities against the level of interest rates with a matching maturity. The shape of the nonparametric regression fitted to the data discards the possibility of a positive correlation between yields and volatilities in our sample period. If any, the relationship appears to be asymmetrically U-shaped, which clearly contrasts with the early 1980s' episode.[8] The volatility is low for the intermediate interest rates range, and increases when rates move to either end of the spectrum. The rise in volatility is more pronounced in low interest rate regimes, and thus explains the negative unconditional correlations between yields and volatilities reported in panel *c* of Table 2.1. The last panel of Figure 2.5, which plots realized volatility conditional on Federal funds rate changes, provides a simple illustration of this point: Federal funds rate cuts induce a stronger upward revision in volatility than do tightenings. The asymmetry is most pronounced for shorter maturities (two years) and decays at the longer end of the curve.[9]

Not surprisingly, evidence on the relationship between interest rates and volatility is mixed and controversial. Figure 2.6 provides one explanation to the lacking consensus. We plot beta-coefficients and R^2's in regressions of realized volatilities on yields with two-, five- and ten-year maturities performed on daily data over a four-month rolling window. The yield-volatility link turns out to be highly state dependent. It reveals large fluctuations in the R^2 and switching signs of the regression coefficients. Yet, despite the obvious instability, it is hard to argue that the link is completely non-existing and should be discarded as a matter of principle.

[Figure 2.5 and Figure 2.6 here.]

For the sake of model design, we can learn how shocks in yields and volatilities are interrelated by estimating a VAR for the joint system. We include three bond portfolios mimicking the level,

[8] David and Veronesi (2009) show that a non-monotonic (V-shaped) relationship between yields and volatilities can arise as a consequence of investors' learning about extreme inflationary and deflationary states.

[9] To understand the type of non-linearity in the yield-volatility relationship, we fit a generalized additive model (GAM) with a linear and a spline part, i.e. $v_t^\tau = \beta_0 + \beta_1 y_t^\tau + s(y_t^\tau) + \varepsilon_t$. The result of the exercise is twofold. For short and intermediate maturities (of two and five years) both components are significant, with a negative β_1 loading and an asymmetric spline component. For the long end of the curve (ten years), in turn, we find no support of a linear component, and a weak confirmation of the U-shaped asymmetry.

slope and curvature of the yield curve plus three realized volatility factors: the level RV_t^{2Y}, the slope $(RV_t^{10Y} - RV_t^{2Y})$ and the covariance $RCov_t^{5Y,10Y}$. The highest cross-correlation between shocks is low and does not exceed 16% (the yield level portfolio and the volatility level).

2.2 The model

A joint model of yields and volatilities is the key to quantifying the relationship between factors driving both curves. The model allows us to answer in a coherent way the empirical question about the spanning of volatility states by the level of yields, and is instrumental in analyzing the economic content of interest rate volatility.

Evidence of the previous section provides guidelines for our modeling approach. First, to generate a sufficiently rich variation in covariances of yields, we allow a multivariate volatility and dynamic interactions between factors. Second, while our empirical findings do invoke the notion of unspanning, we remain cautious about imposing it within the model. In fact, we show that such restriction is not required to fit yields and volatilities jointly. Rather, given findings on the number of factors in yields and volatilities, we do not expect a low-dimensional model (whether or not USV) to perform well on both fronts. In the remainder of this section, we formulate a sufficiently flexible model, and later verify its viability in terms of the econometric fit and economic interpretation of factors.

Our benchmark model is cast in a reduced-form continuous-time framework. In specifying the state dynamics, we take an agnostic view on factor labels, but assign them to two groups: (i) expectations factors X_t, and (ii) covariance factors V_t. The physical dynamics are given by the system:

$$dX_t = (\mu_X + \mathcal{K}_X X_t)dt + \sqrt{V_t}dZ_{X,t}^{\mathbb{P}} \tag{2.2}$$

$$dV_t = \left(\Omega\Omega' + MV_t + V_tM'\right)dt + \sqrt{V_t}dW_t^{\mathbb{P}}Q + Q'dW_t^{\mathbb{P}\prime}\sqrt{V_t}, \tag{2.3}$$

where X_t is a n-vector, and V_t is a $n \times n$ process of symmetric positive definite matrices—a covariance matrix process proposed by Bru (1991) and studied by Gourieroux, Jasiak, and Sufana (2009). Accordingly, $Z_X^{\mathbb{P}}$ and $W^{\mathbb{P}}$ are a n-dimensional vector and a $n \times n$ matrix of independent Brownian motions.[10] μ_X is a n-vector of parameters and $\mathcal{K}_X, M$ and Q are given as $n \times n$

[10]It is straightforward to introduce correlations between Z_X and W by setting $dZ = dW\rho + \sqrt{1-\rho'\rho}dB$ for some constant vector ρ, where dB is an n-dimensional Brownian motion independent of columns in dW. We state the

parameter matrices. To ensure a valid covariance matrix process V_t, we specify $\Omega\Omega' = kQ'Q$ with an integer degrees of freedom parameter k such that $k > n-1$, and require that Q is invertible. This last condition guarantees that V_t stays in the positive definite domain (see e.g., Gourieroux, 2006).

The short interest rate is an affine function of X_t variables, but contains an additional source of persistent shocks:

$$r_t = \gamma_0 + \gamma_X' X_t + \gamma_f f_t. \tag{2.4}$$

The state f_t evolves as:

$$df_t = (\mu_f + \mathcal{K}_f f_t + \mathcal{K}_{fX} X_t)dt + \sigma_f dZ_{f,t}^{\mathbb{P}}, \tag{2.5}$$

with $Z_{f,t}^{\mathbb{P}}$ denoting a single Brownian motion independent of all other shocks in the economy. $\gamma_f, \mathcal{K}_f$ and σ_f are scalars, and γ_X' and $\mathcal{K}_{fX}$ are $(1 \times n)$-vectors of parameters. For convenience, we collect X_t and f_t factors in a vector $Y_t = (X_t', f_t)'$, whose dynamics can be compactly expressed as:

$$dY_t = (\mu_Y + \mathcal{K}_Y Y_t)\, dt + \Sigma_Y(V_t) dZ_t^{\mathbb{P}}, \tag{2.6}$$

with a block diagonal matrix $\Sigma_Y(V_t)\Sigma_Y(V_t)' = \begin{pmatrix} V_t & 0 \\ 0 & \sigma_f^2 \end{pmatrix}$ and $\mathcal{K}_Y = \begin{pmatrix} \mathcal{K}_X & 0_{n\times 1} \\ \mathcal{K}_{fX} & \mathcal{K}_f \end{pmatrix}$.

Bonds in this economy are priced using the standard no-arbitrage argument. By its convenience, we can abstract from a particular preference structure, and specify a general reduced-form compensation $\Lambda_{Y,t}$ required by investors to face shocks in the state vector:

$$\Lambda_{Y,t} = \Sigma_Y^{-1}(V_t)\left(\lambda_Y^0 + \lambda_Y^1 Y_t\right), \tag{2.7}$$

where λ_Y^0 is a $(n+1)$-vector and λ_Y^1 is a $(n+1) \times (n+1)$ matrix of parameters. To be viable, this formulation requires the invertibility of each block of matrix $\Sigma_Y(V_t)$, which is ensured by the positive-definiteness of $\sqrt{V}_t$, and with σ_f different from zero. In Eq. (3.5) we assume that only Z shocks are priced. Therefore, the risk neutral dynamics of X_t follow from the standard drift adjustment:

$$\mu_Y^{\mathbb{Q}} = \mu_Y - \lambda_Y^0 \tag{2.8}$$

$$\mathcal{K}_Y^{\mathbb{Q}} = \mathcal{K}_Y - \lambda_Y^1, \tag{2.9}$$

general solution for $\rho \neq 0$ in the Appendix B.3. However, based on empirical findings of Section 2.1, in particular the very low correlation of shocks between volatilities and yields, we set $\rho = 0$ in the empirical implementation.

and the dynamics of V_t remain unchanged. It is technically possible to introduce a priced volatility risk without loosing the flexibility of the framework. However, in the presence of a weak spanning of volatility states by bonds it is difficult (if not impossible) to identify the market price of risk for volatility from bonds alone. For this, additional volatility-sensitive instruments such as bond options are needed.

Prices of nominal bonds are obtained by solving $P_t^\tau = E_t^{\mathbb{Q}}\left(e^{-\int_0^\tau r_s ds}\right)$. By the Feynman-Kac argument, and using the infinitesimal generator for the joint process $\{Y_t, V_t\}$, the solution for the nominal term structure has a simple affine form (see Appendix B.3):

$$P(t,\tau) = e^{A(\tau)+B(\tau)'Y_t+Tr[C(\tau)V_t]}, \tag{2.10}$$

where $Tr(\cdot)$ denotes the trace operator. The coefficients $A(\tau), B(\tau)$ and $C(\tau)$ solve a system of ordinary differential equations:

$$\frac{\partial A(\tau)}{\partial \tau} = B(\tau)'\mu_Y^{\mathbb{Q}} + \frac{1}{2}B_f^2(\tau)\sigma_f^2 + kTr\left[Q'QC(\tau)\right] - \gamma_0 \tag{2.11}$$

$$\frac{\partial B(\tau)}{\partial \tau} = K_Y^{\mathbb{Q}\prime}B(\tau) - \gamma_Y \tag{2.12}$$

$$\frac{\partial C(\tau)}{\partial \tau} = \frac{1}{2}B_X(\tau)B_X(\tau)' + C(\tau)M + M'C(\tau) + 2C(\tau)Q'QC(\tau), \tag{2.13}$$

where we split the $B(\tau)$ loadings as $B(\tau) = [B_X(\tau)', B_f(\tau)]'$ and $\gamma_Y = (\gamma_X', \gamma_f)'$. The boundary conditions for the system (2.11)–(2.13) are $A(0) = 0_{1\times 1}, B(0) = 0_{(n+1)\times 1}$ and $C(0) = 0_{n\times n}$. The $B(\tau)$ loadings have a simple form typical to the Gaussian models, and allow an immediate solution. The $C(\tau)$ matrix solves a matrix Riccati equation.

Defining $y_t^\tau = -\frac{1}{\tau}\ln P_t^\tau$, the term structure of interest rates has the form:

$$y_t^\tau = -\frac{A(\tau)}{\tau} - \frac{B(\tau)'}{\tau}Y_t - Tr\left[\frac{C(\tau)}{\tau}V_t\right]. \tag{2.14}$$

As a consequence of the dynamics (2.2)–(2.3), yields are an affine function of the entire state vector $(Y_t, \text{vec}(V_t)')'$. Under uncorrelated shocks dZ and dW, and the short rate (2.4), we leave only one channel open through which volatility states appear in the yield curve equation, i.e. the diffusive term in the Y_t dynamics (2.6). Thus, the instantaneous yield covariation is exclusively driven by the covariance factors:

$$
\begin{aligned}
v_t^{\tau_i,\tau_j} &:= \frac{1}{dt}\langle dy_t^{\tau_1}, dy_t^{\tau_2}\rangle \\
&= \frac{1}{\tau_1\tau_2}\{Tr\left[B_X(\tau_2)B_X(\tau_1)' + 4C(\tau_2)Q'QC(\tau_1)\right]V_t + B_f(\tau_1)B_f(\tau_2)\sigma_f^2\}. \qquad (2.15)
\end{aligned}
$$

2.2.A Discussion

In the basic setup, we consider three variables in Y_t, i.e. f_t plus a two-dimensional vector X_t. The latter is equipped with a 2×2 covariance matrix V_t. The form of V_t leads to a two-plus-one-variate process with two volatility plus a covariance factor. Thus, it presents a three-factor model of yield volatilities. The combination of six factors gives us a scope to fit both yields and their volatilities.

We can think of f_t as a short-term monetary policy factor. X_t, instead, represents longer-term forces that reflect expectations about key elements of the economic landscape, e.g. the real and nominal sector. Naturally, they can impact the conditional expectation of f_t. Time-varying volatility enters the model through long term factors: V_t describes the amount of risk present in the economy, with its out-of-diagonal element V_{12} determining the conditional mix between X_t's.

Our split between volatility and expectation variables evokes the $A_m(n)$ classification of Dai and Singleton (2000). Still, at least two differences are worth highlighting. First, V_t represents a complete covariance matrix dynamics (i.e. volatilities plus covariances), and as such involves components which can switch sign. In contrast, independent CIR processes in ATSMs generate stochastic volatility one-by-one. Therefore, the covariances they imply are a linear combination of the volatility factors. Second, to make the roles of factors precise and interpretable, our specification intentionally excludes any interactions between V_t and X_t via the drift. Even though in ATSMs such interactions are usually allowed, we show in estimation that they are not called for by the data.

As dimensions grow, flexibility typically comes at the price of parsimony. The $A_1(4)$ model versions estimated in the literature usually involve over 20 parameters after excluding the market prices of risk. By these standards, the state space we consider is comparably large, but with six factors at work, it involves no more than 13 identified parameters (excluding $\Lambda_{Y,t}$).

The presence of V_t in expression (2.14) sets our approach apart from the USV settings, which explicitly prevent volatility factors from entering the cross section of yields. Collin-Dufresne, Goldstein, and Jones (2009) expose that such separation improves the ATSMs' fit to the spot volatility of yields. Considering our empirical results, however, there appear to be few reasons—except statistical ones—for such constraint to hold in reality. In fact, there are at least two channels through which volatility variables could appear in the term structure, in particular at

its long end. One of them is the well-known convexity bias through which volatility is revealed in bond prices (see also Phoa, 1997; Joslin, 2007). A second, and economically more important one, is the relation between the amount of uncertainty and the expected excess returns. In that the term premia compensate for risks, they should be related to the changing amount of interest rate volatility.

2.3 Model estimation

A practical implication of the weak link between yields and volatilities is that not all factors can be identified directly from yields. Thus, the backing-out technique—inverting Eq. (2.14) from observed yields to latent factors—would not work in our setting. We estimate the model on a weekly frequency ($\Delta t = \frac{1}{52}$) combining pseudo-maximum likelihood with a filtering technique. Since Y_t and V_t factors are unobservable, we express the model in a state-space form. At every date t, we filter the latent state by exploiting information both in yields and in volatilities.

2.3.A Transition dynamics

The transition equation for Y_t is specified as an Euler approximation of the physical dynamics (2.6):[11]

$$Y_{t+\Delta t} = \bar{\mu}_{Y,\Delta t} + \Phi_{Y,\Delta t} Y_t + u^Y_{t+\Delta t}, \tag{2.16}$$

where u^Y_t is a vector of heteroskedastic innovations $u^Y_t = \Sigma_Y(V_t)\sqrt{\Delta t}\epsilon_{t+\Delta t}$, and

$$\bar{\mu}_{Y,\Delta t} = \left(e^{\mathcal{K}_Y \Delta t} - I\right) \mathcal{K}_Y^{-1} \mu_Y \tag{2.17}$$

$$\Phi_{Y,\Delta t} = e^{\mathcal{K}_Y \Delta t}. \tag{2.18}$$

The transition equation for the matrix process V_t is obtained by an exact discretization of the dynamics (2.3):

$$V_{t+\Delta t} = k\bar{\mu}_{V,\Delta t} + \Phi_{V,\Delta t} V_t \Phi'_{V,\Delta t} + u^V_{t+\Delta t}, \tag{2.19}$$

where u^V_t represents a symmetric matrix of heteroskedastic innovations. Parameter matrices $\Phi_{V,\Delta t}$ and $\bar{\mu}_{V,\Delta t}$ are given as:[12]

[11] In Appendix B.2.B, we provide expressions for an exact discretization of the Y_t dynamics. We find that the use of the Euler scheme is virtually immaterial when $\Delta t = 1/52$, but offers a considerable increase in computational speed compared to the exact discretization. Therefore, the results presented here rely on the expression (2.16).

[12] The closed form solution for the integral $\bar{\mu}_{V,\Delta t}$ is given by $\int_0^{\Delta t} \Phi_{V,s} Q' Q \Phi'_{V,s} ds = -\hat{C}_{12}(\Delta t)\hat{C}'_{11}(\Delta t)$, where

$$\bar{\mu}_{V,\Delta t} = \int_0^{\Delta t} \Phi_{V,s} Q'Q \Phi'_{V,s} ds \tag{2.20}$$

$$\Phi_{V,\Delta t} = e^{M\Delta t}. \tag{2.21}$$

Details on transition dynamics are collected in Appendix B.4.

2.3.B Measurements

We introduce two types of measurement equations based on yields (y_t^τ) and their quadratic covariation ($v_t^{\tau_i,\tau_j}$):

$$y_t^\tau = f(S_t;\Theta) + \sqrt{R_y} e_t^y \tag{2.22}$$

$$v_t^{\tau_i,\tau_j} = g(V_t;\Theta) + \sqrt{R_v} e_t^v. \tag{2.23}$$

Functions $f(S_t;\Theta)$ and $g(V_t;\Theta)$ denote model-implied expressions (2.14) and (2.15) corresponding to the observed measurements y_t^τ and $v_t^{\tau_i,\tau_j}$; Θ collects model parameters. $v_t^{\tau_i,\tau_j}$ is obtained from the high-frequency zero curve using estimator (2.1).

The weekly realized covariance matrix estimator is still quite noisy. To alleviate the noise, we construct every week a four week rolling realized covariance matrix. Such an adjustment makes it easier for the Kalman filter to distinguish between the noise and the fundamental volatility.

We assume additive, normally distributed measurement errors e_t^y and e_t^v with zero mean and a constant covariance matrix. In estimation, we use six yields with maturities of six months and two, three, five, seven and ten years. Yields share an identical standard deviation of measurement errors. Hence, their error covariance matrix is $R_y = \sigma_y^2 I_6$. Additionally, we include three volatility measurements which comprise variances of the two- and ten-year bond and the covariance between the five- and ten-year bond. The assumption of constant and identical measurement errors in yields is innocuous and intuitive. Lacking a similar prior for volatility measurements, we allow for different errors across equations, i.e. $R_v = \text{diag}(\sigma_v^i)$, where $i = 1, 2, 3$, and $\text{diag}(\cdot)$ denotes a diagonal matrix.

Implicit in our choice of measurements are two approximations. The first one is the assumption that the realized and the instantaneous covariance matrices of yields are equivalent. Indeed,

$$\begin{pmatrix} \hat{C}_{11}(\Delta t) & \hat{C}_{12}(\Delta t) \\ \hat{C}_{21}(\Delta t) & \hat{C}_{22}(\Delta t) \end{pmatrix} = \exp\left[\Delta t \begin{pmatrix} M & -Q'Q \\ 0 & -M' \end{pmatrix}\right].$$

See Van Loan (1978) for the proof.

neglecting e_t^v, Eq. (2.23) implies that: $\frac{1}{dt}\langle dy_t^\tau \rangle = v_t^\tau$ and $\frac{1}{dt}\langle dy_t^{\tau_i}, dy_t^{\tau_j} \rangle = v_t^{\tau_i,\tau_j}$, where $dt = 1/52$. In a strict sense and in absence of jumps, however, the estimator (2.1) converges to the integrated—rather than instantaneous—covariance matrix of yields. We recognize that the measurement (2.23) is not exact, but on a weekly frequency the error due to the approximation can be assumed negligible.

Our assumption about the absence of jumps necessarily leads the second approximation. Even though extensions including discontinuities are readily possible, we specify the benchmark model as a pure diffusion. Importantly, vast portion of jumps in the term structure appears at deterministic times (scheduled macro announcements). By nature, these jumps differ from the Poisson type of events. This motivates our choice of a purely diffusive model. Here, the weekly horizon comes in handy again: In the context of term structure models, one week appears a sufficiently long period to avoid the break-down of a pure diffusion.

2.3.C Pseudo-maximum likelihood estimation

Despite linearity of the transition and measurement equations, the filtering approach introduced by Kalman (1960) is not directly applicable in our setting due to the non-Gaussian properties of the underlying state dynamics. In order to handle the non-Gaussianity we use the Unscented Kalman Filter (UKF), proposed by Julier and Uhlmann (1997), and recently applied in finance by e.g., Carr and Wu (2007) or Christoffersen, Jacobs, Karoui, and Mimouni (2009, CJKM). To approximate the conditional distribution, the UKF propagates the state through a set of deterministically chosen "sigma" points. Compared to the particle filter, it avoids costly simulations and thus offers a considerable gain in computational speed. This has a particular value for the estimation of multidimensional models like ours. Details on the UKF implementation are provided in Appendix B.6.

Collecting all measurements in vector m_{t+1}, let $\hat{m}_{t+1}^-$ and $\hat{P}_{m,t+1}^-$ denote the time-t forecasts of the time-$(t+1)$ values of the measurement series and of their conditional covariance, respectively, as returned by the filter (for convenience 1 means one week). By normality of measurement errors, we can compute the quasi-log likelihood value for each time point in our sample:

$$l_{t+1}(\Theta) = -\frac{1}{2}\ln|P_{m,t+1}^-| - \frac{1}{2}\left(\hat{m}_{t+1}^- - m_{t+1}\right)' \left(P_{m,t+1}^-\right)^{-1} \left(\hat{m}_{t+1}^- - m_{t+1}\right), \tag{2.24}$$

and obtain parameter estimates by maximizing the criterion:

$$\hat{\Theta} := \arg\min_{\Theta} \mathcal{L}\left(\Theta, \{m_t\}_{t=1}^{T}\right) \quad \text{with} \quad \mathcal{L}\left(\Theta, \{m_t\}_{t=1}^{T}\right) = \sum_{t=0}^{T-1} l_{t+1}\left(\Theta\right). \tag{2.25}$$

with $T = 845$ weeks. The initial log-likelihood is evaluated at the unconditional moments of the state vector (see Appendix B.2 for the expressions).

As a common problem in term structure modeling, the optimization of the loss function (2.25) is complicated by a high dimensionality of the parameter space, presence of multiple stochastic volatility factors, and complex interactions between parameters. Such circumstances leave little hope for standard local optimization methods, even when combined with a grid search. Therefore, to secure against local minima, we use the differential evolution (DE) algorithm designed for an efficient search of global optima in multidimensional, non-monotone, and multimodal problems (Price, Storn, and Lampinen, 2005). We confirm that the algorithm achieves the global minimum by repeating the estimation several times.

The latent nature of factors spells out the possibility that two distinct sets of parameters lead to observationally equivalent yields. To ensure econometric identification, we impose several parameter restrictions. Details of the model identification procedure as well as the discussion of restrictions are relegated to Appendix B.5.

2.4 Estimation results

We discuss the parameter estimates and in-/out-of-sample model performance. Then, we study the dynamics of the filtered states and their respective contributions to the term structure of yields and volatilities.

2.4.A Model performance

Table 2.2 provides parameter estimates for our model indicated by the label G_3SV_3 (meaning: three conditionally Gaussian plus three volatility factors). To obtain a comprehensive picture of its statistical properties, we consider two specifications: (i) a risk-neutral case, in which all parameters in $\Lambda_{Y,t}$ have been set to zero, and (ii) a risk premia case, with some of $\Lambda_{Y,t}$ parameters left free. On the level of explaining yields, we compare the performance of the estimated models to their corresponding Gaussian three-factor counterparts (G_3SV_0), which have a proven track record on this front. We estimate the purely Gaussian models with the standard Kalman filter.

This comparison puts to a test whether volatility factors, while beneficial for capturing second moments of yields, introduce a risk of misspecification on the side of yields.

[Table 2.2 here.]

That we consider models without risk compensation might surprise at first. Zero bond risk premiums are clearly untenable from an empirical perspective (e.g., Campbell and Shiller, 1991). Our motivation for this step is twofold. First, the ability of a model to describe the joint dynamics of yields and volatilities should be revealed from the structure of the underlying states rather than by adding new parameters. Indeed, we expect and demonstrate the risk premium specification to have a second-order effect on our model's performance in matching conditional volatilities. Second, while we certainly recognize that risk compensation exists in the bond market, we also find that the identification of price of risk parameters in $\Lambda_{Y,t}$ is a difficult task. Even for simple models, the estimates of premiums tend to have a low precision. With the inclusion of several parameters in $\Lambda_{Y,t}$, the search for a global optimum of (2.25) becomes slow and cumbersome. For this reason, we adopt a parsimonious form by judiciously selecting only two parameters in $\Lambda_{Y,t}$. We defer the details of this choice to Section 2.7.C. Interestingly, we find X_t's to be the only factors which have a significant impact on the term premiums.

Parameter estimates. After imposing identification restrictions, our preferred model presented in Table 2.2 has nine parameters which drive the yield curve (two of those describe market prices of risk), and six parameters which drive the term structure of volatilities. Even though the unobservability of states makes the interpretation of single estimates relatively uninteresting, several points stand out. Within each factor type at least one variable is more persistent and one faster-moving. While this observation may be common for the yield dynamics, our estimates indicate that it also holds for the volatility curve. Still, the most prominent difference between V_t states comes in their respective volatilities (vol of vol), as captured by the Q parameters. The volatility of the V_{11} and V_{12} state is just about a half of V_{22}. Such distinction is a first indication that a multivariate volatility is actually required by the data.

We also note large differences in the speed of mean reversion between Y_t and V_t factors: as expected, the autocorrelation coefficients decline much more rapidly for the volatility states. That variables generating yields have distinct statistical properties from those generating volatilities is consistent with the empirical findings of Section 2.1. Despite its intuitive appeal, this fact frustrates the volatility fit of many (especially three-factor) term structure models. But it is confirmed in our estimates of the V_t dynamics, specifically in the estimated degrees of freedom

k in (2.3). This single parameter controls the extent of non-Gaussianity present in the volatility factors: the higher the k the more Gaussian the dynamics.[13] Across all estimated versions, the model consistently selects $k = 2$. This low value reflects the need for factors with a pronounced right tail. Although such property suits the observed behavior of volatilities, it is less obvious how it could emerge in the term structure of interest rates. We interpret it as supportive of our design that separates distributional qualities of yields from those of volatilities.

In-sample fit. The last section of Table 2.2 provides the log-likelihood values for each estimated specification. For comparison with the purely Gaussian benchmark, we split the log-likelihood implied by our model into two portions measuring the fit to yields and to volatilities, respectively. A look at the numbers indicates that all versions have similar abilities in explaining yields. In particular, the improvement in volatility modeling offered by G_3SV_3 spells no sacrifice in terms of fitting interest rates (see Figure 2.7). Across a spectrum of maturities, the model replicates the observed weekly dynamics of yields with a high accuracy. Simultaneously, it is able to explain well the evolution of the second moments. Figure 2.8 superimposes the observed and model-implied behavior of the level and the slope of the volatility curve, and the covariance between the five- and ten-year yield. Level and slope are proxied by the two-year volatility and the spread between the ten- and two-year volatility, respectively. The plot shows that the model-implied dynamics track observed quantities very closely in terms of magnitudes, persistence, and signs. Our setting provides a good description of the volatility at the short and the long end of the curve.

[Figure 2.7 and 2.8 here.]

Table 2.3 confirms these conclusions. We provide two yardsticks of in-sample model performance: the root mean squared errors (RMSE) and the percentage of variation in yields and volatilities explained by the model. The latter is measured with the R^2 coefficient from a regression of the observed on the model-implied dynamics. The RMSEs show that on average the model misses the true yield by about two basis points. As such, it is able to explain virtually the total of observed variation in yields. Its performance is very similar to the purely Gaussian case. The match to the volatility dynamics is less perfect than that to yields, still we are able to explain from 94% to 98% of variation in the second moments dynamics.

[Table 2.3 here.]

[13] Buraschi, Cieslak, and Trojani (2008) discuss the role the degrees of freedom parameter in the V_t dynamics. See also Appendix B.2.A for the properties and construction of the V_t process. The representation of V_t as the sum of k outer products of the multivariate Ornstein-Uhlenbeck process provides the key intuition for the role of k.

Yield forecasting performance. When assessed in sample, our model could conceal distortions coming with a richer form of the state space, if misspecified. This should be revealed in a quick deterioration of its performance out of sample. To check this possibility, we re-estimate the model on the data from January 1992 through December 2004, and use the remaining part of the sample (January 2005 through December 2007) for out-of-sample evaluation. For comparison, we take the same approach for the purely Gaussian setting. With new estimates, we obtain predictions of yields over horizons of one, four, 12 and 52 weeks. Table 2.4, panel *b* summarizes the results. The yield forecasting ability implied by our model is very close to the purely Gaussian case. It is well-known that the existing term structure models fail in outperforming the random walk predictions out-of-sample (see e.g. Duffee, 2002; Moench, 2008). While at the very short forecast horizons (i.e. one week) both settings are outpaced by a simple random walk, they do a decent job in forecasting yields over longer term (e.g. one year). The upshot is that the additional volatility factors do not induce instability in the model performance.

[Table 2.4 here.]

Volatility benchmarks. In terms of explaining the volatility curve, the model we propose has no direct benchmark in the literature. The interpretation of the existing results is intricate as approaches differ by the data and sample period used (Treasury or swap rates, with or without interest rate derivatives), restrictions imposed (with or without USV), and estimation methods applied. These discrepancies often lead to different conclusions, thus making a direct comparison difficult.[14] We take $A_1(4)$ models estimated by CDGJ (2008), and Thompson (2008) as a benchmark. A consensus emerging from those papers is that the $A_1(4)$ model can do reasonably well in matching volatility at the short end of the curve, but its performance deteriorates quickly with the yield's maturity. For instance, CDGJ show that instantaneous volatility of the ten-year yield implied by the $A_1(4)$ model is 18 times less volatile than the rolling window volatility estimates, and suggest that more than one state variable is needed to capture the term structure of volatilities. The corresponding number returned by our model is 7%, and is coupled with the ability to explain 98% of the variation in the slope of the volatility curve.

[14] One example comes from comparing the results reported by Thompson (2008) and those reported by CDGJ (2008). Both papers estimate the $A_1(4)$ model with and without USV using swap and LIBOR rates. Thompson (2008) shows the model can track volatility dynamics similarly well whether or not the USV is imposed. CDGJ demonstrate, in contrast, that the USV restriction significantly improves the model's performance. By relying on the spot and option price data, the results for the $A_1(4)$ and $A_2(4)$ specification reported in Joslin (2007) are less well suited for drawing comparisons here.

Similar to the yield forecasting exercise, we compare out-of-sample volatility forecasts to a random walk. Given the fast decay of volatilities, their forecasts at horizons longer than a couple of months are not meaningful. Table 2.4, panel *b* summarizes the results. Our model performs particularly well for four and 12 week horizons, even though the out-of-sample period (2005:01–2007:12) privileges the random walk due a consistently low and stable level of volatilities. Across the maturity spectrum, the volatility of the five- and ten-year yield is predicted with the highest accuracy. This outperformance is explained by the presence of an additional volatility state which, as we argue below, is dedicated to capturing precisely this segment of the curve.

2.4.B Filtered states

This section discusses the properties and roles of filtered factors. Good fit to the data indicates that these factors preserve the substantial portion of the information in the cross-section of yields and volatilities.

Figure 2.9 plots the state variables extracted by the unscented Kalman filter. Panels on the left display the evolution of the three yield curve factors, $Y = (X_1, X_2, f)$, those on the right show the covariance factors $\bar{V} = (V_{11}, V_{12}, V_{22})$.

[Figure 2.9 here.]

This autonomous behavior of the two factor groups reveals no obvious candidate variable that could be acting equally on yields and on their second moments. This does not mean, however, that factors lack common interpretation. Quite the opposite: when fed with the data, the model puts the right pieces of its structure into the right places. Most importantly, the roles it assigns to the X_t states correspond neatly with the covariance interpretation of the respective volatility factors V_t.

[Figure 2.10 here.]

To see this, it is useful to study the effects that different variables have on the yield curve. Figure 2.10 plots the responses of yields to shocks in each element of the state vector. Panels on the left $(a1, a2, a3)$ display the reaction of the curve to perturbations of the Y_t variables; panels on the right $(b1, b2, b3)$—to shocks in V_t variables. In each subplot, the solid line depicts the yield curve when all variables are held at their unconditional means. Instead, the dashed and dotted lines plot the yield curve response when a given factor is set to its 10th and 90th percentile,

respectively. Let us for the moment focus on the impact of the yield curve factors. Several points emerge. The effect of the f_t state is most pronounced at the short end of the curve. A downward (upward) shift in f_t moves the short yield significantly below (above) its unconditional mean. The effect diminishes with the maturity, and is consistent with the filtered time-series pattern of f_t which closely tracks the Federal Funds rate (see Figure 2.9*c*). The two X_t factors influence the longer segment of the curve. X_2 is most active at intermediate maturities (two–three years), inducing changes in the curvature. X_1 acts predominantly at long maturities, thus changing the slope.[15] The identification of X_t states as those impacting on longer maturities concurs with the form of the market price of risk we adopt (see discussion in Section 2.7.C). Recall that our preferred specification chooses the X_t factors to be informative about risk compensations, which are most pronounced in the longer segment of the curve.

[Table 2.5 here.]

The roles of X_t factors are in line with the identified volatility states. Indeed, V_{11} generates the volatility at longer maturities, while V_{22} is responsible for capturing the shorter end of the volatility curve. To support this conclusion, Table 2.5 reports t-statistics obtained by regressing observed yield volatilities and covariances on the elements of the state vector. The significance of V_{11} and V_{22} factors for explaining the observed volatility curve has an opposite pattern: The effect of V_{22} is decreasing with the maturity, in contrast the effect V_{11} is increasing with the maturity. The out-of-diagonal element V_{12} is overwhelmingly significant, but has the largest impact on the yield covariance dynamics.

2.4.C Short and long-end volatility: two episodes

The split of volatility factors into the short- and long-end components is also evident in their filtered dynamics. The long-term factor V_{11} reflects major moves in the volatility of the ten-year yield. The short-term factor V_{22}, in turn, captures the two-year volatility. To illustrate this point, let us consider two salient moments in the volatility history (marked with circles in Figure 2.9). In the last months of 2003, the large move of the V_{11} factor mirrors a rapid increase in volatility that affected the long end of the curve, but was barely visible at its short end (panel *d*). The main trigger for the elevated volatility in this period came from the real part of the economy. In June,

[15] Since factors are latent, we do not discuss the direction of their impacts when at the tenth or 90th percentile. The particular sign is a consequence of the identifying normalization we impose. f_t factor is an exception, given its close resemblance to the Federal Funds rate.

the Fed lowered rates to 1% and was communicating no tightening for a foreseeable future. This anchored certainty about the short end of the curve. At the same time, the second half of 2003 welcomed first signs of recovery, but they continued to be mixed: Increases in retail sales and productivity gains mingled with still weak payroll figures. Investors were struggling to interpret the implications of these numbers for the long term, spurring a new thrust of volatility at the long segment of the curve.

Quite a different, but consistent, story is related to an earlier volatility episode in June 1995 highlighted in panel f of Figure 2.9. This moment, which marks the largest move in volatility of the two-year rate, is manifestly captured by the V_{22} dynamics. Yet, it does not appear more than a usual blip either in the other two factors, or in the volatility of longer yields. Accordingly, this time the trigger to the event had a short-end nature. In mid-1995 the monetary policy easing was imminent. The market was expecting that the Fed would take an aggressive pace to prop up the economy, as indicated by the BlueChip Economic Indicators survey. In May the expectations of the three-month T-bill rate plunged remarkably by over 1%. Though the Fed did ease in the end, it did so by much less then expected. This raised confusion at the short end of the curve, but left the long-end almost intact.

However incidental, the two volatility episodes expose well the roles of factors that move the volatility curve. In particular, they back our finding of long- versus short-run volatility components. To provide further support for this interpretation, in Section 2.5 we establish more rigorously the link between factors and observable economic quantities. Before doing so, however, we turn to exploring whether and how the volatility states influence the yield curve.

2.4.D Are volatility factors revealed by the yield curve?

To understand how our model answers this question, let us focus again on Figure 2.10. Its right-hand panels $(b1, b2, b3)$ plot the response of the yield curve to large perturbations in the volatility states. Relative to the strong reaction induced by the Y_t factors, the impact of the volatility states on the cross section of yields turns out much weaker. Even spectacular shifts in the V_{22} and V_{12} factors—recall that we consider their tenth and 90th percentile values—do not exert a visible effect on interest rates of any maturity. The impact of V_{11} is only revealed at maturities beyond five years, still it does not exceed a few (on average seven) basis points. This fact has an intuitive meaning: Abstracting from the potential volatility impact on the term premiums, we would expect the long-term volatility factor V_{11} to show up at longer maturities via the convexity effect. Since

this effect is mainly present in long yields, the response to V_{22} should be less pronounced, or even negligible.

To illustrate this effect over time, Figure 2.11 plots the model-implied yields $y_t^\tau = -\frac{1}{\tau}\{A(\tau) + B(\tau)'X_t + Tr[C(\tau)V_t]\}$ against hypothetical yields produced by removing V_t, i.e. $y_t^\tau = -\frac{1}{\tau}\{A(\tau) + B(\tau)'X_t\}$. The graph makes our earlier point more explicit: Volatility factors contribute a tiny margin to the dynamics of yields. The curve is virtually indistinguishable whether or not V_t factors are included. With $\tau \to 0$, the impact of V_t is excluded by construction of the short rate in Eq. (2.4). It turns out, however, that also in matching longer yields the model chooses not to employ the volatility states.

[Figure 2.11 here.]

Our estimates suggest that matrix of factors loadings, whose rows are given by $\left[-\frac{B(\tau_i)}{\tau_i}', -vech\left(\frac{C(\tau_i)}{\tau_i}\right)'\right]$, $i = 1, \ldots, 6$, is close to singular (condition number of 9000). It is thus hardly possible to back out a complete set of volatility states from yields. This is due to the negligible contribution of volatility factors to yields. It shows identical picture to what we see in Figure 2.10. The biggest contribution to yields comes from factor V_{11} for the long term yields which is mainly due to the convexity effect. However, also in this case the average contribution does not exceed mere 7 basis points.

2.4.E Filtered factors versus principal components

Finally, we would like to understand how our extracted factors relate to the usual yield curve metrics—the orthogonal principal components. How does their interdependence and the presence of V_t states impact the interpretation? Table 2.6 presents contemporaneous regressions of filtered states on the five principal components (PCs).

[Table 2.6 here.]

Each of our yield curve factors is strongly related to two PCs. Factor X_1 is linked to PC_1 (level) and PC_2 (slope) exactly as factor f is. But while X_1 loads with the same sign on the level and slope, factor f loads positively on level but negatively on the slope. Our intuition is that a positive shock to f—a positive monetary policy shock—is linked to a higher level but a lower slope due to, e.g., the decreased inflation expectations. Alternatively, one can interpret factors X_1 and f as two levels, a long- and a short-term, respectively. These levels respond to

economically distinct shocks (see Section 2.5) and are connected in a nontrivial way through the monetary policy transmission mechanism. Interestingly, support to our two-level interpretation of factors is given in a recent Wall Street Journal article by the ex-Fed Chairman Alan Greenspan, as he discusses the decoupling of the monetary policy from the long-term rates after the year 2000 (Greenspan, 2009). Factor X_2 is related to the level and to PC_3 (curvature) with the opposite signs, and captures the monetary transmission up to the medium term.

Our volatility states pick up tiny movements in the curve embedded in principal components of higher order (PC_4 and PC_5). But not only. V_t's are also related to the slope (PC_2). Empirical evidence in Section 2.1 helps explain this fact: Monetary easing in our sample, raising the slope of the term structure, coincides also with a pronounced boost of the volatility curve (see e.g., Figure 2.1). Therefore, and despite their very different roles for the yield curve, both volatility states and the slope happen to respond to a common factor. In terms of higher-order PCs, the long-end volatility V_{11} and the covariance term V_{12} load with large and significant coefficient on the PC_4. While more weakly linked to the filtered states, PC_5 is also significant across the board.

This section leads us to several conclusions. There appears to be little evidence that one and the same variable generates both the term structure of interest rates and of interest rate volatilities. In combination with the knowledge that three factors are needed to explain just yields, such discrepancy is a tall order for a three-factor model. Precisely for this reason, the USV restriction may come in handy as a tool for separating roles of factors in low-dimensional settings. Without such restriction, three-factor models face misspecification risk on the volatility modeling front. Consequently, more factors are needed, and just one more may not be enough. There is little support that volatility factors can be identified from the cross section of yields. Rather, help of extra information from interest rate options, instantaneous volatility proxies (like the ones we apply), and estimation methods (filtering) is required.

2.5 Yield volatility states and macroeconomic conditions

Table 2.7 gathers major moves in the yield and volatility curves from 2000 to 2004, and provides snapshots of the underlying economic circumstances. This interesting period covers the longest easing cycle in our sample, during which the Fed brought interest rates down to a (then) unprecedented level of just 1%. When the uncertainty about economic outcomes is high, the volatility curve shows motion. Indeed, at least part of the variation in interest rate risk should

be associated with the varying perceptions about key macro variables (e.g. Kim, 2007a). Since our volatility measures arise from active trading in liquid bonds, we would expect them to give a timely reflection of changing market expectations and fears. Fed officials watch this uncertainty to take their interest rate decisions. The transcripts from the FOMC meetings provide a good record of this point, e.g. "Uncertainty about key interactions in the economy is a good reason to wait" (Mr. Donald Kohn, October 3, 2000).[16] Our event Table 2.7 shows that the volatility curve can be at least as rich a source of information as yields themselves.

[Table 2.7 here.]

The current section takes a step to investigate the economic content of factors. We seek additional evidence for our interpretation of latent states using surveys, and relate the filtered dynamics to expectations and uncertainties about macro aggregates. We rely on surveys rather than realized numbers motivated by the notion that prices and state variables shall reflect economic prospects better than past events. Our monthly survey data are spliced from two sources: BlueChip Economic Indicators (BCEI) survey and BlueChip Financial Forecasts (BCFF). As evident in the transcripts of the FOMC meetings, these surveys are regularly used by policymakers at the Fed to read market expectations. Appendix B.1.D provides details about the survey data including their timing within month. We use responses of individual panelists to construct proxies for the consensus forecast and for the uncertainty. Each month, the consensus is computed as the median survey reply. The uncertainty is measured with the mean absolute deviation of individual forecasts.

2.5.A Expectations, uncertainties, and the duration of volatilities

While it seems natural that different regions of the yield curve are driven by different underlying forces, current discussion suggests that the same holds true for their volatilities. Therefore, it is tempting to understand which economic variables might stand behind those moves. Table 2.8 presents the regressions of factors on consensus and uncertainty proxies about macroeconomic quantities. For clarity, we include only significant loadings. The macro variables we use represent three standard domains: (*i*) real activity is captured by real GDP growth (RGDP), industrial production (IP), unemployment (UNEMPL), and housing starts (HOUST); (*ii*) the federal funds rate (FFR) describes the stance of the monetary policy, and (*iii*) inflation is reflected through

[16] See the FOMC transcript at `http://federalreserve.gov/monetarypolicy/files/FOMC20001003meeting.pdf`.

the CPI forecasts. The combination of expectations and uncertainty proxies constructed from these measures turns out to be overwhelmingly informative about our filtered states. We are able to explain up to about 90% and 50% of variation in the yield curve and volatility factors, respectively.

[Table 2.8 here.]

The picture emerging from those regressions has many angles, but some general observations can be made. The distribution of loadings among the yield curve factors confirms their distinct responses to the economic environment. The long-term yield factor X_1 is strongly driven by the expectations about the real sector, with unemployment emerging as its key mover. Being an intermediate state, X_2 straddles both the real and nominal effects and puts additional weight onto uncertainties. X_1 and X_2 load both on consensus and uncertainty about the real GDP. Still, the magnitude of their reactions to either component is different, and stronger for X_1. Finally, in addition to being almost completely explained by the monetary policy, the short-end factor f_t responds somewhat to the CPI inflation and the current stance of the real activity. The latter is timely captured by the industrial production. Not surprisingly, the monetary policy is significant for all yield curve states, but its impact diminishes with the state's maturity, telling us how the monetary impulse is transmitted to the longer segments of the curve.

Macroeconomic conditions are also reflected in the V_t states. V_{11} shows the strongest response to measures of real activity, and is pronouncedly influenced by the real GDP growth, the uncertainty about unemployment, and expectations on housing starts. Such configuration is consistent with its interpretation as the long-run volatility state. Quite different variables have explanatory content for the short-term volatility state, instead. V_{22} is linked to the uncertainty about inflation, industrial production and monetary policy. Notably, shocks to these variables are almost contemporaneously observed and rather short-lived, thus providing a contrast to more sticky metrics as the real GDP.[17] Indeed, variables inducing long- versus short-run volatility movements form virtually disjoint sets. Expectations of a monetary policy easing, for instance, spur an increase in V_{22}, but do not appear to have any visible effect on V_{11}. This gives a meaning to our finding in Section 2.1 that volatilities of shorter yields respond more to interest rate cuts

[17] Ranking inflation as short-lived may be surprising. This can be partially explained by the sample period we use (1992:01–2007:12). Looking at the term structure of inflation expectations provided by BCFF surveys, we notice that while short-run forecasts (for the current quarter) are surprisingly volatile, their volatility dies out very rapidly. Starting already from one-quarter-ahead, inflation expectations become extremely smooth. In his speech on January 3, 2004 at the American Economic Association meetings, Mr. Bernanke mentions this feature explicitly.

than do volatilities of longer yields: It is the short-run volatility component that provides for the asymmetry.

[Figure 2.12 here.]

To better understand the difference between the volatility states, in panels a through e of Figure 2.12 we plot their impulse-responses to various sources of macro uncertainty. The respective reactions differ in terms of magnitude and persistence as well as types of variables which are important. Uncertainties about FFR and inflation have an effect on short volatility which can last up to six months, but their impact on the long volatility remains negligible. This pattern reverts for the real GDP uncertainty, whose role persists at the long end but is only contemporaneously important at the short end. Interestingly, V_{11} has a significant impact on V_{22}, but not vice versa, as visible in panel f. The intuition for this result comes with a simple example. Imagine a situation in which uncertainty about the GDP growth picks up, and elevates V_{11}. The same news likely spurs uncertainty about near-term activity measures such as industrial production that affect V_{22} on the way. However, inverting the scenario and recognizing a higher duration of the long-term volatility, we do not necessarily expect a symmetric effect in the opposite direction to take place.

Admittedly, interactions between model-implied factors and macro variables deserve a deeper discussion than the one we venture here. A simple analysis is unlikely to reflect the complexity of lead-lag relations between different variables. While exploring these relations in depth is beyond the current scope of this paper, here we provide one example. Housing starts are significant across the board, and as a sole variable simultaneously appear in the short- and long-run volatility components. The key to understand this fact lies in the role that housing plays for the real economy and thus monetary policy in the US. The slowdown in the housing market is thought to have a significant effect on the real GDP growth via reductions in consumer spending (negative household wealth effect and decline in mortgage equity withdrawals, see e.g. Mishkin (2007)). In conditions of contained inflation as in our sample, the Fed is expected to counteract this negative development by lowering the interest rates. This double-edged sword effect of housing on the real economy and monetary policy shows that the apparently confusing presence of this variable in both V_{11} and V_{22} can actually be consistent with the overall interpretation of these factors.

Turning to the interpretation of the out-of-diagonal covariance factor, we find that—under our identification scheme[18]—V_{12} is highly positively linked to the realized correlation of the long-to-medium part of the yield curve (see Figure 2.2, panel *c*). Contingent on this observation, we can interpret its signs in survey regressions. V_{12} is associated with uncertainty measures about the real GDP, unemployment, housing starts, and monetary policy, all of which load with a negative sign. An increase in uncertainty about the macroeconomy implies a decline in V_{12}, and a lower correlation between the long and medium part of the yield curve. Intuitively, the different segments tend to move more independently when the curve changes shape. This, in turn, usually occurs in times of increased uncertainty such as April–June 2001, which is the only NBER-proclaimed recession in our sample (see our collection of major curve moves in Table 2.7). V_{12} captures precisely those moves.

2.6 Yield volatility and market-wide liquidity

Despite their links to macro variables, around half of volatility dynamics remains unexplained by fundamentals. We argue that yield volatility states share common variation with the risk factors that underlie on-the-run liquidity premia, price pressures or Treasury supply shocks. Small magnitudes and a relatively fast mean-reversion of these factors make it hard to extract them from low frequency yield data. The Treasury market in the US is considered one of the most liquid markets worldwide, therefore one would expect the deviations from the fundamental price to be short-lived and difficult to spot at low sampling frequencies. Realized covariance matrix of yields aggregates this high frequency variation; and our model serves as a tool that transforms it into a low number of factors.

We distinguish between two notions of liquidity. The first one, the funding liquidity, is linked to the tightness of financial conditions and credit availability in the economy; as such, it captures lower frequency movements in the market-wide liquidity. The second, a shorter term aspect, reflects the deviations from the fundamental price caused by investors' desire for immediacy. We document an interesting correlation pattern between those two aspects of liquidity and the volatility factors of different duration.

[18] Our latent states are subject to identification restrictions which preclude direct interpretation of factor signs (except for V_{11} and V_{22}, which are always positive). Recognizing this fact, we seek to establish factor signs that give their natural interpretation in terms of observable quantities.

Intuitively, one would expect the funding liquidity to share common features with the long-end volatility state, V_{11}. In our sample, this link runs through the monetary policy channel. High value of V_{11} is observed when distant prospects for the economy are weak, and surrounded by a high degree of uncertainty (see Table 2.8). In such circumstances, the Fed counteracts the worsening economic outlook by easing the monetary policy.[19] Lowering short-term interest rates improves the funding conditions for financial intermediaries thus increasing the liquidity (Jensen and Moorman, 2010). To reconcile our interpretation, we associate volatility factors with a measure of funding liquidity recently proposed by Fontaine and Garcia (2010).[20] As expected, the funding liquidity displays the strongest link to the long volatility state V_{11} which explains 45% of its variation. In line with our interpretation, the funding liquidity co-moves with the monetary policy (Panel A, Table 2.9): Survey-based expectations about the Fed funds rate capture 42% of its fluctuations.[21]

[Table 2.9 here.]

The short-term aspect of liquidity captures its transitory dry-ups. Consequently, a negative transient liquidity shock either dissipates or transforms into a more persistent funding liquidity squeeze as it was the case in the 2007–2009 crisis. Recall that at the end of 2007, only V_{22} increased rapidly while V_{11} initially remained at low levels. It stands to a reason that V_{22} captures part of the variation in short-term liquidity. An example of the transient liquidity variation is the noise illiquidity measure by Hu, Pan, and Wang (2010).[22] The factor captures temporary deviations from the efficient yield which is represented by the Nelson-Siegel-Svensson model estimates. Importantly, the transient illiquidity is positively linked to V_{22} which explains 21% of its contemporaneous variation (Panel B, Table 2.9). The positive sign is intuitive given

[19] We can think of monetary easing in broader terms than just lowering the Fed funds rate. For instance, large scale asset purchases conducted by the central bank fall into this category.

[20] Fontaine and Garcia (2010) filter the funding liquidity factor using on-the-run and and off-the run bond yields. We thank Jean-Sébastien Fontaine for sharing the data with us.

[21] In the regression of funding liquidity on the monetary policy expectations and V_{11}, both regressors are highly significant and the R^2 increases to 63%.

[22] Hu, Pan, and Wang (2010) define the noise illiquidity measure as the root mean squared yield pricing error on a given day. Every day, they estimate a Nelson-Siegel-Svensson model using the CRSP Treasuries dataset. Higher value of illiquidity signals the shortage of arbitrage capital which would otherwise align the prices of individual bonds. We follow their procedure to construct the short-term illiquidity measure.

that the value of the noise illiqidity measure is high in periods of high uncertainty and financial distress.[23]

To quantify the additional explanatory power of liquidity for rate volatility, we augment the regressions of volatility factors on surveys (Table 2.8) with the respective liquidity gauge. Including the funding liquidity measure into survey regression for V_{11} increases the R^2 by 6%; the noise illiquidity measure adds 7% in the V_{22} regression. We do not report the full regression results for brevity.

The results in this section provide a broader view on the unspanned factors in the yield curve. Our filtered volatility states aggregate information that is present in the on-the-run Treasury yield curve, but can be extracted only at higher frequencies. The model provides a structure for decomposing the multivariate interest rate risk according to its duration. Both, macro survey and liquidity regressions demonstrate that a multivariate and duration-based interpretation of volatility factors is economically intuitive, and required to describe the data.

2.7 Discussion and robustness

This section discusses the robustness and efficiency of the realized second moment estimator (2.1) that are critical for our results. We also address the market price of risk specification, and its selection approach we follow in estimation.

2.7.A Realized covariance matrix estimation

In case of asynchronous trading, the realized covariance matrix estimator defined in Eq. (2.1) can be biased toward zero (see e.g. Hayashi and Yoshida, 2005; Audrino and Corsi, 2007). The bias is to a large extent generated by the interpolation of non-synchronously traded assets, and its severity depends on the difference in liquidity of the assets considered.[24] Hayashi and Yoshida (2005, HY) propose a covariance estimator which corrects for the bias in (2.1). The estimator sums up all cross-products of returns which have an overlap in their time spans, and thus no data

[23] The strong link between noise illiquidity measure and realized volatility is preserved at daily frequency. A regression of daily illiquidity on realized volatility of a two-year bond gives an R^2 of 13% (Panel B, Table 2.9). Similar result is obtained from the regression of the noise illiquidity measure on Merrill Lynch Treasury volatility index (MOVE). MOVE index is constructed as a weighted average of implied volatilities from one month Treasury options with maturities two, five, ten and 30 years. In rough terms, it is the Treasury bond equivalent of the equity VIX index.

[24] Audrino and Corsi (2008) offer a thorough discussion of the bias in realized covariance.

is thrown away. The covariance of two bond yields reads:

$$RCov_{i,j}^{HY}(t,t+h;N_i,N_j)=\sum_{k=1}^{N_i}\sum_{l=1}^{N_j}\left(y_{t_i+\frac{kh}{N_i}}-y_{t_i+\frac{(k-1)h}{N_i}}\right)\left(y_{t_j+\frac{lh}{N_j}}-y_{t_j+\frac{(l-1)h}{N_j}}\right)\mathbb{I}\left(\tau_i\cap\tau_j\neq\emptyset\right) \quad (2.26)$$

where τ_i and τ_j denote the interval of the return on the first and second bond, respectively.

To verify the robustness of our realized covariance estimator (2.1), we implement the HY approach for the realized covariance of ten- and five-year bond. Both estimators deliver very similar results in terms of magnitude and covariance dynamics. They are highly correlated (90%) and the t-test for the difference in means does not reject the null that $\mu_{\text{outer}}=\mu_{HY}$ (p-val = 0.57).

There at least two reasons why we stick to the simple outer-product realized covariance estimator (2.1). For one, estimators of Hayashi and Yoshida (2005), Audrino and Corsi (2007) are not directly applicable in our case because for the construction of the zero curve we require a synchronized set of yield changes. More importantly, on-the-run Treasury bonds are largely homogenous in terms of liquidity, which is well proxied by the average number of quotes/trades per day reported in Table B.1 of Appendix B.1.B.

2.7.B Alternatives to the realized covolatility estimator

Instead of the realized second moments, one could resort to other volatility measures. To assess the gain from using high-frequency data, we consider two alternatives: (*i*) the DCC-GARCH model proposed by Engle and Sheppard (2001), and (*ii*) the realized covolatility computed from the daily data.

In a first step, we estimate the DCC-GARCH(1,1) specification on demeaned yield changes at daily, weekly and monthly frequency. Estimation results are available upon request. We see several reasons to prefer the realized estimator to the GARCH. Most of all, the realized covolatility has a richer economic content that is embedded in the intraday yield movements. The residuals from a regression of realized volatility on the GARCH volatility still carry significant economic information. Almost 30% of the variation in residuals can be explained by macroeconomic surveys. Second, the realized measure is nonparametric and thus not prone to a misspecification. Third, GARCH estimates cannot be easily aggregated to lower frequencies (Drost and Nijman, 1993). Based on monthly data, GARCH estimates are imprecise compared to the realized volatility. Thus, if using GARCH, our analysis in Section 2.5 would be less reliable. Finally, the DCC-

GARCH model has difficulties in matching the abrupt changes in correlation evident in the realized correlation estimates.

In a second step, we compare our realized covolatility measure to the one estimated on the daily data. This exercise gives us a view on the additional value coming from the intraday data. The daily estimator can explain a significant part (around 70%) of the variation in the realized covolatility. However, similar to above, the residuals from the regression of realized volatility measures on the volatility from the daily data are still informative. Up to 20% of their variation can be explained by the economic forecasts. The importance of including intraday observations increases with maturity.

Those comparisons show that the reason for using high-frequency data is beyond the standard statistical efficiency arguments. The incremental economic content of intraday observations is consistent with our interpretation of realized volatility as providing complementary piece of information towards yields themselves.

2.7.C Specification of bond premiums

Both economically and statistically, it is hard to identify all parameters of market prices of risk Λ_Y driving excess returns of bonds. Therefore, we reduce the number of parameters to be estimated to a necessary minimum. This section describes our selection procedure.

We write log excess holding period return from a simple strategy of buying an n-year bond at time t and selling it in one year, i.e. for weekly data at $t+52$, the strategy is financed by rolling over one year bond:

$$rx_{t+52}^{(n)} \equiv p_{t+52}^{(n-1)} - p_t^{(n)} - y_t^{(1)}, \tag{2.27}$$

where $p_t^{(n)}$ denotes log of bond price $P_t^{(n)}$, and n is expressed in years. To understand how excess returns are related to the yield curve, we regress the realized excess returns of two- to ten-year zero bonds defined in Eq. (2.27) on the vector of filtered states $(Y_t', vech(V_t)')'$:

$$rx_{t+52}^{(n)} = \alpha^{(n)} + \beta_Y^{(n)} Y_t + \beta_V^{(n)} vech(V_t) + \epsilon_{t+52}^{(n)}. \tag{2.28}$$

Because at this point we are agnostic about the structure of model-implied risk prices Λ_Y, we obtain factors for the above regression from a model estimated with zero market prices of risk for all shocks. There is a reason for keeping all states in the predictive regression: The six factors are in the information set of an investor at time t, i.e. having high-frequency bond data the investor

can filter all of them. While those regressions explain between 41% (two-year) and 33% (ten-year) of variation in realized excess returns, based on Hansen-Hodrick standard errors only X_1 and X_2 turn out statistically significant at the 1% level.

The part of bond excess return that can be predicted using the yield curve information has the following form:

$$E_t\left[rx_{t+52}^{(n)}\right] = \alpha^{(n)} + \beta_Y^{(n)} Y_t + \beta_V^{(n)} vech(V_t). \tag{2.29}$$

In a first step, we determine the main sources of the time-variation in excess returns. Similar to Cochrane and Piazzesi (2008), we stack expected excess returns across maturities and extract principal components. A single factor—the first PC—can explain about 98% of the return variation and its loading increases with the bond maturity. Importantly, around 80% of the movements in this dominant PC can be captured by X_1 and X_2. This finding has immediate implications for modeling premiums. In particular, the return-forecasting factor is largely spanned by our first two yield curve states.[25] Hence, we let X_1 and X_2 drive the variation in excess returns, both of which exhibit stochastic volatility and a nontrivial correlation.

In a second step, to establish which shocks are priced, we estimate a VAR system for the filtered yield curve factors Y_t, and compute covariances of factor shocks with the realized excess returns across maturities. Covariances of shocks to X_1 and X_2 with the realized excess returns are much larger than those of f, and in both cases rise with maturity: covariance of X_1 increases linearly and of X_2 has a hump-shape. This pattern suggests that shocks to factors X_1 and X_2 are priced. Therefore, from Eq. (3.5), we specify:

$$\Sigma_{Y,t}\Lambda_{Y,t} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_{11}^1 & 0 & 0 \\ 0 & \lambda_{22}^1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_{1,t} \\ X_{2,t} \\ f_t \end{pmatrix}. \tag{2.30}$$

In order to corroborate this choice, we turn again to regressions of yield factors on surveys (Section 2.5.A, Table 2.8). Yield factors are to a large degree explained by the expectations and uncertainties about the macro conditions. As such, their innovations should subsume shocks in those expectations and uncertainties. Factor f captures rather short-lived shocks stemming from industrial production and inflation. Instead, the likely candidates for impacting premiums on

[25] Recall that the states we filter out differ from the traditional PCs extracted from yields. Therefore, the explanatory power of X_t's for the return-forecasting factor is not in contradiction with the results of Cochrane and Piazzesi (2005, 2008), who show that their factor goes beyond the standard PCs.

longer-duration bonds in our sample—output growth, unemployment and housing market—enter the yield curve precisely through factors X_1 and X_2.

While with specification (2.30) we aim to provide a decent description of the time-varying portion of premiums, we remain moderately optimistic about its ability to capture the full economic content of expected excess returns. Several symptoms foster our caution. For instance, the residuals from regressions (2.28) of realized excess returns on yield curve factors are highly autocorrelated (0.96). This points to the existence of additional return predictors outside the yield and volatility factor span. Moreover, we find it possible to significantly increase the R^2's of those regressions by including the cross-products of yield curve factors. Unfortunately, such market prices of risk are hard to implement in our model without sacrificing its tractability. The convenient affine property of both the physical and risk neutral dynamics is potentially traded off for some misspecification on the side of term premiums. With our approach to selecting Λ_Y, we try to introduce a minimal misspecification while following the basic economic intuition, and ensuring stable parameter estimates. Indeed, including the unconstrained matrix Λ_Y can considerably hamper the optimization of the loss function (2.25), and thus distort the identification of latent states.

2.8 Conclusions

We study and model the joint behavior of the yield and the volatility curve. Understanding the interest rate volatility requires an additional effort as its dynamics is hidden from the perspective of the cross-section of yields. However, with 16 years' worth of tick-by-tick Treasury bond transaction data, we obtain a good and contemporaneous description of the evolution of yields and their realized covariances. Several facts emerge from the analysis of the realized series: The level, slope and curvature portfolios comove in a nontrivial way, so that an unconditional PCA neglects potentially important interactions between states forming the curve. Interest rate volatility curve does not move on a single factor, rather several states are needed to describe its dynamics. Finally, while the yield-volatility link is weak on average, we also show that it is time-varying and subject to the influence of the monetary policy.

The model disentangles clear long- and short-run volatility components which correspond to the respective yield curve states, and consequently, yield durations. An additional covolatility factor accommodates the interactions between the long and intermediate region of the curve. Neither of those states manifests itself in the cross section of yields: If at all, the most pronounced

impact comes from the long-end volatility, but it does not exceed a few basis points and is only visible at maturities of beyond five years.

We start by being agnostic about factor names, and only after having shown their explanatory content, we link them to the forward-looking economic quantities using surveys. Factors differ by shocks to which they respond. The short-run volatility reflects the impacts of changing expectations and uncertainties about short-lived, at least in our sample, economic variables (e.g. inflation, monetary policy, and industrial production). The longer-run volatility is associated with more sticky measures of macroeconomic conditions (e.g. real GDP growth, unemployment). The covolatility term is sensitive to uncertainties about the real activity and the Fed, all of which contribute negatively to the correlation between the middle-to-long segment of the yield curve.

Still, while yield curve states are well captured with the combination of surveys, a large portion (about half) of volatility remains unexplained from the perspective of the economic variables we consider. We identify different liquidity measures as significantly related with the variation in volatility factors. Interactions with the stock and other markets, that are beyond the scope of this study, are potential ingredients contributing to the unexplained part.

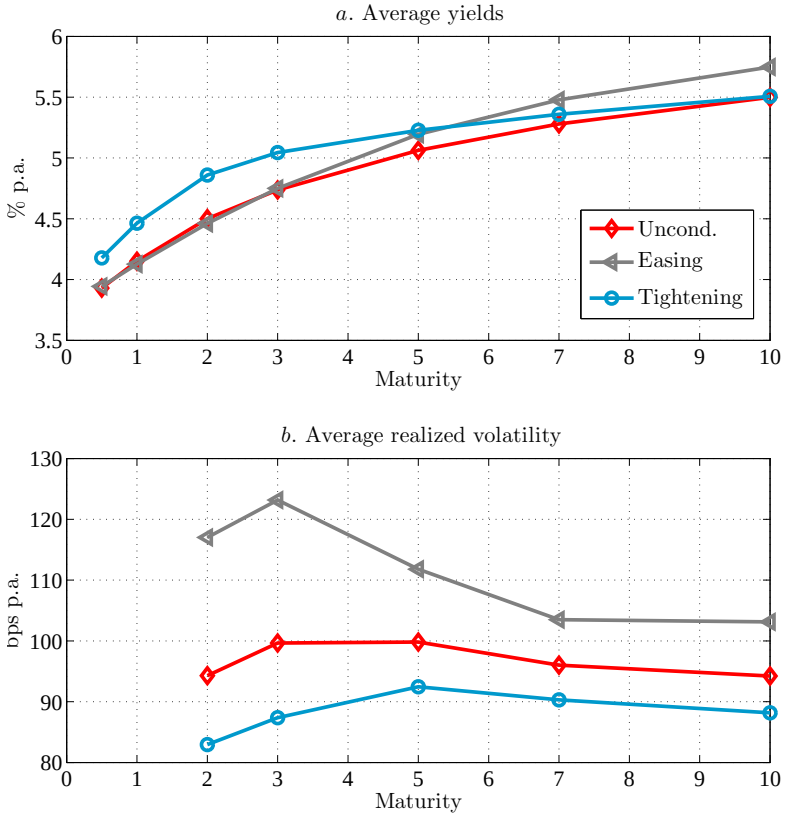


Figure 2.1: Yield and volatility curves

The figure plots the yield and volatility curves: unconditional mean for the whole sample 1992:01–2007:12, and means conditional on the Fed's tightening and easing cycles. The cycles are regarded as easing or tightening if at least three subsequent moves in the federal funds rate target have been in the respective direction. In our sample, out of total 3955 days, we identify 646 days as the easing regime, and 958 days as the tightening regime. The realized volatility curves are computed from the daily data and annualized (×250).

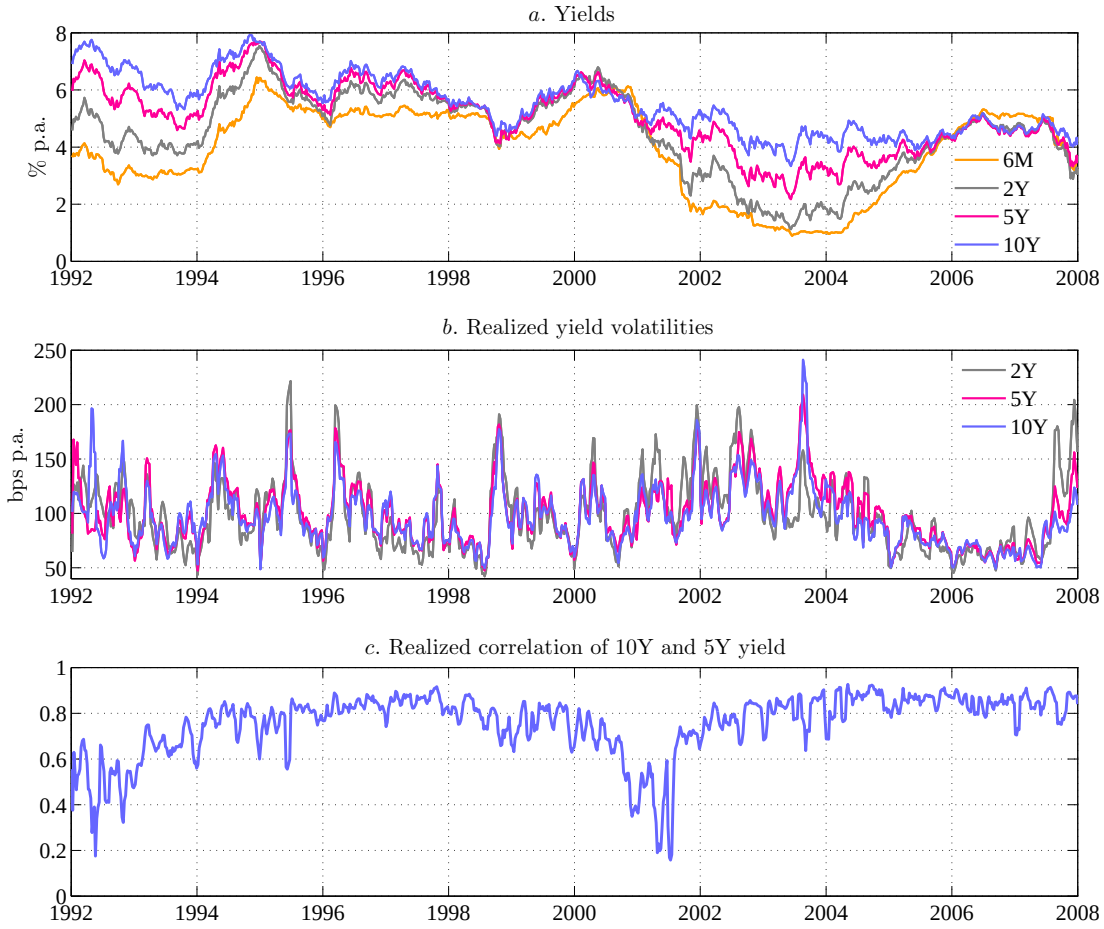


Figure 2.2: Evolution of yields and volatilities

The figure plots the dynamics of weekly yields, realized volatilities and correlation over the 1992:01–2007:12 period. Yields include maturities of six months and two, five, and ten years. Realized volatilities are constructed from the actively traded bonds of two, five and ten years to maturity.

Figure 2.3: Realized correlations of yield factors

The figure plots realized correlations between yield curve factors (left-hand axis) constructed from the high-frequency zero curve. Level, slope and curvature factor loadings are obtained from the PCA decomposition of the unconditional covariance matrix of yields with maturities: two, three, five and ten years. Daily realized correlations are smoothed over a three-month window. In each panel, correlation dynamics are juxtaposed against the Fed regimes: cut, no change, hike (right-hand axis).

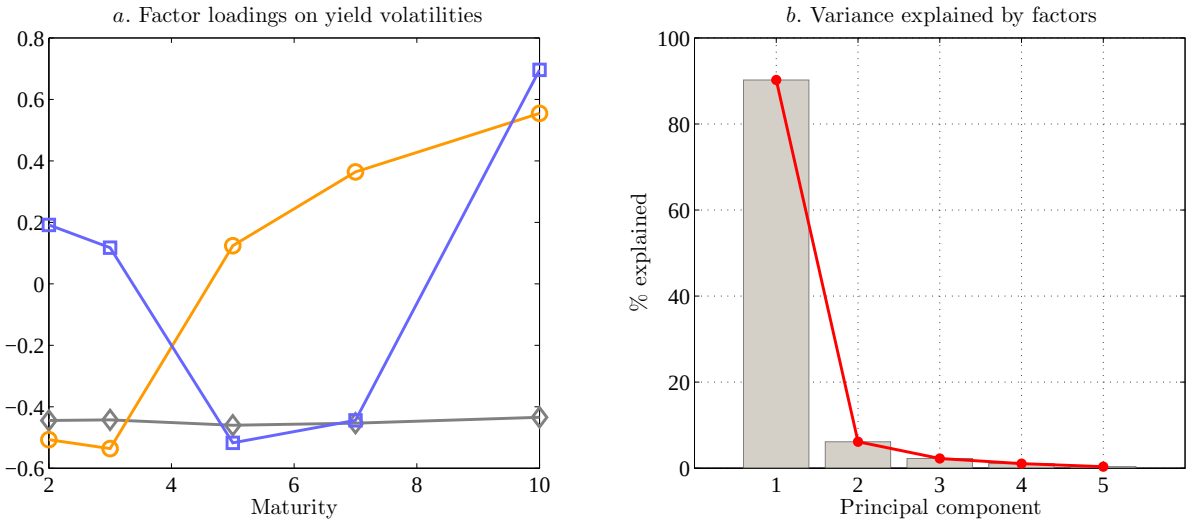


Figure 2.4: Factors in yield volatilities

The figure shows the principal component decomposition of the yield volatility curve. We use the unconditional correlation matrix of realized weekly volatilities computed for two-, three-, five-, seven- and ten-year zero bonds. Panel *a* plots the loadings of factors on volatilities. Panel *b* displays the percentage of variance explained by each factor.

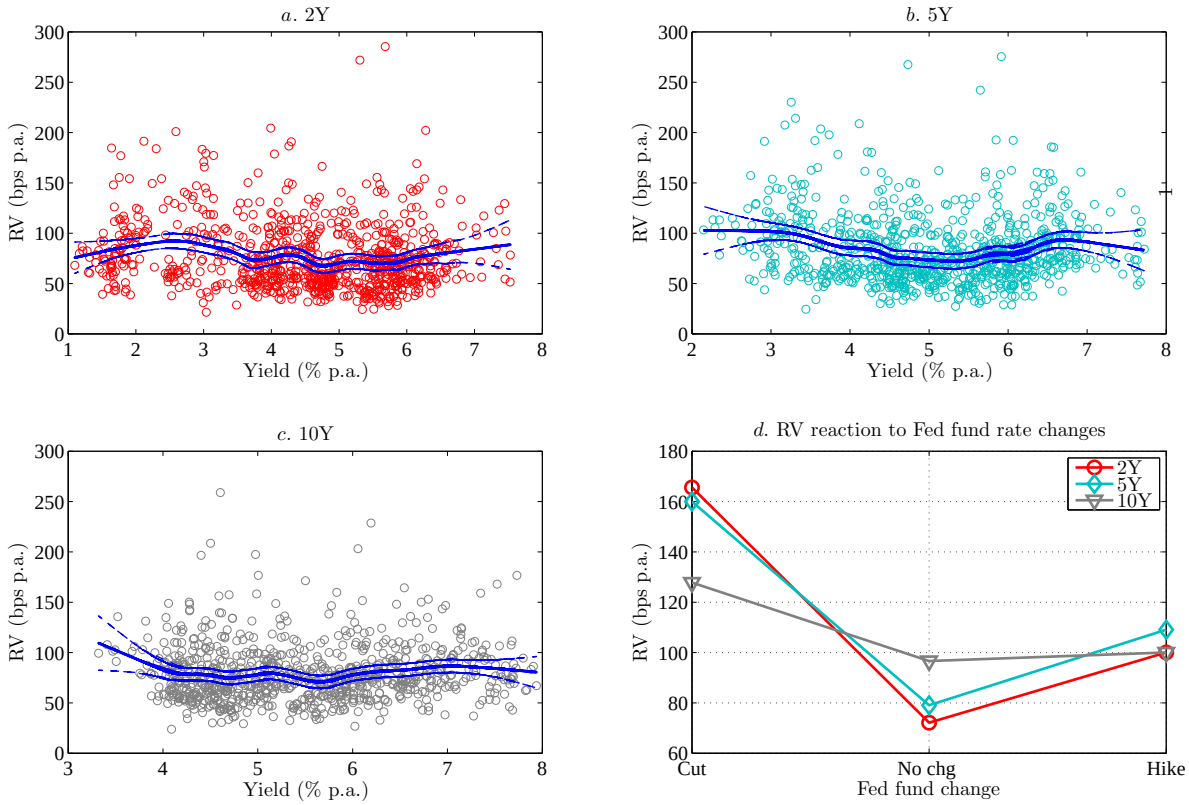


Figure 2.5: Yield level versus yield volatility

The figure depicts the relationship between yield level and yield volatility for bonds with two-, five- and ten-year maturity (based on weekly data). In panels *a*–*c*, circles denote data points, and the lines represent the fits of the nonparametric kernel regression together with the 99% confidence bound. Panel *d* shows the level of the realized volatility conditional on the change in the Fed funds target rate (cut, no change, hike). During our sample period 1992:02–2007:12 there were 23 rate cuts, 31 hikes, and 3901 days on which the rate did not change.

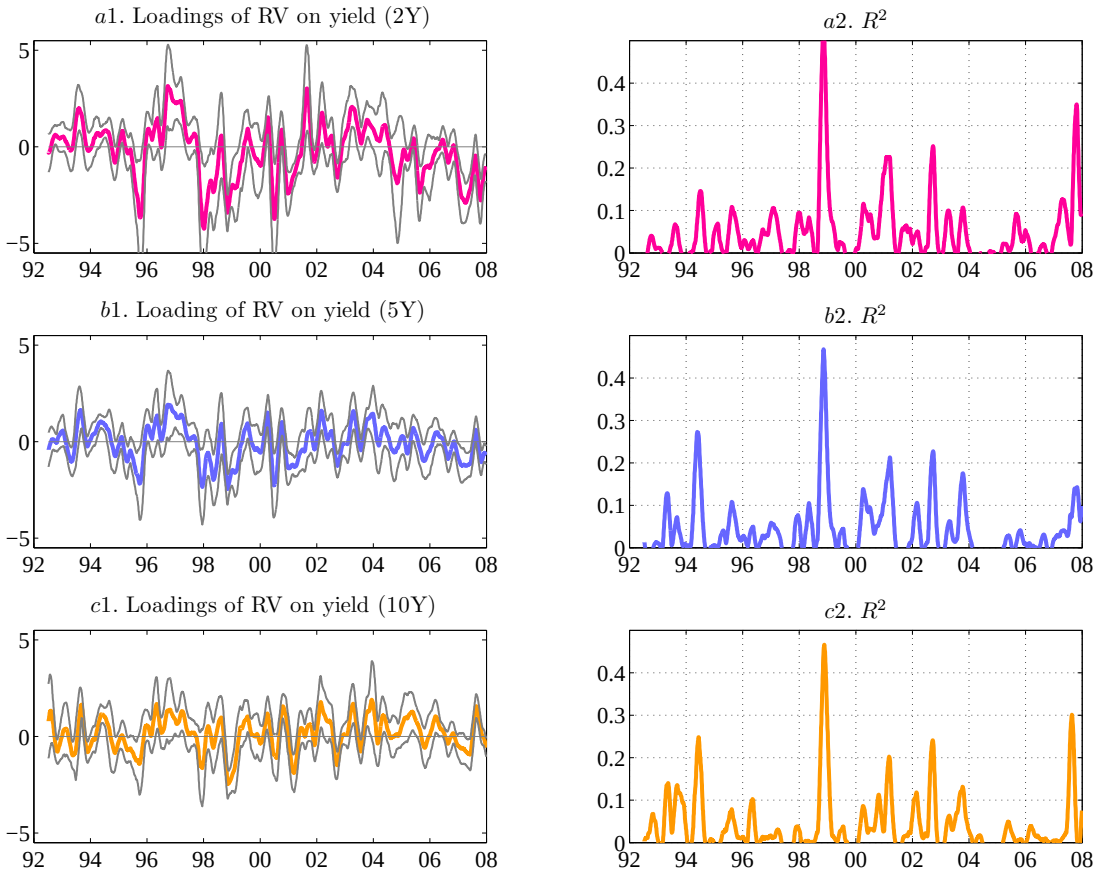


Figure 2.6: Regressions of realized volatility on the yield level

The figure plots the rolling regression coefficients, 90% confidence bands (left panels) and R^2's (right panels) in regressions of realized volatilities on yields of corresponding maturities. The regressions use daily data on yields and realized volatilities rolled over the window of 3 months. For clarity, all regression statistics have additionally been smoothed using a window of 50 observations. All variables have been standardized.

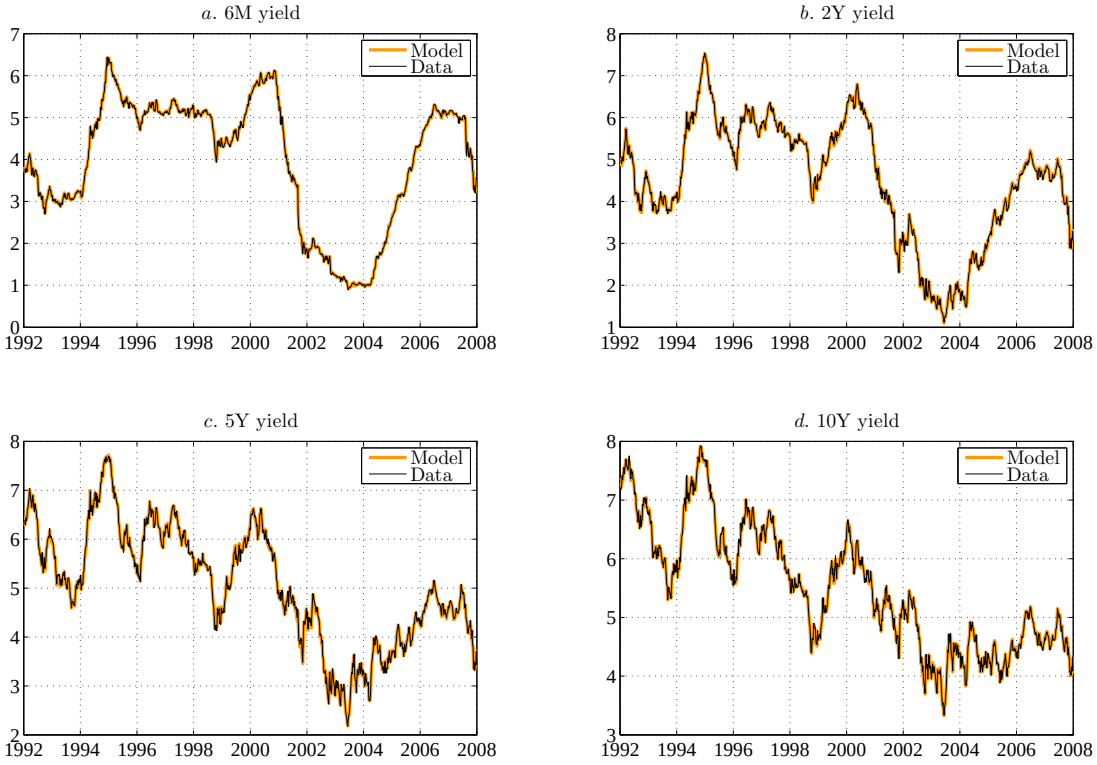


Figure 2.7: In-sample fit to yields

The figure compares the in-sample fit of the model-implied yields to the observed yields with maturities: six months, two, five, and ten years over the period 1992:01–2007:12. Yields are presented in percentage p.a., and are measured on weekly frequency.

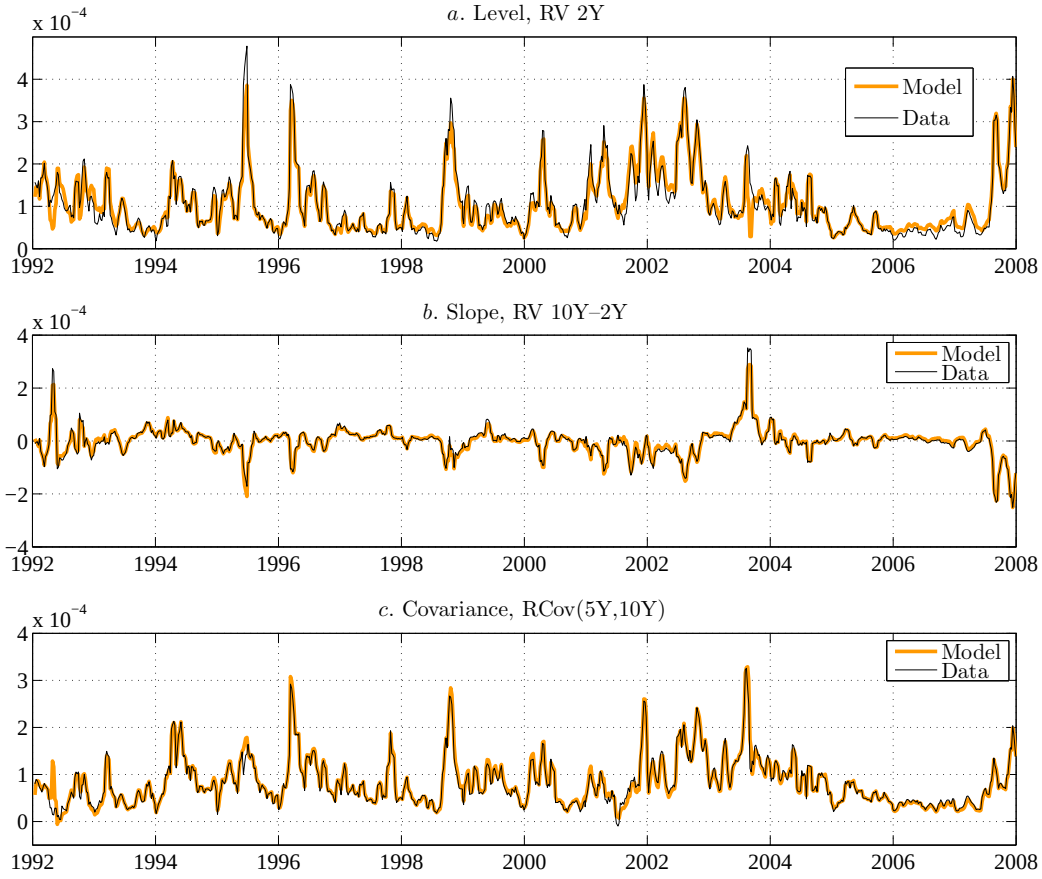


Figure 2.8: In-sample fit to the second moment dynamics of yields

The figure plots the model-implied in-sample fit to the yield second moments. The level is defined as the variance of the two-year yield. The slope is computed as the difference between the realized variance of the ten- and the two-year yield. Finally, the covariance is between the five- and the ten-year yield.

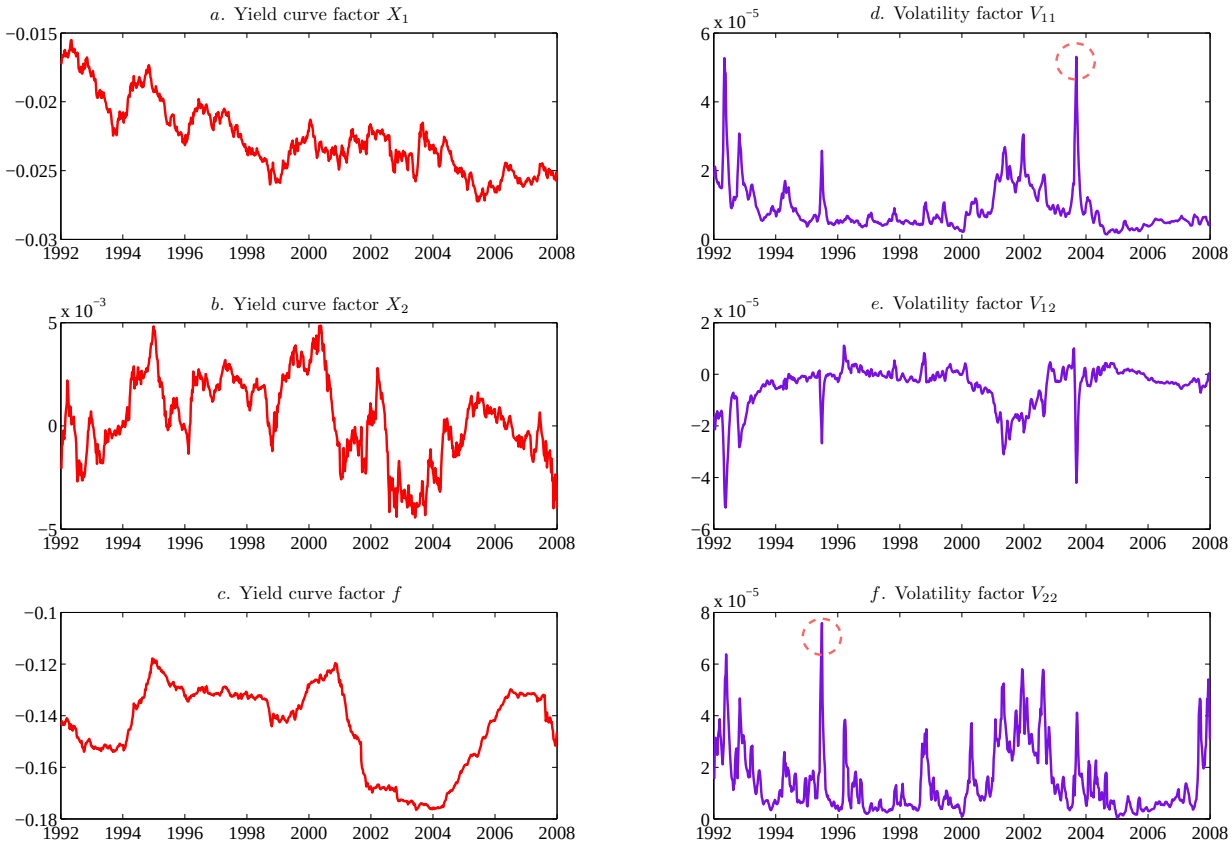


Figure 2.9: Filtered factor dynamics

The figure plots the model-implied factor dynamics extracted by the unscented Kalman filter under the estimated parameters of the G_3SV_3 model. The left-hand panels present factors driving the yield curve. The right-hand panels display factors generating time-variation in conditional second moments. The circles in panels d and f mark examples of important events that had a strong impact on the evolution of the volatility curve in our sample period. We discuss them in Section 2.4.B. For presentation, the X states have been multiplied by -1.

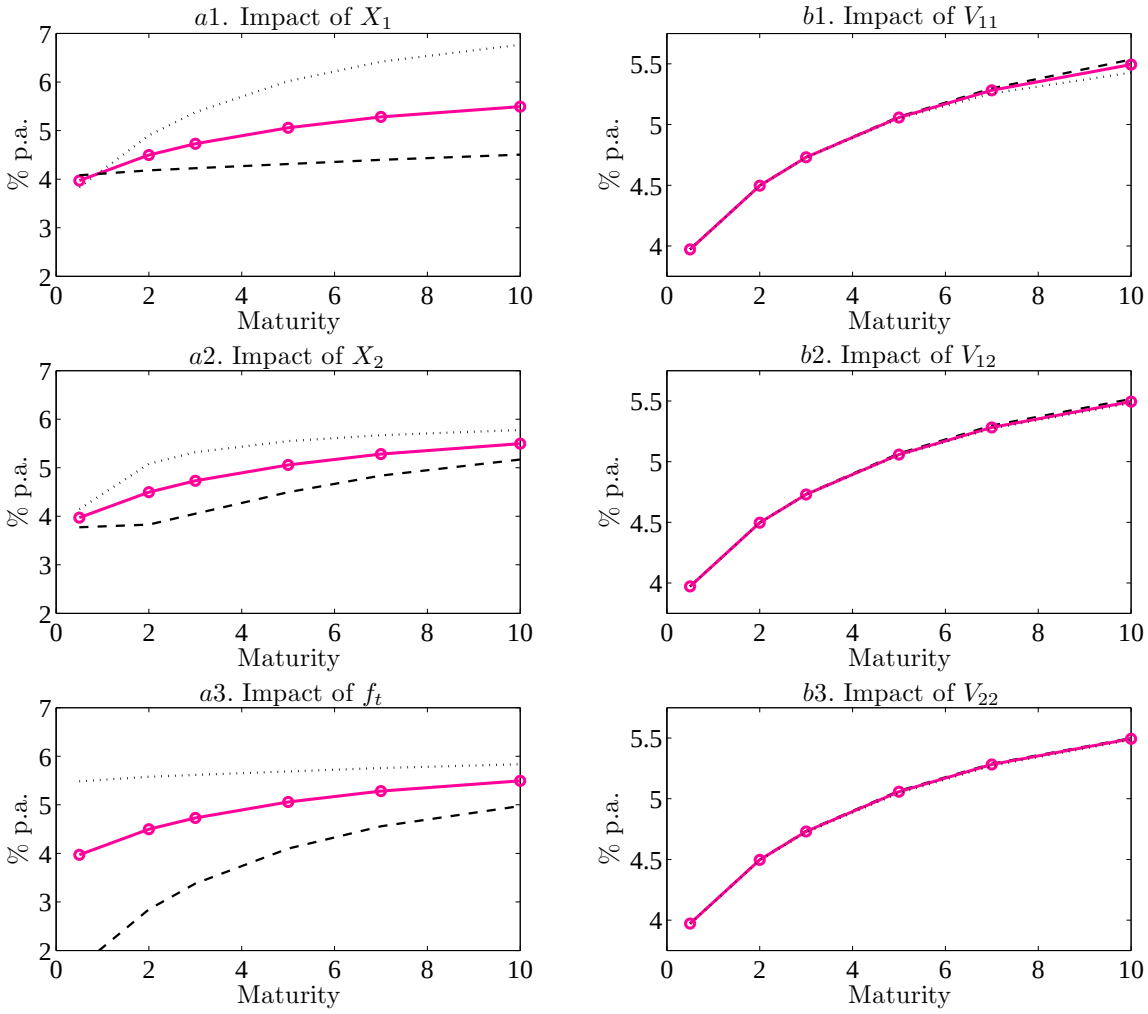


Figure 2.10: Factor shocks and yield curve responses

The graph presents the response of the yield curve to shocks in the state variables. The left-hand panels $(a1, a2, a3)$ display the effect of the yield curve factors $Y_t = (X_{1t}, X_{2t}, f_t)'$, the panels on the right $(b1, b2, b3)$ display the effects of the volatility factors $\bar{V}_t = (V_{11t}, V_{12t}, V_{22t})'$. In each subplot, the solid line shows the yield curve generated by setting all state variables to their unconditional means. Circles indicate the maturities used in estimation, i.e. six months, two, three, five, seven, and ten years. The dashed and dotted lines are obtained by setting a given state variable to its tenth and 90th percentile, respectively, and holding the remaining factors at their unconditional average. For presentation, factors X_1, X_2 have been multiplied by -1 so that both correlate positively with yields.

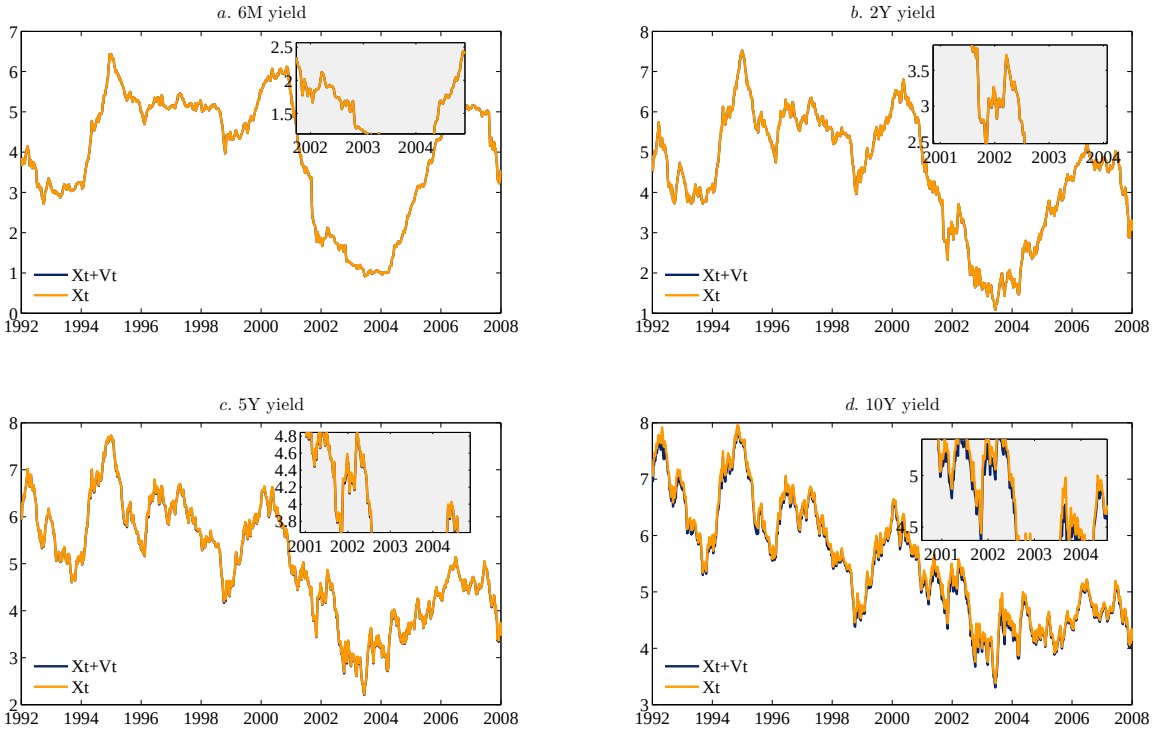


Figure 2.11: Contribution of X_t and V_t to model-implied yields

The figure plots the model-implied yields for maturities of six months and two, five and ten years. Each panel compares hypothetical yields that would be generated by X_t factors after V_t factors have been excluded ($y_t^\tau = -\frac{1}{\tau}\{A(\tau) + B(\tau)'X_t\}$) against those produced by X_t and V_t jointly ($y_t^\tau = -\frac{1}{\tau}\{A(\tau) + B(\tau)'X_t + Tr[C(\tau)V_t]\}$). For the ease of comparison, we magnify selected regions of each graph.

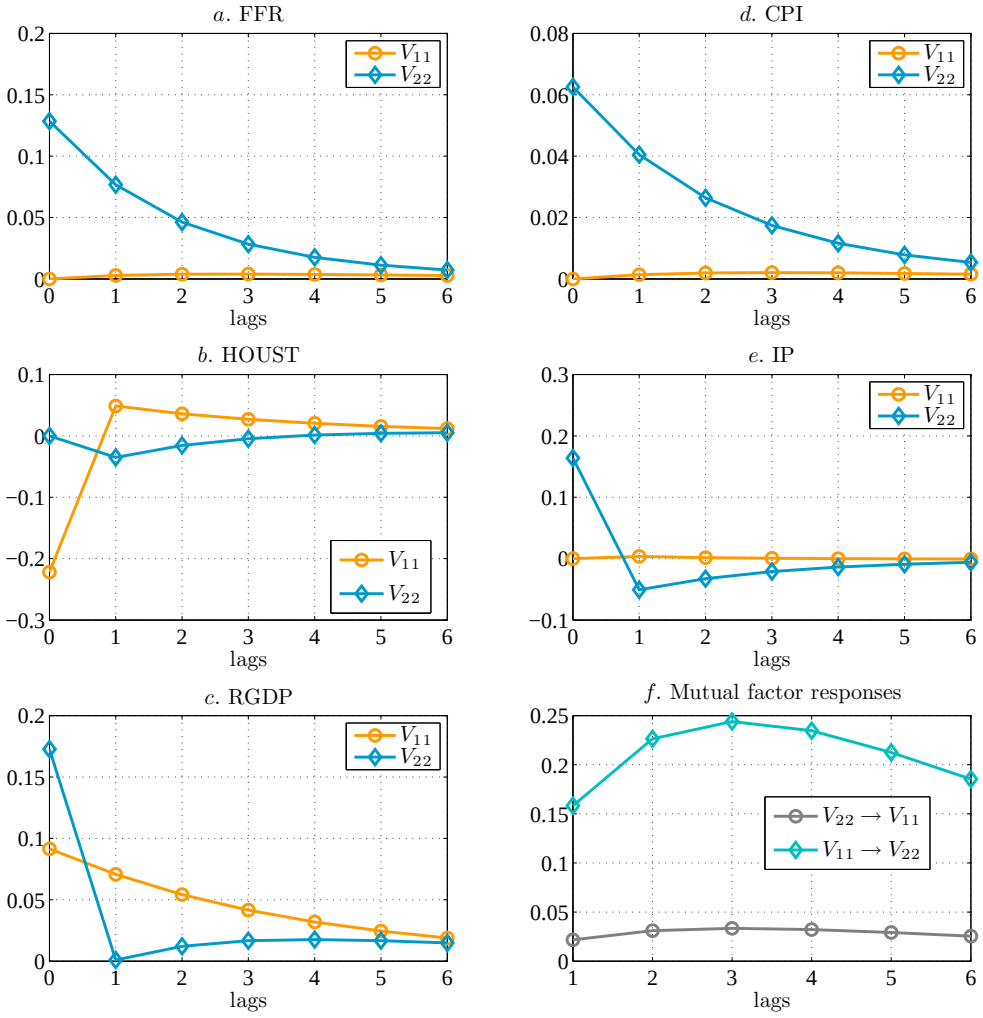


Figure 2.12: Impulse-response functions of volatility factors V_{11} and V_{22} to macroeconomic uncertainty proxies

Panels *a*–*e* display the impulse-response functions of volatility factors to macroeconomic uncertainties extracted from the surveys. Uncertainty is proxied with the mean absolute deviation of forecasts made by individual panelists. Panel *f* shows mutual responses of factors to each other. The impulse responses are based on a VARMAX model with first order AR component, allowing for contemporaneous and lagged effects of exogenous macro variables. The impulse responses exclude insignificant exogenous regressors. All variables have been standardized; lags are in months.

Table 2.1: Descriptive statistics of weekly yields and realized volatilities

The table contains descriptive statistics of weekly yields (panel *a*) and realized yield volatilities (panel *b*) based on the period 1992:01–2007:12. JB denotes the Jarque-Bera normality test (critical value 5.91). Panel *c* shows unconditional correlations between yields and volatilities. Numbers in italics indicate correlations which are not significant at the 1% level.

Panel *a*. Yields (% p.a.)

	6M	2Y	3Y	5Y	7Y	10Y
Mean	3.97	4.50	4.73	5.06	5.28	5.50
Skew	-0.60	-0.46	-0.35	-0.12	0.09	0.31
Kurt	2.15	2.46	2.47	2.30	2.17	2.13
JB	76.22	40.62	27.69	19.08	25.21	40.03

Panel *b*. Realized yield volatilities (bps p.a.)

	2Y	3Y	5Y	7Y	10Y
Mean	108.33	114.44	113.31	108.35	104.62
Stdev	46.73	49.75	45.87	40.75	38.52
Skew	1.47	1.57	1.40	1.18	1.72
Kurt	6.41	7.23	5.98	4.78	9.17
JB	712.90	977.33	589.98	309.12	1755.34

Panel *c*. Unconditional correlations of yields and realized volatilities

	y_t^{6M}	y_t^{2Y}	y_t^{3Y}	y_t^{5Y}	y_t^{7Y}	y_t^{10Y}	v_t^{2Y}	v_t^{3Y}	v_t^{5Y}	v_t^{7Y}	v_t^{10Y}
y_t^{6M}	1										
y_t^{2Y}	0.94	1									
y_t^{3Y}	0.89	0.99	1								
y_t^{5Y}	0.77	0.93	0.97	1							
y_t^{7Y}	0.66	0.85	0.91	0.99	1						
y_t^{10Y}	0.54	0.76	0.84	0.95	0.99	1					
v_t^{2Y}	-0.21	-0.18	-0.15	*-0.10*	*-0.05*	*-0.01*	1				
v_t^{3Y}	-0.24	-0.18	-0.15	*-0.07*	*-0.01*	*0.05*	0.96	1			
v_t^{5Y}	-0.27	-0.22	-0.20	-0.14	*-0.09*	*-0.04*	0.90	0.89	1		
v_t^{7Y}	-0.29	-0.23	-0.20	-0.13	*-0.08*	*-0.02*	0.85	0.86	0.97	1	
v_t^{10Y}	-0.26	-0.19	-0.15	*-0.08*	*-0.02*	*0.03*	0.77	0.78	0.86	0.89	1

Table 2.2: Parameter estimates

This table reports parameter estimates for four model specifications. "G_3SV_0 risk-neutral" is a Gaussian three-factor model without risk premiums. "G_3SV_0" denotes a Gaussian three-factor model with essentially affine market prices of risk. "G_3SV_3 risk-neutral" is our six-factor model without risk premiums with three conditionally Gaussian factors for the yield curve and three volatility factors. Model specification "G_3SV_3" additionally contains affine risk compensations. The purely Gaussian specifications follow Duffee (2002), and are diffusion-normalized according to the identification scheme of Dai and Singleton (2000). This type of normalization allows to treat γ_Y as a free parameter vector. BHHH standard errors are in parentheses. The last section of the table shows the log-likelihood values. For comparison, in models with stochastic volatility, we split the log-likelihood values into the yield component and the volatility component: Loglik y_t^τ and Loglik v_t^τ.

	G_3SV_0 risk–neutral	G_3SV_0	G_3SV_3 risk–neutral	G_3SV_3
$\mathcal{K}_{Y,11}$	-0.001	-0.506	-0.014	-0.033
	(0.001)	(0.289)	(0.001)	(0.008)
$\mathcal{K}_{Y,21}$	-0.181	-1.653	–	–
	(0.151)	(0.077)	–	–
$\mathcal{K}_{Y,22}$	-0.786	-0.434	-1.233	-1.947
	(0.266)	(0.103)	(0.032)	(0.309)
$\mathcal{K}_{Y,31}$	-1.696	-2.900	-2.797	-2.768
	(0.115)	(0.116)	(0.182)	(0.175)
$\mathcal{K}_{Y,32}$	-2.828	-0.800	-7.275	-7.668
	(0.096)	(0.034)	(0.433)	(0.479)
$\mathcal{K}_{Y,33}$	-0.786	-0.021	-0.500	-0.498
	(0.267)	(0.001)	(0.009)	(0.008)
σ_f	–	–	0.006	0.006
	–	–	(6.96e-5)	(7.05e-5)
k	–	–	2.000	2.000
	–	–	(1.335)	(1.067)
Q_{11}	–	–	0.001	0.001
	–	–	(1.14e-4)	(1.14e-4)
Q_{22}	–	–	0.004	0.003
	–	–	(2.68e-4)	(2.63e-4)
M_{11}	–	–	-0.069	-0.113
	–	–	(0.044)	(0.038)
M_{21}	–	–	-0.260	-0.185
	–	–	(0.184)	(0.177)
M_{22}	–	–	-0.518	-0.685
	–	–	(0.090)	(0.127)
γ_0	1.306	0.118	0.172	0.159
	(1.121)	(0.004)	(0.038)	(0.026)
γ_{11}	–	0.002	–	–
	–	(2.58e-4)	–	–
γ_{12}	0.001	0.003	–	–
	(5.64e-4)	(2.24e-4)	–	–
γ_{13}	0.006	0.006	–	–
	(1.18e-4)	(1.42e-4)	–	–
$\lambda^1_{Y,11}$	–	0.629	–	-0.018
	–	(0.290)	–	(0.020)
$\lambda^1_{Y,22}$	–	0.150	–	-0.703
	–	(0.101)	–	(0.310)
σ_y^2	8.34e-8	8.01e-8	8.02e-8	7.84e-8
	(1.42e-9)	(1.33e-9)	(1.44e-9)	(1.40e-9)
Loglik y_t^τ	35443	35530	35315	35341
Loglik v_t^τ	–	–	26264	26260

Table 2.3: In-sample model fit

This table gives the in-sample fit for four model specifications presented in Table 2.2. The performance of our model (G_3SV_3) in matching yields is compared with the purely Gaussian setting (G_3SV_0). Additionally, we report its ability to match the volatility dynamics. In panels $a1$ and $b1$, we report root mean squared error (RMSE) in bps per annum. The fit of volatilities is also stated as RMSE. To obtain easily comparable numbers, we convert the covariance matrix of yields into covolatilities: Since the slope and covariance can be negative, we take the square root of their absolute values. This serves as an input for computing the RMSE for the term structure of volatilities. Panels $a2$ and $b2$ show the percentage of variation in observed yields and volatilities explained by the respective model-implied quantities. The numbers represent the R^2's from the regression of the observed variable on the fitted one.

	G_3SV_0 risk-neutral	G_3SV_0	G_3SV_3 risk-neutral	G_3SV_3
Panel a. Term structure of yields				
$a1$. RMSE in bps				
6 months	1.3	1.3	1.3	1.5
2 year	2.4	2.3	2.3	2.2
3 year	1.6	1.6	1.5	1.6
5 year	2.9	2.8	2.7	2.6
7 year	1.8	1.7	1.7	1.7
10 year	2.8	2.8	2.5	2.5
$a2$. Explained variation (%)				
6 months	99.99	99.99	99.99	99.99
2 year	99.97	99.97	99.98	99.98
3 year	99.99	99.99	99.99	99.99
5 year	99.94	99.94	99.95	99.97
7 year	99.97	99.98	99.98	99.99
10 year	99.93	99.93	99.94	99.96
Panel b. Term structure of volatilities				
$b1$. RMSE in bps				
v_t^{2Y}	–	–	6.1	5.8
$v_t^{10Y} - v_t^{2Y}$	–	–	6.1	5.8
$v_t^{5Y,10Y}$	–	–	4.7	4.7
$b2$. Explained variation (%)				
v_t^{2Y}	–	–	95.92	96.22
$v_t^{10Y} - v_t^{2Y}$	–	–	98.23	98.30
$v_t^{5Y,10Y}$	–	–	96.92	96.89

Table 2.4: Out-of-sample yield forecasting

This table reports RMSEs in bps for forecasting horizons $h = 1, 4, 12$ and 52 weeks. The models are estimated using weekly data from 1992:01 through 2004:12. Forecasts are performed over the period 2005:01 through 2007:12.

Panel *a*. Yield forecasting

Yield Maturity	Forecast Horizon (h weeks)	Random Walk	G_3SV_0	G_3SV_3
6 months	1	9.29	13.54	13.62
2 years	1	10.93	14.89	15.10
3 years	1	11.36	14.80	14.81
5 years	1	10.77	14.13	14.30
7 years	1	9.98	13.56	13.67
10 years	1	9.17	13.11	13.59
6 months	4	21.16	23.03	22.81
2 years	4	22.74	26.99	27.16
3 years	4	23.27	27.04	27.08
5 years	4	22.37	25.63	25.85
7 years	4	20.88	24.11	24.38
10 years	4	19.08	22.27	23.23
6 months	12	43.18	39.69	38.66
2 years	12	43.47	43.16	43.33
3 years	12	42.40	42.37	42.58
5 years	12	38.28	38.99	39.80
7 years	12	34.64	36.12	37.53
10 years	12	31.05	33.19	36.21
6 months	52	146.56	89.37	82.77
2 years	52	110.02	70.27	67.87
3 years	52	92.73	63.88	63.57
5 years	52	66.82	53.43	59.00
7 years	52	53.12	47.10	57.99
10 years	52	44.64	41.74	60.48

Panel *b*. Volatility forecasting

Volatility Maturity	Forecast Horizon (h weeks)	Random Walk	G_3SV_0	G_3SV_3
2 years	1	11.28	n.a.	13.62
5 years	1	11.75	n.a.	10.96
10 years	1	9.76	n.a.	9.71
2 years	4	25.81	n.a	22.33
5 years	4	26.12	n.a.	16.47
10 years	4	22.78	n.a.	15.19
2 years	12	33.14	n.a.	29.08
5 years	12	34.13	n.a.	26.35
10 years	12	30.90	n.a.	26.41

Table 2.5: Regressions of yield realized second moments on filtered states

This table reports the t-statistics and adjusted R^2's obtained by regressing the yield realized variances (first five columns) and covariances (last three columns) on the latent factors extracted from our model, i.e. $X_1, X_2, f, V_{11}, V_{12}, V_{22}$. The sample covers the 1992:01–2007:12 period, which amounts to 845 weekly observations. t-statistics are corrected using the Newey-West adjustment with 12 lags.

RV & RCov	2Y	3Y	5Y	7Y	10Y	(5Y,2Y)	(10Y,2Y)	(10Y,5Y)
const.	2.53	2.86	-0.08	1.45	1.32	2.52	-0.10	-0.03
X_1	1.27	-2.56	1.03	-1.17	1.44	0.52	0.73	-2.51
X_2	1.25	1.94	1.12	2.45	1.16	1.32	-2.22	0.17
f	3.83	1.25	0.68	0.61	3.75	2.75	-0.13	-2.74
V_{11}	5.81	3.70	7.74	9.06	19.48	13.76	12.90	17.46
V_{12}	28.52	8.58	19.50	24.96	29.54	34.95	27.87	68.45
V_{22}	43.01	29.89	12.67	11.74	7.43	25.12	16.29	25.78
Adj. R^2	0.97	0.88	0.87	0.90	0.93	0.94	0.93	0.97

Table 2.6: Regressions of filtered states on yield curve principal components

This table reports regressions of model-implied factors on principal components extracted from the yield curve. The sample covers the 1992:01–2007:12 period, which amounts to 845 weekly observations. Standard errors are corrected using Newey-West adjustment with 12 lags and are reported in parentheses. Both left- and right-hand side variables are standardized to make the regression coefficients directly comparable. X_1 and X_2 have been multiplied by -1.

Regressors	X_1	X_2	f	V_{11}	V_{12}	V_{22}
PC_1	0.544	0.784	0.875	-0.102	0.013	-0.138
	(0.003)	(0.004)	(0.001)	(0.070)	(0.068)	(0.087)
PC_2	0.818	-0.124	-0.458	0.533	-0.397	0.358
	(0.004)	(0.005)	(0.001)	(0.080)	(0.085)	(0.079)
PC_3	0.184	-0.604	0.155	0.158	-0.233	0.184
	(0.003)	(0.005)	(0.001)	(0.061)	(0.057)	(0.086)
PC_4	-0.002	0.028	-0.012	-0.345	0.463	-0.205
	(0.003)	(0.004)	(0.001)	(0.062)	(0.064)	(0.082)
PC_5	-0.012	0.013	0.006	-0.182	0.151	-0.109
	(0.003)	(0.005)	(0.001)	(0.060)	(0.057)	(0.089)
Adj. R^2	1.00	1.00	1.00	0.47	0.45	0.23

Table 2.7: Major moves in the U.S. Treasury yield and volatility curves in the period 2000–2004 (1st part)

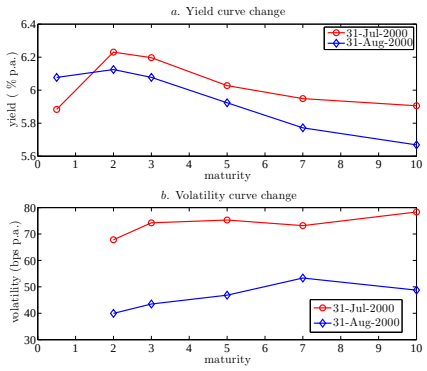
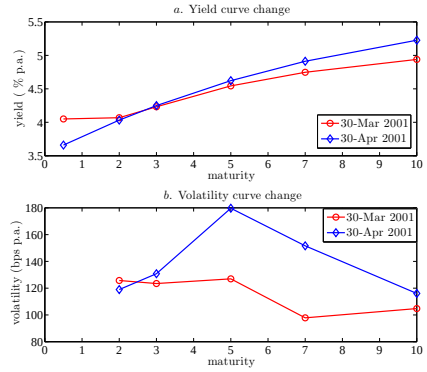
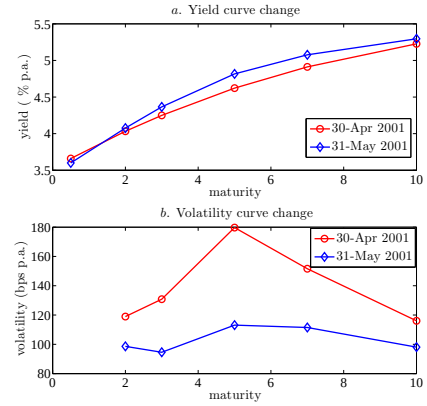
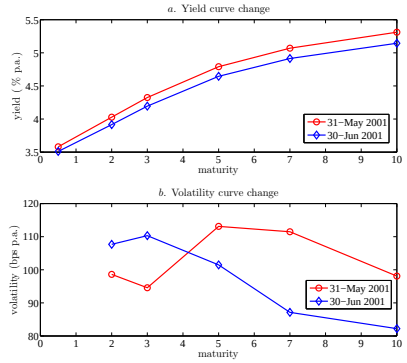
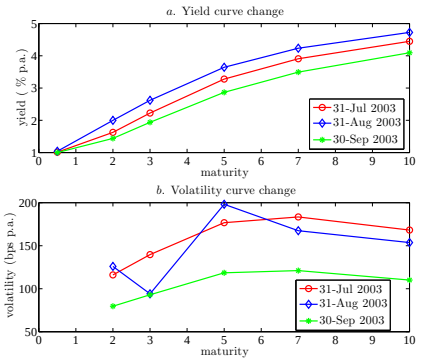
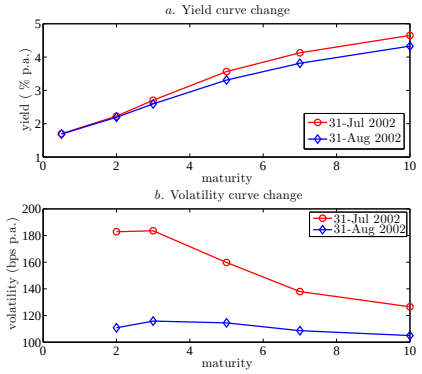
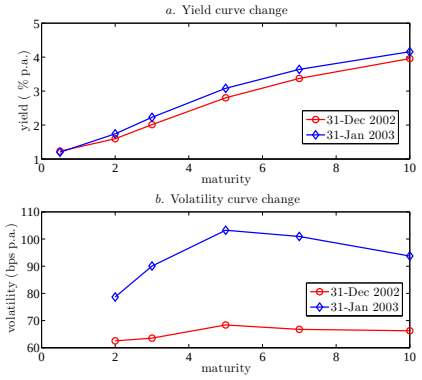
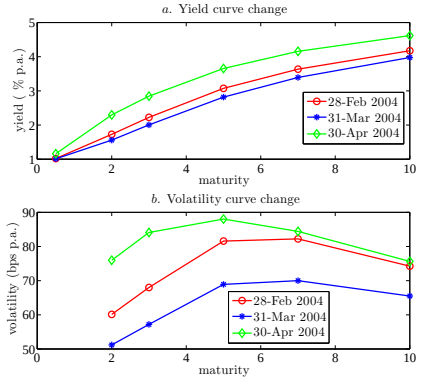
31-Jul-2000, 31-Aug-2000. The main topics are the labor productivity, level of NAIRU and change in the trend of potential output of the US economy. Market participants expect that the increased productivity should keep the inflation lower. At the same time, the economy is slowing from the record GDP growth. All these facts contribute to the rapid decline in the long term interest rates.	
30-Mar-2001, 30-Apr-2001. Leading indicators point to a recession. This leads to the repricing of the short end of the curve as markets expect the Fed to cut interest rates significantly even after the surprising -50 bps move in April 2001.	
30-Apr-2001, 31-May-2001. The flow of disappointing data starts already in April: payrolls, unemployment, all pointing toward recession. Long term yields increase due the concerns that the Fed might be easing too much. In the context of a general market belief: the unemployment data lags the business cycle. The volatility curve moves in the opposite direction to the yield curve and gets hump-shaped.	
31-May-2001, 30-Jun-2001. Recession is being priced-in, as the Fed cuts the rate only by 25 bps instead of 50 bps expected. The volatility increases at the short end and changes its shape.	

Table 2.7: Major moves in the US Treasury yield and volatility curves in the period 2000–2004 (2nd part)

31-Jul-2003, 31-Aug-2003, 30-Sep-2003. The upward shift in yields is caused by the strong labor productivity data, whereas the subsequent downward shift in September is due to the weak August employment data. The FOMC minutes from August meeting are published on September 19th, FOMC members stress the output gap. This leads market participants (together with the worsening data) to revise their expectations.	
31-Jul-2002, 31-Aug-2002. The data flow worsens: ISM manufacturing survey and the unemployment show alarming signals.	
31-Dec-2002, 31-Jan-2003. The ISM data read better than expected, the new orders index increases to 63.3% from 49.9%—the third biggest monthly gain since data is available. The positive news on ISM takes market participants by surprise, this in turn increases the volatility substantially.	
28-Feb-2004, 31-Mar-2004, 30-Apr-2004. The first downward shift in yields is caused by the disappointing data on jobless claims and retail sales. The upward move in yields in April is due to the improved economic data (especially manufacturing), but markets still do not expect Fed's tightening.	

Sources: Bloomberg, Reuters.

Table 2.8: Regressions of filtered states on macroeconomic surveys

This table reports regressions of the model-implied factors on macroeconomic surveys. Monthly factors are obtained by averaging the weekly numbers returned from the estimation. We splice monthly data from two surveys: BCFF and BCEI compiled over the period 1992:01–2007:12. This amounts to 192 monthly observations. The following variables are used: real GDP (RGDP), unemployment (UNEMPL), housing starts (HOUST), federal funds rate (FFR), industrial production (IP), consumer price index (CPI). RGDP and FFR forecasts are obtained from BCFF, while the remaining variables are from BCEI. $E(\cdot)$ denotes the consensus, defined as a median forecast; $\sigma(\cdot)$ proxies for the uncertainty, and is computed as the mean absolute deviation of individual forecasts. Standard errors are corrected using Newey-West adjustment with four lags and are reported in parentheses. Both the latent factors and the survey data are standardized in order to make the regression coefficients directly comparable. The table contains only significant variables. For the ease of interpretation, X_1 and X_2 have been multiplied by -1, so that they have a positive correlation with the two-year yield and ten-year yield, respectively.

Regressor	X_1	X_2	f	V_{11}	V_{12}	V_{22}
E(RGDP)	0.350	0.236	–	0.664	–	–
	(0.065)	(0.077)	–	(0.065)	–	–
σ(RGDP)	0.248	0.188	–	–	-0.257	–
	(0.063)	(0.057)	–	–	(0.117)	–
E(UNEMPL)	0.770	–	–	0.238	–	–
	(0.093)	–	–	(0.089)	–	–
σ(UNEMPL)	–	–	–	0.385	-0.197	–
	–	–	–	(0.081)	(0.100)	–
E(HOUST)	-0.417	–	–	-0.310	–	-0.315
	(0.098)	–	–	(0.080)	–	(0.052)
σ(HOUST)	–	-0.232	–	–	-0.386	–
	–	(0.053)	–	–	(0.079)	–
E(FFR)	0.516	0.862	0.868	–	–	-0.285
	(0.086)	(0.088)	(0.026)	–	–	(0.098)
σ(FFR)	–	–	0.041	–	-0.144	0.299
	–	–	(0.021)	–	(0.078)	(0.073)
E(IP)	–	–	-0.108	–	–	–
	–	–	(0.028)	–	–	–
σ(IP)	–	–	-0.057	–	–	0.305
	–	–	(0.023)	–	–	(0.078)
E(CPI)	–	–	0.074	–	–	–
	–	–	(0.021)	–	–	–
σ(CPI)	–	-0.172	–	–	–	0.254
	–	(0.061)	–	–	–	(0.071)
Adj. R^2	0.81	0.73	0.96	0.52	0.29	0.47

Table 2.9: Liquidity and the volatility curve

Panel *a* reports the regression results of funding liquidity on filtered volatility factors and survey-based proxy for monetary policy expectations $E(FFR)$. The funding liquidity is represented with the Fontaine and Garcia (2010) factor. Panel *b* reports the regression output of noise illiquidity measure on volatility factors. The last column reports the regression of daily noise illiquidity measure on the two-year realized volatility obtained from the high frequency data. The construction of the noise illiquidity measure follows Hu, Pan, and Wang (2010). The sample is 1992:01–2007:12 comprising 192 observations. Monthly filtered factors are obtained by averaging the weekly numbers returned from the estimation. t-statistics in parentheses are obtained using Newey-West adjustment with 12 lags. All variables are standardized.

Panel *a*: Funding liquidity				
V_{11}	-0.67			
	(-5.16)			
V_{12}		0.63		
		(5.12)		
V_{22}			-0.60	
			(-4.05)	
$E(FFR)$				0.65
				(4.40)
$\bar{R}^2$	0.45	0.40	0.35	0.42

Panel *b*: Noise illiquidity				
	monthly			**daily**
V_{11}	0.39			
	(3.02)			
V_{12}		-0.36		
		(-3.14)		
V_{22}			0.46	
			(4.08)	
RV 2Y				0.24
				(10.12)
$\bar{R}^2$	0.15	0.12	0.21	0.13

Appendix B

B.1 Data appendix

This Appendix gives a brief description of the high-frequency Treasury data, our zero curve construction methodology, and macroeconomic surveys.

B.1.A GovPX and BrokerTec

Table B.1 reports some basic statistics on the Treasury bond transaction data in the period 1992:01 through 2007:12. For the GovPX period (1992:01–2000:12), we report the average number of quotes per trading day and for the BrokerTec period (2001:01–2007:12) we report the average number of transactions per trading day. The number of Treasury bonds and bills totals to 1005 in our sample period. These were transacted or quoted more than 37.7 million times in the on-the-run secondary market.

Table B.1: Average number of quotes/trades per day in the GovPX and BrokerTec databases

Bond maturity	GovPX period	BrokerTec period
3M	374	–
6M	352	–
2Y	2170	1035
3Y	1385	773
5Y	3128	1987
7Y	637	–
10Y	2649	1956
30Y	793	633

B.1.B Testing for microstructure noise

To avoid potential bias in the estimates of the realized volatility using high-frequency data, we apply several tests for the presence of noise caused by the market microstructure effects. In a first step, we compute the first order autocorrelation in high-frequency price returns. Table B.2 reports the first order autocorrelation of equally-spaced ten-minute yield changes in the US Treasury zero curve in the period 1992:01–2007:12. The autocorrelation is statistically significant for the maturities of three, five and ten years. However, the magnitude of all autocorrelations is very small, which makes them economically insignificant.

Table B.2: Autocorrelation of high-frequency yield changes

	2Y	3Y	5Y	7Y	10Y
Autocorrelation	0.0039	0.0154	-0.0064	-0.0040	-0.0175
p-value	(0.0648)	(0.0001)	(0.0025)	(0.0589)	(0.0001)

In a second step, we use the volatility signature plots displaying the average realized volatility against the sampling frequency (Figure B.1). In the presence of microstructure noise, the average realized volatility increases with the sampling frequency. The reason is the dominance of noise at the very high-frequency sampling (see e.g., Bandi and Russell, 2008). None of the above diagnostics suggests that the microstructure noise present in our data is large and could overwhelm our results.

B.1.C Extracting zero coupon yield curve from high-frequency data

We fit the discount curve using smoothing splines. One of the important steps in the procedure is to select the appropriate number of knot points. We make the number of knot points dependent on the number of available bonds and locate them at the bond maturities. In our setting, the number of knot points varies between three and six. The fact that we consider only one specific part of the zero coupon yield curve allows us to use constant roughness penalty as in Fisher, Nychka, and Zervos (1994) for estimating the whole curve. Waggoner (1997) proposes a varying roughness penalty for the smoothing splines procedure with a low penalty at the short end and a high penalty at the very long end of the curve. In the period 2001:01–2007:12, the intraday quotes on Treasury bills are not available from the BrokerTec database. In order to anchor the very short end for the smoothing splines procedure, we include the daily data on the three-month Treasury bill obtained from the FRED database at the FRB St. Louis. Before using the constructed zero curve for the realized volatility estimation, we compare our zero coupon yields with the daily Constant Maturity Treasury rates (CMT) from the Fed, as well as with zero yields compiled by Gürkaynak, Sack, and Wright (2006) (GSW). Our daily yields are virtually perfectly correlated with the CMTs as well as with the GSW yields. Table B.3 summarizes the results.

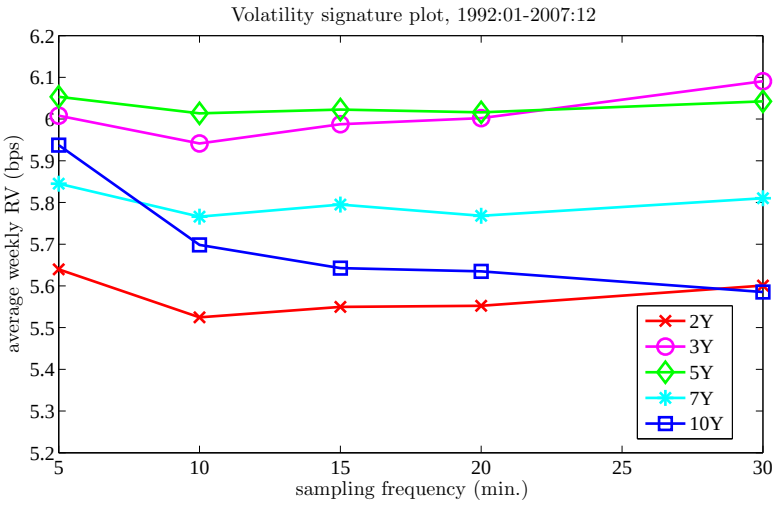


Figure B.1: Volatility signature plot

We plot the average weekly realized volatility (RV) against the sampling frequency for the whole sample 1992:01–2007:12. We consider five maturities in the zero coupon curve: two, three, five, seven, and ten years.

Table B.3: Correlation of zero coupon yields with CMT and GSW yields

	2Y	3Y	5Y	7Y	10Y
Corr CMT	1.000	1.000	0.999	0.996	0.997
Corr GSW	1.000	0.999	0.999	0.998	0.997

B.1.D Survey data

BlueChip Financial Forecasts. BlueChip Financial Forecasts (BCFF) survey contains monthly forecast of yields, inflation and GDP growth given by approximately 45 leading financial institutions. The BCFF is published on the first day of each month, but the survey itself is conducted over a two-day period, usually between the 23rd and 27th of each month. The exception is the survey for the January issue which generally takes place between the 17th and 20th of December. The precise dates as to when the survey was conducted are not published. The BCFF provides forecasts of constant maturity yields across several maturities: three and six months, one, two, five, ten, and 30 years. The short end of the term structure is additionally covered with the forecasts of the Fed funds rate, prime bank rate and three-month LIBOR rate. The forecasts are quarterly averages of interest rates for the current quarter, the next quarter out to five quarters ahead. The figures are expressed as percent per annum. In addition, panelist provide forecasts for macroeconomic quantities: real GDP, GDP price index and Consumer Price Index (CPI). The numbers are seasonally adjusted quarter-on-quarter changes.

BlueChip Economic Indicators. The BlueChip Economic Indicators (BCEI) survey contains individual and consensus forecasts of about 50 professional economists from leading financial and advisory institutions. The survey is compiled on a monthly basis, and contains predictions of key financial and macroeconomic indicators, e.g. real and nominal GDP, GDP deflator, CPI, three-month T-bill rate, industrial production, unemployment, housing starts. The survey is conducted over two days, generally beginning on the first business day of each month. The newsletter is typically finished on the third day following completion of the survey and published on the tenth of a month. Every month, panelists provide two types of forecasts: (i) average figure for the current calendar year and (ii) average figure for the next calendar year. For instance, in January 2001 the survey contains forecasts for 2001 and 2002. In February 2001, the forecast horizon shrinks to 11 months for the current year, and to 23 months for the next year, and so on. The diminishing forecast horizon implies that the cross-sectional uncertainty measures computed from the individual responses display a visible seasonal pattern. To gauge uncertainty, every month we use the mean absolute deviation of individual forecasts. To remove the problem of seasonality, we adjust the series with a X-12 ARIMA filter. Consensus forecast is defined as the median of individual forecasts in a given month.

B.2 Technical appendix

This Appendix derives moments of the state variables necessary for the implementation of the unscented Kalman filter.

B.2.A Moments of the V_t process

The first conditional moment of the volatility process V_t is given as:

$$E_t\left(V_{t+\Delta t}\right) = k\bar{\mu}_{V,\Delta t} + \Phi_{V,\Delta t} V_t \Phi'_{V,\Delta t}, \tag{B.1}$$

where

$$\Phi_{V,\Delta t} = e^{M\Delta t} \tag{B.2}$$

$$\bar{\mu}_{V,\Delta t} = \int_0^{\Delta t} e^{M\Delta t} Q'Q e^{M'\Delta t} ds = -\frac{1}{2}\hat{C}_{12}(\Delta t)\hat{C}'_{11}\left(\Delta t\right), \tag{B.3}$$

with

$$\begin{pmatrix} \hat{C}_{11}(\Delta t) & \hat{C}_{12}(\Delta t) \\ \hat{C}_{21}(\Delta t) & \hat{C}_{22}(\Delta t) \end{pmatrix} = \exp\left[\Delta t \begin{pmatrix} M & -2Q'Q \\ 0 & -M' \end{pmatrix}\right].$$

Assuming stationarity (i.e. negative eigenvalues of M), the unconditional first moment of V_t follows as:

$$\lim_{\Delta t\to\infty} \text{vec}E_t\left(V_{t+\Delta t}\right) = k\text{vec}\left(\bar{\mu}_{V,\infty}\right) = -k\left[\left(I\otimes M\right)+\left(M\otimes I\right)\right]^{-1}\text{vec}(Q'Q). \tag{B.4}$$

The conditional and unconditional covariance matrix of V_t reads:

$$Cov_t\left[\text{vec}\left(V_{t+\Delta t}\right)\right] = \left(I_{n^2}+K_{n,n}\right)\left[\Phi_{V,\Delta t}V_t\Phi'_{V,\Delta t}\otimes\bar{\mu}_{V,\Delta t} + k\left(\bar{\mu}_{V,\Delta t}\otimes\bar{\mu}_{V,\Delta t}\right)+\bar{\mu}_{V,\Delta t}\otimes\Phi_{V,\Delta t}V_t\Phi'_{V,\Delta t}\right]. \tag{B.5}$$

$$\lim_{\Delta t\to\infty} Cov_t\left[\text{vec}\left(V_{t+\Delta t}\right)\right] = \left(I_{n^2}+K_{n,n}\right)k\left(\bar{\mu}_{V,\infty}\otimes\bar{\mu}_{V,\infty}\right). \tag{B.6}$$

$K_{n,n}$ is the commutation matrix with the property that $K_{n,n}\text{vec}(A) = \text{vec}(A')$. These moments are derived in Buraschi, Cieslak, and Trojani (2008) and thus are stated without a proof.

Gourieroux, Jasiak, and Sufana (2009) show that when $\Omega\Omega' = kQ'Q$, k integer, the dynamics of V_t can be represented as the sum of outer products of k independent Ornstein-Uhlenbeck processes with a zero long-run mean:

$$V_t = \sum_{i=1}^{k} v_t^i v_t^{i\prime} \tag{B.7}$$

$$v_{t+\Delta t}^i = \Phi_{V,\Delta t}v_t^i + \epsilon_{t+\Delta t}^i, \quad \epsilon_t^i \sim N(0,\bar{\mu}_{V,\Delta t}). \tag{B.8}$$

Taking the outer-product implies that the exact discretization of V_t has the form:

$$V_{t+\Delta t} = k\bar{\mu}_{V,\Delta t} + \Phi_{V,\Delta t}V_t\Phi'_{V,\Delta t} + u_{t+\Delta t}^V, \tag{B.9}$$

where the shock $u_{t+\Delta t}^V$ is a heteroskedastic martingale difference sequence.

B.2.B Moments of the Y_t dynamics

We assume that the dimension of X_t is $n = 2$ and f_t is a scalar process. Let $Y_t = (X_t', f_t)'$:

$$dY_t = \left(\mu_Y + \mathcal{K}_Y Y_t\right)dt + \Sigma\left(V_t\right)dZ_t. \tag{B.10}$$

It is straightforward to show that the conditional and unconditional first moment of Y_t has the form:

$$E_t\left(Y_{t+\Delta t}\right) = \left(e^{\mathcal{K}_Y\Delta t} - I\right)\mathcal{K}_Y^{-1}\mu_Y + e^{\mathcal{K}_Y\Delta t}Y_t \tag{B.11}$$

$$\lim_{\Delta t\to\infty} E_t\left(Y_{t+\Delta t}\right) = -\mathcal{K}_Y^{-1}\mu_Y, \tag{B.12}$$

where $\mathcal{K}_Y$ is assumed to be lower triangular with negative eigenvalues.

To compute the conditional covariance of Y_t, let $V_Y\left(t,T\right) := Cov_t\left(Y_T\right)$. Following Fisher and Gilles (1996a), the application of Ito's lemma to $\hat{Y}(t,T) := E_t\left(Y_T\right)$ reveals that:

$$d\hat{Y}\left(t,T\right)=\hat{\sigma}_{Y}\left(t,T\right)dZ_{t}, \tag{B.13}$$

where $\hat{\sigma}_{Y}(t,T):=\Phi_{Y}\left(t,T\right)\Sigma\left(V_{t}\right)$, with

$$\Phi_{Y}\left(t,T\right)=e^{\mathcal{K}_{Y}(T-t)} \tag{B.14}$$

and

$$\Sigma_{Y}\left(V_{t}\right)=\left(\begin{array}{cc}\sqrt{V_{t}} & 0\\ 0 & \sigma_{f}^{2}\end{array}\right). \tag{B.15}$$

Then, integrating $d\hat{Y}\left(t,T\right)$ yields:

$$Y_{T}=\hat{Y}_{T,T}=\hat{Y}_{t,T}+\int_{s=t}^{T}\hat{\sigma}_{Y}\left(s,T\right)dZ_{s}. \tag{B.16}$$

Therefore, we have:

$$V_{Y}(t,T)=Cov_{t}\left[\int_{s=t}^{T}\hat{\sigma}_{Y}(s,T)dZ_{s}^{Y}\right]=E_{t}\left[\int_{s=t}^{T}\hat{\sigma}_{Y}(s,T)\hat{\sigma}_{Y}(s,T)'ds\right] \tag{B.17}$$

$$=\int_{s=t}^{T}\Phi_{Y}(s,T)E_{t}\left(\begin{array}{cc}V_{s} & 0\\ 0 & \sigma_{f}^{2}\end{array}\right)\Phi_{Y}'(s,T)ds. \tag{B.18}$$

Note that since $\mathcal{K}_{Y}$ is lower triangular, $\Phi_{Y}(t,T)=e^{\mathcal{K}_{Y}(T-t)}$ is also lower triangular, and we have:

$$\Phi_{Y}\left(t,T\right)=\left(\begin{array}{cc}\Phi_{X}\left(t,T\right) & 0\\ \Phi_{Xf}\left(t,T\right) & \Phi_{f}\left(t,T\right)\end{array}\right). \tag{B.19}$$

Let us for convenience define two matrices:

$$\mathcal{M}_{1Y}(t,T)=\left(\begin{array}{c}\Phi_{X}(t,T)\otimes\Phi_{X}(t,T)\\ \Phi_{X}(t,T)\otimes\Phi_{fX}(t,T)\\ \Phi_{fX}(t,T)\otimes\Phi_{X}(t,T)\\ \Phi_{fX}(t,T)\otimes\Phi_{fX}(t,T)\end{array}\right)\text{ and }\mathcal{M}_{0Y}=\left(\begin{array}{c}0_{8\times1}\\ \Phi_{f}^{2}(t,T)\sigma_{f}^{2}\end{array}\right). \tag{B.20}$$

With help of simple matrix algebra applied to (B.68), the conditional covariance of Y_t has the (vectorized) form

$$\text{vec}V_{Y}(t,T)=\int_{s=t}^{T}\mathcal{M}_{1Y}(s,T)\left[\Phi_{V}(s,T)\otimes\Phi_{V}(s,T)\right]ds\times\text{vec}\left(V_{t}\right)+\int_{s=t}^{T}k\mathcal{M}_{1Y}\text{vec}\left[\bar{\mu}_{V}(t,s)\right]ds+\int_{s=t}^{T}\mathcal{M}_{0Y}(s,T)ds. \tag{B.21}$$

The unconditional covariance of Y is given as:

$$\lim_{T\to\infty} \text{vec}V_Y(t,T) = \lim_{T\to\infty} \int_{s=t}^{T} k\mathcal{M}_{1Y}(s,T)\text{vec}\left[\bar{\mu}_V(t,s)\right] ds + \int_{s=t}^{T} \mathcal{M}_{0Y}(s,T)ds. \tag{B.22}$$

This expression exists if the mean reversion matrices M and $\mathcal{K}_Y$ are negative definite.

The expressions for the conditional mean (B.11) and covariance (B.21) give rise to an exact discretization of the process Y_t.

Remark 1. In order to avoid the numerical integration, we can resort to a discrete-time approximation of the unconditional covariance matrix of Y factors. To this end, we discretize the dynamics

$$dY_t = \left(\mu_Y + \mathcal{K}_Y Y_t\right) dt + \Sigma_Y\left(V_t\right) dZ_t \tag{B.23}$$

as

$$Y_{t+\Delta t} = \bar{\mu}_{Y,\Delta t} + \Phi_{Y,\Delta t} Y_t + \Sigma_Y\left(V_t\right)\sqrt{\Delta t}\varepsilon_{t+\Delta t}, \tag{B.24}$$

where $\bar{\mu}_{Y,\Delta t} = \left(e^{\mathcal{K}_Y \Delta t} - I\right)\mathcal{K}_Y^{-1}\mu_Y$. The second moment of the discretized dynamics is straightforward to obtain as:

$$\begin{aligned}
\text{vec}E\left(YY'\right) &= \left(I - \Phi_{Y,\Delta t}\otimes\Phi_{Y,\Delta t}\right)^{-1}\times \\
&\times\text{vec}\left\{\bar{\mu}_{Y,\Delta t}\bar{\mu}'_{Y,\Delta t} + \bar{\mu}_{Y,\Delta t}E\left(Y'\right)\Phi'_{Y,\Delta t} + \Phi_{Y,\Delta t}E\left(Y\right)\bar{\mu}'_{Y,\Delta t} + E\left[\Sigma_Y\left(V_t\right)\Sigma_Y\left(V_t\right)'\right]\Delta t\right\} \\
\text{vec}\left[Var\left(Y\right)\right] &= \text{vec}E\left(YY'\right) - \text{vec}E\left(Y\right)\left[\text{vec}E\left(Y\right)\right]'.
\end{aligned} \tag{B.25}$$

We check that for the weekly discretization step $\Delta t = \frac{1}{52}$ this approximation works well, and implies a significant reduction of the computational time.

B.3 General solution

We provide a solution for the general version of the model, which incorporates both correlation between the dW and dZ shocks and a general form of the market prices of risk. Based on arguments presented in the body of the paper, we analyze a restricted version of the model, in which the correlation parameter is set to zero and only dZ shocks are priced.

B.3.A Dependence between X and V factors

In the general case, X_t and V_t can be correlated, i.e.:

$$dZ_X = dW\rho + \sqrt{1-\rho'\rho}\, dB \tag{B.26}$$

$$= dW\rho + \widetilde{\rho} dB, \tag{B.27}$$

where dB is a (2×1)-vector of Brownian motions which is independent from dW, and ρ is a (2×1)-vector such that $\rho \in [-1,1]$ and $\rho'\rho < 1$ (e.g., da Fonseca, Grasselli, and Tebaldi, 2006; Buraschi, Porchia, and Trojani, 2009). We use short notation $\widetilde{\rho} := \sqrt{1-\rho'\rho}$.

B.3.B General form of the market prices of risk

Let us write the shocks to Y under the physical dynamics as (for brevity we omit the superscript $\mathbb{P}$):

$$dZ = \begin{pmatrix} dZ_X \\ dZ_f \end{pmatrix} = \begin{pmatrix} dW\rho + \widetilde{\rho} dB \\ dZ_f \end{pmatrix} = \begin{pmatrix} dW\rho \\ 0_{1\times 2} \end{pmatrix} + \underbrace{\begin{pmatrix} \widetilde{\rho} I_{2\times 2} & 0_{2\times 1} \\ 0_{1\times 2} & 1_{1\times 1} \end{pmatrix}}_{R} \underbrace{\begin{pmatrix} dB \\ dZ_f \end{pmatrix}}_{d\widetilde{Z}} = \begin{pmatrix} dW\rho \\ 0_{1\times 2} \end{pmatrix} + R d\widetilde{Z}, \tag{B.28}$$

where

$$R = \begin{pmatrix} \widetilde{\rho} I_{2\times 2} & 0_{2\times 1} \\ 0_{1\times 2} & 1_{1\times 1} \end{pmatrix}, \qquad d\widetilde{Z} = \begin{pmatrix} dB \\ dZ_f \end{pmatrix}. \tag{B.29}$$

The change of drift is specified as:

$$d\widetilde{Z} = d\widetilde{Z}^{\mathbb{Q}} - \Lambda_{Y,t} dt \tag{B.30}$$

$$dW = dW^{\mathbb{Q}} - \Lambda_{V,t} dt \tag{B.31}$$

$$\Lambda_{Y,t} = \Sigma_Y^{-1}(V_t)\left(\lambda_Y^0 + \lambda_Y^1 Y_t\right) \tag{B.32}$$

$$\Lambda_{V,t} = \left(\sqrt{V_t}\right)^{-1} \Lambda_V^0 + \sqrt{V_t}\Lambda_V^1, \tag{B.33}$$

where λ_Y^0 and λ_Y^1 are a $(n+1)$-vector and $(n+1)\times(n+1)$ matrix of parameters, and Λ_V^0 and Λ_V^1 are $n\times n$ constant matrices. To exclude arbitrage, the market price of risk requires that the parameter matrix Q be invertible, so that V_t stays in the positive-definite domain. This specification implies the risk-neutral dynamics of Y_t given by:

$$dY_t = \left[\left(\mu_Y - \begin{pmatrix} \Lambda_V^0\rho \\ 0 \end{pmatrix} - R\lambda_Y^0\right) + \left(\mathcal{K}_Y - R\lambda_Y^1\right) Y_t - \begin{pmatrix} V_t\Lambda_V^1\rho \\ 0 \end{pmatrix}\right] dt + \Sigma_Y\left(V_t\right) dZ_t^{\mathbb{Q}}. \tag{B.34}$$

Let:

$$\mu_Y^Q = \mu_Y - \begin{pmatrix} \Lambda_V^0\rho \\ 0 \end{pmatrix} - R\lambda_Y^0 \tag{B.35}$$

$$\mathcal{K}_Y^Q = \mathcal{K}_Y - R\lambda_Y^1. \tag{B.36}$$

The dynamics of V_t is given as:

$$dV_t = \left[\left(\Omega\Omega' - \Lambda_V^0 Q - Q'\Lambda_V^{0\prime}\right) + \left(M - Q'\Lambda_V^{1\prime}\right) V_t + V_t\left(M' - \Lambda_V^1 Q\right)\right] dt + \sqrt{V_t} dW_t^{\mathbb{Q}} Q + Q' dW_t^{\mathbb{Q}\prime}\sqrt{V_t}. \tag{B.37}$$

Let

$$\Omega^{\mathbb{Q}}\Omega^{\mathbb{Q}\prime} = \Omega\Omega' - \Lambda_V^0 Q - Q'\Lambda_V^{0\prime} = (k-2v)\, Q'Q \tag{B.38}$$

$$M^{\mathbb{Q}} = M - Q'\Lambda_V^{1\prime}, \tag{B.39}$$

where, to preserve the same distribution under $\mathbb{P}$ and $\mathbb{Q}$, we assume $\Lambda_V^0 = vQ'$ for a scalar v such that $(k-2v) > n-1$.

B.3.C Solution for bond prices

Since both components of the state vector, i.e. Y_t, V_t, are affine, bond prices are of the form:

$$F\left(Y_t, V_t; t, \tau\right) = \exp\left\{A\left(\tau\right) + B\left(\tau\right)' Y_t + Tr\left[C\left(\tau\right) V_t\right]\right\}. \tag{B.40}$$

By discounted Feynman-Kac theorem, the drift of dF equals rF, thus:

$$\mathcal{L}_{\{Y,V\}} F + \frac{\partial F}{\partial t} = rF, \tag{B.41}$$

where $\mathcal{L}_{\{Y,V\}}$ is the joint infinitesimal generator of the couple $\{Y_t, V_t\}$ under the risk neutral measure. We have:

$$\mathcal{L}_{\{Y,V\}}F = \left(\mathcal{L}_Y + \mathcal{L}_V + \mathcal{L}_{Y,V}\right)F \tag{B.42}$$

$$\mathcal{L}_Y F = \frac{\partial F}{\partial Y'}\left[\mu_Y^{\mathbb{Q}} + \mathcal{K}_Y^{\mathbb{Q}} Y - \begin{pmatrix} V_t \Lambda_V^1 \rho \\ 0 \end{pmatrix}\right] + \frac{1}{2} Tr\left[\frac{\partial F}{\partial Y \partial Y'} \Sigma_Y(V)\, \Sigma_Y'(V)\right] \tag{B.43}$$

$$\mathcal{L}_V F = Tr\left[\left(\Omega^{\mathbb{Q}}\Omega^{\mathbb{Q}\prime} + M^{\mathbb{Q}}V + V M^{\mathbb{Q}\prime}\right)\mathcal{R}F + 2V\mathcal{R}Q'Q\mathcal{R}F\right] \tag{B.44}$$

$$\mathcal{L}_{Y,V} F = 2Tr\left[\left(\mathcal{R}Q'\rho\frac{\partial}{\partial X'}\right)FV\right]. \tag{B.45}$$

$\mathcal{R}$ is a matrix differential operator: $\mathcal{R}_{ij} := \left(\frac{\partial}{\partial V_{ij}}\right)$. Substituting derivatives of (B.40) into (B.41) gives:

$$B_\tau'\left(\mu_Y^{\mathbb{Q}} + \mathcal{K}_Y^{\mathbb{Q}} Y\right) - Tr\left(\Lambda_V^1 \rho B_{X,\tau}' V\right) + \frac{1}{2} Tr\left(B_{X,\tau} B_{X,\tau}' V\right) + \frac{1}{2} B_{f,\tau}^2 \sigma_f^2 \tag{B.46}$$

$$+ Tr\left(\Omega^{\mathbb{Q}}\Omega^{\mathbb{Q}\prime} C_\tau\right) + Tr\left[\left(C_\tau M^{\mathbb{Q}} + M^{\mathbb{Q}\prime} C_\tau + 2C_\tau Q' Q C_\tau\right) V\right] \tag{B.47}$$

$$+ Tr\left[\left(C_\tau Q' \rho B_{X,\tau}' + B_{X,\tau}\rho' Q C_\tau\right) V\right] \tag{B.48}$$

$$= \frac{\partial A_\tau}{\partial \tau} + \frac{\partial B_\tau}{\partial \tau}' Y + Tr\left(\frac{\partial C_\tau}{\partial \tau} V\right) + \gamma_0 + \gamma_Y' Y \tag{B.49}$$

By matching coefficients, we obtain the system of equations:

$$\frac{\partial A}{\partial \tau} = B_\tau' \mu_Y^{\mathbb{Q}} + \frac{1}{2} B_{f,\tau}^2 \sigma_f^2 + Tr\left(\Omega^{Q}\Omega^{Q\prime} C_\tau\right) - \gamma_0 \tag{B.50}$$

$$\frac{\partial B}{\partial \tau} = \mathcal{K}^{\mathbb{Q}\prime} B_\tau - \gamma_Y \tag{B.51}$$

$$\frac{\partial C}{\partial \tau} = \frac{1}{2} B_{X,\tau} B_{X,\tau}' + C_\tau\left(M^{\mathbb{Q}} + Q'\rho B_{X,\tau}'\right) + \left(M^{\mathbb{Q}\prime} + B_{X,\tau}\rho' Q\right) C_\tau + 2C_\tau Q' Q C_\tau - \Lambda_V^1 \rho B_{X,\tau}' \tag{B.52}$$

To obtain the solution provided in the text, set $\rho = 0_{2\times 1}$, $\Lambda_V^0 = 0_{2\times 2}$ and $\Lambda_V^1 = 0_{2\times 2}$.

B.3.D Instantaneous volatility of yields

The instantaneous volatility of yields is given as:

$$\frac{1}{dt}\left\langle dy_t^{\tau_1}, dy_t^{\tau_2}\right\rangle = \frac{1}{\tau_1 \tau_2} Tr\left[B_{f,\tau_2} B_{f,\tau_1} \sigma_f^2 + \left(B_{X,\tau_1} B_{X,\tau_2}' + 2C_{\tau_2} Q'\rho B_{X,\tau_1}' + 2C_{\tau_1} Q'\rho B_{X,\tau_2}' + 4C_{\tau_1} Q' Q C_{\tau_2}\right) V_t\right]. \tag{B.53}$$

Proof. The only term which requires clarification is $B_{\tau_1}' dY_t \times Tr\left[C_{\tau_2} dV_t\right] = B_{X,\tau_1}' dX_t \times Tr\left[C_{\tau_2} dV_t\right]$

$$\begin{aligned} B_{X,\tau_1}' dX_t \times Tr\left[C_{\tau_2} dV_t\right] &= B_{X,\tau_1}' \sqrt{V} dZ_X \times Tr\left[C_{\tau_2}\left(\sqrt{V} dW Q + Q' dW' \sqrt{V}\right)\right] \\ &= B_{X,\tau_1}' \sqrt{V}\left(dW\rho + \widetilde{\rho} dB\right) \times 2Tr\left(Q C_{\tau_2} \sqrt{V} dW\right) \\ &= 2Tr\left(C_{\tau_2} Q'\rho B_{X,\tau_1}' V\right) \end{aligned}$$

Where we use the following fact:

$$Tr\left[C\left(\sqrt{V}dWQ + Q'dW'\sqrt{V}\right)\right] = 2Tr\left(QC\sqrt{V}dW\right). \tag{B.54}$$

□

B.3.E Conditional covariance of X and V

We consider the conditional covariance matrix of X and V

$$d\left\langle \begin{pmatrix} X_{t,1} \\ X_{t,2} \\ f_t \end{pmatrix}, \begin{pmatrix} V_{t,11} \\ V_{t,12} \\ V_{t,22} \end{pmatrix} \right\rangle = \begin{pmatrix} d(X_1, V_{11}) & d(X_1, V_{12}) & d(X_1, V_{22}) \\ d(X_2, V_{11}) & d(X_2, V_{12}) & d(X_2, V_{22}) \\ d(f, V_{11}) & d(f, V_{12}) & d(f, V_{22}) \end{pmatrix} \tag{B.55}$$

The elements of the covariance matrix are given by:

$$d\langle X_k, V_{ij}\rangle = \rho'\left(Q_{:,j}V_{ik} + Q_{:,i}V_{jk}\right), \tag{B.56}$$

where $Q_{:,j}$ denotes the j-th column of matrix Q.

Proof. The expression follows by simple algebra:

$$\begin{aligned}
\frac{1}{dt}d\langle V_{ij}, X_k\rangle &= \left[e_i'\left(\sqrt{V}dWQ\right)e_j + e_i'\left(Q'dW'\sqrt{V}\right)e_j\right]\left(e_k'\sqrt{V}dW\rho\right) \\
&= Tr\left(e_j e_i'\sqrt{V}dWQ\right) \times Tr\left(\rho e_k'\sqrt{V}dW\right) + Tr\left(e_j e_i'Q'dW'\sqrt{V}\right) \times Tr\left(\rho e_k'\sqrt{V}dW\right) \\
&= \text{vec}\left(\sqrt{V}e_i e_j'Q'\right)' \text{vec}\left(\sqrt{V}e_k\rho'\right) + \text{vec}\left(\sqrt{V}e_j e_i'Q'\right)' \text{vec}\left(\sqrt{V}e_k\rho'\right) \\
&= Tr\left(Qe_j e_i'Ve_k\rho'\right) + Tr\left(Qe_i e_j'Ve_k\rho'\right) \\
&= \rho'\left(Q_{:,j}V_{ik} + Q_{:,i}V_{jk}\right),
\end{aligned} \tag{B.57}$$

where e_i is the i-th column of the identity matrix. □

B.3.F Discrete approximation to the unconditional covariance matrix of X and V

We can use the discretized dynamics of X and V to compute the unconditional covariance matrix:

$$X_{t+\Delta t} = \bar{\mu}_{X,\Delta t} + \Phi_{X,\Delta t}X_t + \sqrt{V_t\Delta t}\left(U_{t+\Delta t}\rho + \widetilde{\rho}b_{t+\Delta t}\right) \tag{B.58}$$

$$V_{t+\Delta t} = k\bar{\mu}_{V,\Delta t} + \Phi_{V,\Delta t}V_t\Phi_{V,\Delta t}' + \sqrt{V_t\Delta t}U_{t+\Delta t}Q + Q'U_{t+\Delta t}'\sqrt{V_t\Delta t}, \tag{B.59}$$

where U_t is a 2×2 matrix of Gaussian shocks, and b_t is a 2-vector of Gaussian shocks. The covariance between X and V is computed as :

$$Cov\left[X, \text{vec}(V)\right] = E\left[X\text{vec}\left(V\right)'\right] - E\left(X\right)E\left[\text{vec}\left(V\right)'\right]. \tag{B.60}$$

The element $E\left[X\text{vec}\left(V\right)'\right]$ reads:

$$\text{vec}E\left[X\left(\text{vec}V\right)'\right] = \left[I_{n^3} - \left(\Phi_V \otimes \Phi_V\right) \otimes \Phi_X\right]^{-1}\left(\text{vec}A + \text{vec}B\right), \tag{B.61}$$

where A is given as:

$$A = \bar{\mu}_X \text{vec}\left(k\bar{\mu}_V\right)' + \bar{\mu}_X \text{vec}\left[\Phi_V E\left(V_t\right)\Phi_V'\right] + \Phi_X E(X_t)\text{vec}(k\bar{\mu}_V)', \tag{B.62}$$

and the elemtent (k, ij) of matrix B, associated with the covariance of X_k and V_{ij} has the form:

$$B_{k,ij} = \rho'\left(Q_{:,j}V_{ik} + Q_{:,i}V_{jk}\right)\Delta t, \tag{B.63}$$

where $B = \begin{pmatrix} B_{1,11} & B_{1,12} & B_{1,21} & B_{1,22} \\ B_{2,11} & B_{2,12} & B_{2,21} & B_{2,22} \end{pmatrix}$. Note that the second and third columns of B are identical.

B.4 Transition dynamics

This Appendix collects the details about the vectorization of transition dynamics for V_t.

We recast Eq. (2.19) in a vector form:

$$\text{vec}(V_{t+\Delta t}) = k\text{vec}(\bar{\mu}_{V,\Delta t}) + (\Phi_{V,\Delta t} \otimes \Phi_{V,\Delta t})\text{vec}(V_t) + \text{vec}(u^V_{t+\Delta t}). \tag{B.64}$$

Since the process V_t lives in the space of symmetric matrices, its lower triangular part preserves all information. Let us for convenience define two linear transformations of some symmetric matrix A: (i) an elimination matrix: $\mathcal{E}_n\text{vec}(A) = \text{vech}(A)$, where $\text{vech}(\cdot)$ denotes half-vectorization, (ii) a duplication matrix: $\mathcal{D}_n\text{vech}(A) = \text{vec}(A)$. Using half-vectorization, we define $\bar{V}_t := \text{vech}(V_t) = \mathcal{E}_n\text{vec}(V_t)$, which contains $\bar{n} = n(n+1)/2$ unique elements of V_t:

$$\bar{V}_{t+\Delta t} = k\mathcal{E}_n\text{vec}(\bar{\mu}_{V,\Delta t}) + \mathcal{E}_n(\Phi_{V,\Delta t} \otimes \Phi_{V,\Delta t})\mathcal{D}_n\bar{V}_t + \mathcal{E}_n\text{vec}(u^V_{t+\Delta t}). \tag{B.65}$$

Collecting all elements, we can redefine the state as: $S_t = (Y_t', \bar{V}_t')'$, whose transition is described by the conditional mean:

$$E_t\left(S_{t+\Delta t}\right) = \begin{pmatrix} \left(e^{\mathcal{K}_Y \Delta t} - I\right)\mathcal{K}_Y^{-1}\mu_Y + e^{\mathcal{K}_Y \Delta t}Y_t \\ k\mathcal{E}_n\text{vec}(\bar{\mu}_{V,\Delta t}) + \mathcal{E}_n(\Phi_{V,\Delta t} \otimes \Phi_{V,\Delta t})\mathcal{D}_n\bar{V}_t \end{pmatrix}, \tag{B.66}$$

and the conditional covariance of the form:

$$Cov_t\left(S_{t+\Delta t}\right) = \begin{pmatrix} Cov_t(Y_{t+\Delta t}) & 0_{n\times\bar{n}} \\ 0_{\bar{n}\times n} & Cov_t\left(\bar{V}_{t+\Delta t}\right) \end{pmatrix}. \tag{B.67}$$

The block diagonal structure in the last expression follows from our assumption that shocks in Y_t be independent of shocks in V_t. The respective blocks are given as:

$$Cov_t(Y_{t+\Delta t}) = \Sigma_Y(V_t)\Sigma_Y(V_t)'\Delta t \tag{B.68}$$

$$\begin{aligned} Cov_t\left(\bar{V}_{t+\Delta t}\right) &= \mathcal{E}_n Cov_t\left(V_{t+\Delta t}\right)\mathcal{E}_n' \\ &= \mathcal{E}_n\left(I_{n^2}+K_{n,n}\right)\left[\Phi_{V,\Delta t}V_t\Phi_{V,\Delta t}'\otimes\bar{\mu}_{V,\Delta t}+k\left(\bar{\mu}_{V,\Delta t}\otimes\bar{\mu}_{V,\Delta t}\right)+\bar{\mu}_{V,\Delta t}\otimes\Phi_{V,\Delta t}V_t\Phi_{V,\Delta t}'\right]\mathcal{E}_n', \end{aligned} \tag{B.69}$$

where $K_{n,n}$ denotes a commutation matrix (see e.g., Magnus and Neudecker, 1979). Buraschi, Cieslak, and Trojani (2008) provide the derivation of the last expression.

B.5 Identification

This Appendix details our econometric identification procedure and parameter restrictions.

To ensure econometric identification, we consider invariant model transformations of the type $\tilde{Y}_t = v + LY_t$ and $\tilde{V}_t = LV_tL'$, for a scalar v and an invertible matrix L. Such transformations result in the equivalence of the state variables, the short rate and thus yields (Dai and Singleton, 2000). If allowed, they can invalidate the results of an estimation.

To prevent the invariance, we adopt several normalizations for the physical dynamics of the process Y_t: (i) Setting $\mu_Y = 0$ allows to treat γ_0 as a free parameter. (ii) Restricting $\gamma_f = 1$ makes σ_f identified. (iii) Since both $\mathcal{K}_X$ and V_t determine interactions between the elements of X_t, they are not separately identifiable. We set $\mathcal{K}_X$ to a diagonal matrix, and allow correlations of the X_t factors to be generated solely by V_t. By the same token, the last row of matrix $\mathcal{K}_Y$, i.e. $(\mathcal{K}_{fX}, \mathcal{K}_f)$ is left unrestricted, as f_t does not interact with X_t via the diffusion term.

The identification of volatility factors V_t is ensured with three restrictions: (i) M is lower triangular and (ii) Q is diagonal with positive elements. (iii) The diagonal elements of Q are uniquely determined by setting $\gamma_X = \mathbf{1}_{n\times 1}$, where $\mathbf{1}_{n\times 1}$ is a vector of ones. These normalizations protect V_t against affine transformations and orthonormal rotations of Brownian motions. Finally, to guarantee the stationarity of the state, we require that the mean reversion matrices $\mathcal{K}_Y$ and M be negative definite. Due to the lower triangular structure of both, this is equivalent to restricting the diagonal elements of each matrix to be negative.

B.6 Estimation Appendix

This Appendix summarizes the algorithm for the unscented Kalman filtering.

We recast the transition and measurement equations above into one state space. The compound transition equation is given by:

$$S_{t+\Delta t} = A + BS_t + \varepsilon_{t+\Delta t}, \tag{B.70}$$

and the compound measurement equation is given by:

$$m_t = h(S_t; \Theta) + \vartheta_t. \tag{B.71}$$

$S_t = (Y_t', \bar{V}_t')'$ and A are $(n + \bar{n} + 1) \times 1$-dimensional vectors, A is given by:

$$A = \begin{pmatrix} (\Phi_{Y,\Delta t} - I)\, \mathcal{K}_Y^{-1} \mu_Y \\ k \cdot \mathcal{E}_n vec(\bar{\mu}_{V,\Delta t}) \end{pmatrix}. \tag{B.72}$$

B is a block-diagonal matrix of the form:

$$B = \begin{pmatrix} \Phi_{Y,\Delta t} & \mathbf{0}_{n\times\bar{n}} \\ \mathbf{0}_{\bar{n}\times n} & \mathcal{E}_n(\Phi_{V,\Delta t} \otimes \Phi_{V,\Delta t})\mathcal{D}_n \end{pmatrix}. \tag{B.73}$$

The vector shocks is of the form:

$$\varepsilon_{t+\Delta t} = \begin{pmatrix} u^Y_{t+\Delta t} \\ \mathcal{E}_n \text{vec}(u^V_{t+\Delta t}) \end{pmatrix}, \tag{B.74}$$

and its covariance matrix is given by a block-diagonal matrix:

$$Cov_t(\varepsilon_{t+\Delta t}) = \begin{pmatrix} Cov_t(Y_{t+\Delta t}) & \mathbf{0}_{n\times\bar{n}} \\ \mathbf{0}_{\bar{n}\times n} & Cov_t(\bar{V}_{t+\Delta t}) \end{pmatrix}. \tag{B.75}$$

m_t is a vector of observed yields and volatility measures given by $m_t = (y_t^\tau, v_t^{\tau_i,\tau_j})'$. Model implied yields and volatilities are affine in the state vector. Function $h(\cdot)$ translates the state variables to model implied yields and volatilities:

$$h(S_t; \Theta) = \begin{pmatrix} f(S_t; \Theta) \\ g(V_t; \Theta) \end{pmatrix}. \tag{B.76}$$

The vector of measurement errors:

$$\vartheta_t = \begin{pmatrix} \sqrt{R_y} e_t^y \\ \sqrt{R_v} e_t^v \end{pmatrix} \tag{B.77}$$

is Gaussian with the covariance matrix, for six yields and three volatility measurements, is given by:

$$Cov(\vartheta_t) = \begin{pmatrix} \sigma_y^2 \mathbf{I}_6 & \mathbf{0}_{6\times 3} \\ \mathbf{0}_{3\times 6} & diag(\sigma_{i,v}^2)_{i=1,2,3} \end{pmatrix}. \tag{B.78}$$

The core of UKF is the unscented transformation which approximates a distribution of a nonlinear transformation of any random variable by a set of sample points. In the UKF framework, we apply the unscented transformation recursively to B and $h(\cdot)$.

We define $L_S := n + \bar{n} + 1$. Assume that we know the mean $\bar{S}$ and the covariance P_S of S_t at each point in time t. We form a matrix $\mathcal{S}$ of $2L_S + 1$ sigma vectors:

$$\mathcal{S}_0 = \bar{S} \tag{B.79}$$

$$\mathcal{S}_i = \bar{S} + \left(\sqrt{(L_S + \lambda)P_S}\right)_i, i = 1, ..., L_S \tag{B.80}$$

$$\mathcal{S}_i = \bar{S} - \left(\sqrt{(L_S + \lambda)P_S}\right)_{i-L_S}, i = L_S + 1, ..., 2L_S, \tag{B.81}$$

where $\lambda = \alpha^2(L_S - \kappa) - L_S$ is a scaling parameter governing the spread of sigma points around the mean and $\left(\sqrt{(L_S + \lambda)P_S}\right)_i$ is the i-th column of matrix P_S. Sigma points $\mathcal{S}$ are propagated through function $h(\cdot)$ to get $\mathcal{M}$. The first two moments of m_t are approximated by:

$$\bar{m} \approx \sum_{i=0}^{2L_S} W_i^{\mu} \mathcal{M}_i \tag{B.82}$$

$$P_S \approx \sum_{i=0}^{2L_S} W_i^{\sigma} (\mathcal{M}_i - \bar{m})(\mathcal{M}_i - \bar{m})', \tag{B.83}$$

where W^{μ} and W^{σ} denote weights for the mean and the covariance matrix, respectively and are defined as:

$$W_0^{\mu} = \frac{\lambda}{L_S + \lambda} \tag{B.84}$$

$$W_0^{\sigma} = \frac{\lambda}{L_S + \lambda} + 1 - \alpha^2 + \beta, i = 1, ..., L_S \tag{B.85}$$

$$W_i^{\mu} = W_i^{\sigma} = \frac{\lambda}{2(L_S + \lambda)}, i = L_S + 1, ..., 2L_S. \tag{B.86}$$

Parameters α and β, mainly determine higher moments of the distribution.

The UKF Algorithm

1. Initialize at unconditional moments:[1]

$$\hat{S}_0 = \mathbb{E}[S_0] \tag{B.87}$$

$$P_{S0} = \mathbb{E}[(S_0 - \hat{S}_0)(S_0 - \hat{S}_0)'] \tag{B.88}$$

   for $k \in 1, ..., \infty$:

2. Compute the sigma points:

$$\mathcal{S}_{k-1} = \left[\begin{array}{ccc} \hat{S}_{k-1} & \hat{S}_{k-1} + \sqrt{(L_S + \lambda) P_{S,k-1}} & \hat{S}_{k-1} - \sqrt{(L_S + \lambda) P_{S,k-1}} \end{array} \right] \tag{B.89}$$

3. Time update:

$$\mathcal{S}^a_{k|k-1} = B(\mathcal{S}_{k-1}) \tag{B.90}$$

$$\hat{S}^-_k = \sum_{i=0}^{2L_S} W^\mu_i \mathcal{S}^a_{k|k-1} \tag{B.91}$$

$$P^-_{Sk} = \sum_{i=0}^{2L_S} W^\sigma_i (\mathcal{S}^a_{ik|k-1} - \hat{S}^-_k)(\mathcal{S}^a_{ik|k-1} - \hat{S}^-_k)' + Cov_t(\varepsilon_{t+\Delta t}) \tag{B.92}$$

4. Augment sigma points:

$$\mathcal{S}_{k|k-1} = \left[\begin{array}{ccc} \mathcal{S}^a_{k|k-1} & \mathcal{S}^a_{0k|k-1} + \sqrt{(L_S + \lambda) Cov_t(\varepsilon_{t+\Delta t})} & \mathcal{S}^a_{0k|k-1} - \sqrt{(L_S + \lambda) Cov_t(\varepsilon_{t+\Delta t})} \end{array} \right] \tag{B.93}$$

$$\mathcal{M}_{k|k-1} = h(\mathcal{S}_{k|k-1}) \tag{B.94}$$

$$\hat{m}^-_k = \sum_{i=1}^{2L_S} W^\sigma_i \mathcal{M}_{i,k|k-1} \tag{B.95}$$

5. Measurement equations update:

[1] We borrow the algorithm from Wan and van der Merwe (2001).

$$P^{-}_{m_k} = \sum_{i=0}^{2L_S} W_i^{\sigma}(\mathcal{M}_{ik|k-1} - \hat{m}_k^{-})(\mathcal{M}_{ik|k-1} - \hat{m}_k^{-})' + Cov_t(\vartheta_{t+\Delta t}) \tag{B.96}$$

$$P_{S_k m_k} = \sum_{i=0}^{2L_S} W_i^{\sigma}(\mathcal{S}_{ik|k-1} - \hat{S}_k^{-})(\mathcal{M}_{ik|k-1} - \hat{m}_k^{-})' \tag{B.97}$$

$$\mathcal{K}_k = P_{S_k m_k} P^{-1}_{m_k} \tag{B.98}$$

$$\hat{F}_k = \hat{F}_k^{-} + \mathcal{K}_k(m_k - \hat{m}_k^{-}) \tag{B.99}$$

$$P_k = P_k^{-} - \mathcal{K}_k P^{-}_{m_k} \mathcal{K}_k'. \tag{B.100}$$

Chapter 3

Correlation risk and the term structure of interest rates

We study a completely affine continuous-time yield curve model, in which risk factors are stochastically correlated.[1] The co-movement among factors provides an element of flexibility in modeling the first and second moments of yields. This setting complements the standard affine class of term structure models, and can reconcile several regularities in the dynamics of US Treasury yields.[2] First, excess bonds returns are on average close to zero, but vary systematically with the term structure. The expectations hypothesis is violated in that excess returns can be predicted with yield curve variables—the slope, the spot-forward spread, or linear combination of forward rates. Second, the term structure of forward rate and cap implied volatilities peaks for intermediate maturities, and is moderately downward sloping for longer yields. Third, the term structure of conditional second moments of yields is time-varying, and exhibits a multi-factor structure.[3]

[1]This chapter is based on the paper under the same title written in collaboration with Andrea Buraschi at Imperial College and Fabio Trojani at the University of Lugano. We thank Jaime Casassus, Mikhail Chernov, Qiang Dai, Jerome Detemple, Darrell Duffie, Christian Gourieroux, Christian Julliard, Ilaria Piatti, Paolo Porchia, Peter Gruber, Ken Singleton, Paul Söderlind, Davide La Vecchia, Andrea Vedolin, Pietro Veronesi, and Liuren Wu for their comments, and the participants at the meetings of the Western Finance Association in Big Sky Montana (2007), the European Finance Association in Ljubljana (2007), the Swiss Society of Econometrics and Statistics in St. Gallen (2007), the Adam Smith Asset Pricing Workshop in London (Fall 2007), the 6th Swiss Doctoral Workshop in Gerzensee (2007), Imperial College Financial Econometrics Conference in London (2007), VIII Workshop in Quantitative Finance in Venice (2007).

[2] By "standard affine" we will denote the state space $\mathbb{R}_+^m \times \mathbb{R}^{n-m}$ of a regular affine n-dimensional process (see Duffie, Filipovic, and Schachermayer, 2003).

[3]The literature discussing these features is voluminous. See, among others, Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2005) for the properties of excess bond returns; Amin and Morton (1994), Piazzesi (2001), Leippold and Wu (2003)—for the term structure of volatilities of forward rates, yields and caps;

The complexity of modeling jointly the first and second moments of yields is well illustrated by comparing the term structure reaction to two monetary tightenings by the US Fed which occurred a decade apart, 1994/95 and 2004/05. In both periods, the response at the long end of the curve provided for a surprise, yet for opposing reasons: The first interest rate hike puzzled many observers because long term yields rose, the latter did so because they fell. The difference in co-movement between short and long yields in the two periods is remarkable and hard to rationalize by otherwise successful models (see Figure 3.1 for a summary of the different bond market behavior during these two periods). In search for an economic explanation, various studies have argued about the differences in risk compensation and in the amount of risk, manifest in shifting bond market volatility.[4] The consensus view suggests that while premia on long bonds were high and accompanied by an increased yield volatility in the early episode, they turned low or even negative in the more recent and calmer period. This simple account illustrates both the importance and the challenge of designing yield curve models that combine a flexible specification of bond excess returns with a sufficient time-variation in correlations and volatilities of risk factors.

A vast literature has explored the ability of term structure models to account for the time-series and cross-sectional properties of bond market dynamics. The research has focused on analytically tractable models that ensure economically meaningful behavior of yields and bond returns. This combination of theoretical and empirical requirements poses a significant challenge. In affine term structure models (ATSMs), for instance, the tractability in pricing and estimation comes with restrictions that guarantee admissibility of the underlying state processes and their econometric identification. Dai and Singleton (2000) emphasize that under the risk-neutral measure admissibility implies a trade off between factors' dependence and their stochastic volatilities. In order to have both correlations and time-varying volatilities, positive square-root processes need to be combined with conditionally Gaussian ones. The inclusion of Gaussian dynamics allows for an unconstrained sign of factor correlations, but gives up the ability to accommodate stochastic yield volatilities.

Andersen and Benzoni (2008)—for the dynamic properties of bond volatilities. Dai and Singleton (2003), Piazzesi (2003), and Singleton (2006) provide excellent surveys of the empirical characteristics of interest rates and discuss the ability of affine models to capture them.

[4]See Campbell (1995), Backus and Wright (2007), and Cochrane and Piazzesi (2008) for an interesting account of these tightening events. Rudebusch, Swanson, and Wu (2006) report the failure of two established macro-finance Gaussian models in fitting the 2004/05 "conundrum" period. They ascribe a large portion of the conundrum to declines in long-term bond volatility, and contrary to common argument, find little or no role for foreign official purchases of US Treasuries in explaining the puzzle.

In order to match the physical dynamics of the yield curve, reduced-form models have exploited different specifications of the market price of risk. The "completely affine" literature assumed the market price of risk to be a constant multiple of factor volatility.[5] Summarizing the empirical failure of this specification, Duffee (2002) proposed an "essentially affine" extension, in which the market prices of risk are inversely proportional to factor volatility in the case of unrestricted (conditionally Gaussian) factors and have a switching sign, but preserve the completely affine form for the square-root (volatility) factors. More recently, Cheridito, Filipovic, and Kimmel (2007) suggested an "extended affine" generalization making the market price of risk of all factors, both Gaussian and square-root, inversely proportional to their volatilities.

The advantages obtained by augmenting the market price of risk are twofold. First, by incorporating Gaussian factors, the expected excess bond returns can switch sign. Second, correlations between some factors can take both positive and negative values. This gain in flexibility is crucial for matching the behavior of yields over time. Duffee (2002) and Dai and Singleton (2002) stress the role of correlated factors for the model's ability to forecast yield changes and excess bond returns. They note that despite a good fit to some features of the data, the essentially affine models face the trade-off between stochastic volatilities and correlations of factors.[6] While the extended affine market price of risk helps mitigate some of these tensions, recent research has exposed its restrictions in terms of matching higher order moments of yields (Feldhütter, 2007). Being useful for fitting the data, the extensions of the market price of risk are not innocuous from the perspective of theory and applications. In an equilibrium setting, the market price of risk reflects investor risk attitudes. More complex formulations are therefore equivalent to increasingly complex investor preferences which can be difficult to justify by standard arguments.[7] Gains in empirical performance implied by a richer market price of risk come with new parameters, many of which turn out difficult to identify from yield curve data alone.

[5] Examples of completely affine models are Vasicek (1977), or Cox, Ingersoll, and Ross (1985b). The models have been systematically characterized by Dai and Singleton (2000).

[6] See e.g., Brandt and Chapman (2002); Bansal and Zhou (2002); Dai and Singleton (2003).

[7] Some equilibrium motivation for the essentially affine form of the market price of risk can be found in term structure models with habit formation, like in Dai (2003), and Buraschi and Jiltsov (2006). The extended affine family of Cheridito, Filipovic, and Kimmel (2007), instead, seems difficult to reconcile with the standard expected utility maximization, because it entails that agents become more concerned about risk precisely when it goes away. While potentially inconsistent with standard preferences, such behavior can arise quite naturally in an economy with ambiguity aversion as demonstrated by Gagliardini, Porchia, and Trojani (2007). In their non-affine term structure model, the contribution of the ambiguity premium to excess bond returns dominates the standard risk premia precisely when the aggregate risk in the economy is low.

Rather than focussing on market prices of risk, we start with a non-standard assumption about the state space. Specifically, we assume the risk factors in the economy to follow a continuous-time affine process of positive definite matrices whose transition probability is Wishart. By construction, such factors allow correlations and volatilities both to be stochastic. This approach builds on the work of Gourieroux and Sufana (2003) who propose the Wishart process as a convenient theoretical framework to represent yield factors. We start from their insight and take the first attempt to investigate the properties of a continuous-time Wishart yield curve model. Exploring the properties of the state space, we do not introduce a new form of the market price of risk, but we resurrect the parsimonious completely affine class. We study the properties of the term structure, bond returns and standard interest rate derivatives, and document the ability of this setting to match several features of the data.

First, our completely affine market price of risk specification involves elements that take both positive and negative values. Hence, it translates into excess bond returns that can switch sign. We find that the variation in the model-implied term premia is consistent in size and direction with the historical deviations from the expectations hypothesis.

Second, the model lessens the volatility-correlation trade-off. Both under the physical and risk-neutral measures, it accommodates switching sign of conditional and unconditional correlations among state variables. We show that the model-generated yields bear a degree of mutual co-movement and volatility persistence which is compatible with historical evidence.

Third, the framework is interesting from the derivatives pricing perspective. We find that it produces a hump-shaped term structure of forward interest rate and cap implied volatilities. The direct presence of stochastic correlations in the state dynamics, both under the risk-neutral and physical measure, does not require to introduce them with an additional independent state variable.

Finally, the Wishart setting affords a direct extension to a larger dimension at no harm to analytical solutions. With the state space enlarged to six variables, the framework additionally supports the single return-forecasting factor of Cochrane and Piazzesi (2005), is able to incorporate realistic dynamic correlations of yield changes, and the comovement of yield volatilities. Parallel to providing a comprehensive description of the spot yield curve, the enlarged model also leads to an appearance of unspanned factors in interest rate derivatives. The flexibility gained with extra factors does not seriously impair the parsimony as the six-factor setting requires fewer parameters than the standard three-factor specifications.

In our analysis, we exploit several results on the non-central Wishart distribution. Even though the statistical properties of this distribution have been extensively covered in the early multivariate statistics literature as summarized for instance by Muirhead (1982), the extension involving time dependence has only been explored of late. The Wishart process has been first proposed by Bru (1991), and more recently studied by Donati-Martin, Doumerc, Matsumoto, and Yor (2004) and Gourieroux (2006) in continuous time and Gourieroux, Jasiak, and Sufana (2009) in discrete time. This work has provided a foundation for such applications to finance as derivatives pricing or portfolio choice with stochastically correlated assets (see e.g., Gourieroux and Sufana, 2004; Buraschi, Porchia, and Trojani, 2009; da Fonseca, Grasselli, and Tebaldi, 2006).

The plan of the paper is as follows. Section 3.1 introduces our completely affine market price of risk specification, derives the short interest rate, provides the solution for the term structure and discusses its asset pricing implications. Section 3.2 presents the empirical approach, derives the moments of the Wishart process and yields, and investigates the properties of factors based on the model fitted to the US Treasury zero curve. In Section 3.3, the most simple three-factor specification is scrutinized for its consistency with the stylized facts of yield curve literature. Section 3.4 extends the discussion to a six-factor framework and highlights the features of the enlarged model. Section 3.5 concludes. All proofs and figures are in Appendices.

3.1 The Economy

In analogy to the standard completely affine models, we motivate a parsimonious form of the market price of risk within the general Cox, Ingersoll, and Ross (1985a) framework.

Assumption 1 (Preferences)**.** *The representative agent maximizes an infinite horizon utility function:*

$$E_t\left[\int_t^\infty e^{-\rho(s-t)}\ln(C_s)ds\right], \tag{3.1}$$

where $E_t(\cdot)$ is the conditional expectations operator, ρ is the time discounting factor, and C_t is consumption at time t.

We depart from the affine literature only in the specification of risk factors driving the production technology dynamics. These are assumed to follow an affine continuous-time process of symmetric positive definite matrices.

Assumption 2 (Production technology)**.** *The return to the production technology evolves as:*

$$\frac{dY_t}{Y_t} = Tr\left(D\Sigma_t\right)dt + Tr\left(\sqrt{\Sigma_t}dB_t\right), \tag{3.2}$$

where dB_t is a $n \times n$ matrix of independent standard Brownian motions; Σ_t is a $n \times n$ symmetric positive definite matrix of state variables, and $\sqrt{\cdot}$ denotes the square root in the matrix sense; D is a symmetric $n \times n$ matrix of deterministic coefficients. Tr indicates the trace operator.

In the above diffusion, the drift $\mu_Y = Tr(D\Sigma_t)$ and the quadratic variation $\sigma_Y^2 = Tr(\Sigma_t)$ are of affine form. Therefore, the univariate process of returns $\frac{dY_t}{Y_t}$ is affine in $\frac{n(n+1)}{2}$ distinct elements of the symmetric Σ_t matrix.

Assumption 3 (Risk factors)**.** *The physical dynamics of the risk factors are governed by the Wishart process Σ_t, given by the matrix diffusion system:*

$$d\Sigma_t = \left(\Omega\Omega' + M\Sigma_t + \Sigma_t M'\right)dt + \sqrt{\Sigma_t}dB_t Q + Q'dB_t'\sqrt{\Sigma_t}, \tag{3.3}$$

where Ω, M, Q, Ω invertible, are square $n \times n$ matrices. Throughout, we assume that $\Omega\Omega' = kQ'Q$ with integer degrees of freedom $k > n-1$, ensuring the Σ_t matrix is of full rank.

The Wishart process is a multivariate extension of the well-known square-root (CIR) process. In a special case when k is an integer, it can be interpreted as a sum of outer products of k independent copies of an Ornstein-Uhlenbeck process, each with dimension n. A number of qualities make the process particularly suitable for modeling multivariate sources of risk in finance (Gourieroux, Jasiak, and Sufana, 2009; Gourieroux, 2006). First, the conditional Laplace of the Wishart and the integrated Wishart process are exponential affine in Σ_t. As such, these processes are affine in the sense of Duffie, Filipovic, and Schachermayer (2003). This feature gives rise to convenient closed-form solutions to prices of bonds or options, and simplifies the econometric inference. Second, the Wishart process lives in the space of positive definite matrices. Thus, it is distinct from the affine processes defined on $\mathbb{R}_+^m \times \mathbb{R}^{n-m}$. The restriction $\Omega\Omega' \gg Q'Q$ guarantees that Σ_t is positive semi-definite. By additionally assuming that $k > n-1$, it is ensured that the state matrix is of full rank. Thus, the diagonal elements of Σ_t (and $\sqrt{\Sigma_t}$) are always positive, but the out-of-diagonal elements can take on negative values. Moreover, if $\Omega\Omega' = kQ'Q$ for some $k > n-1$ (not necessarily an integer), then Σ_t follows the Wishart distribution (Muirhead, 1982, p. 443). Third, the elements of the Wishart matrix feature rich dependence structure, and their conditional and unconditional correlations are unrestricted in sign (see Section 3.2 and Result 1 in Appendix C.4 for details). This feature implies a term structure setting with weak restrictions on

the stochastic cross-sectional dependence between yields. Finally, the coefficients of the dynamics (3.3) admit intuitive interpretation: The M matrix is responsible for the mean reversion of factors, and the Q matrix—for their conditional dependence. Typically, in order to ensure non-explosive features of the process, M is assumed negative definite.

3.1.A The Short Interest Rate and the Market Price of Risk

Since we are primarily interested in exploring the term structure implications of the state dynamics (3.3), we maintain the simple completely affine form of the market price of risk implied by our assumptions. Given Assumption 1, the optimal consumption plan of the representative agent is $C_t^* = \rho Y_t$, and the short interest rate can be computed by the equation $r_t = -\mathbb{E}_t\left[\frac{du'(C_t^*)}{u'(C_t^*)}\right]$, with $u(C) = \ln(C)$. The $n \times n$ matrix of market prices of risk Λ_t follows as the unique solution of the equation $\mu_Y - r_t = Tr\left[\sqrt{\Sigma_t}\Lambda_t\right]$. A standard application of Itô's Lemma implies the following expressions for r_t and Λ_t.

Proposition 1 (The short interest rate and the market price of risk)**.** *Under Assumptions 1–3, the short interest rate is given by:*

$$r_t = Tr\left[(D - I_n)\Sigma_t\right], \tag{3.4}$$

where I_n is an $n \times n$ identity matrix. The market price of risk equals the square root of the matrix of Wishart factors:

$$\Lambda_t = \sqrt{\Sigma_t}. \tag{3.5}$$

The short interest rate in (3.4) can be equivalently written as:

$$r_t = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}\Sigma_{ij,t},$$

where d_{ij} denotes the ij-th element of matrix $D - I_n$. Thus, r_t is a linear combination of the Wishart factors, which are conditionally and unconditionally dependent, with correlations being unrestricted in sign.[8] While the short rate comprises both positive factors on the diagonal of Σ_t and out-of-diagonal factors that can take both signs, its positivity is ensured with a restriction

[8] By restricting D in equation (3.2) to be a diagonal 2×2 matrix, this expression for the short interest rate resembles the two-factor model of Longstaff and Schwartz (1992). Even in this special case, however, the model has a richer structure, as the variance of the changes in the interest rate is driven by all three factors $(\Sigma_{11}, \Sigma_{12}, \Sigma_{22})$, which are pairwise correlated.

that the $D-I_n$ matrix be positive definite (see Result 4 in Appendix C.4). In Section 3.1.D below, we prove that the same condition is necessary and sufficient to impose the positivity on the whole term structure of interest rates. The instantaneous variance V_t of changes in the short rate is obtained by applying Ito's Lemma to equation (3.4), and computing the quadratic variation of the process dr, i.e. $V_t = \langle dr \rangle_t$. Similarly, the instantaneous covariance CV_t between the changes in the short rate and their variance can be found as the quadratic co-variation of dr and dV, i.e. $CV_t = \langle dr, dV \rangle_t$. The instantaneous variance of the interest rate changes is given by:

$$\frac{1}{dt}V_t = 4Tr\left[\Sigma_t(D-I_n)Q'Q(D-I_n)\right]. \tag{3.6}$$

The covariance between changes in the level and changes in the volatility of the interest rate becomes:

$$\frac{1}{dt}CV_t = 4Tr\left[\Sigma_t(D-I_n)Q'Q(D-I_n)Q'Q(D-I_n)\right]. \tag{3.7}$$

Expressions (3.6) and (3.7) show that both quantities preserve the affine form in the elements of Σ_t. Appendix C.1.A provides the derivation.

A well-recognized critique of the completely affine models is their inability to match the empirical properties of yields due to (i) the sign restriction on the market price of risk, and (ii) its one-to-one link with the volatility of factors. Despite analogous derivation, the market price of risk in (3.5) is distinct from the standard completely affine specification in the sense that it reflects not only volatilities but also co-volatilities of factors, and thus involves elements that can change sign. Positive factors on the diagonal of matrix Σ are endowed with a positive market price of risk, while the market price of risk of the remaining out-of-diagonal factors is unrestricted. Intuitively, the signs of the elements of Λ_t reflect different perceptions of volatility and co-volatility risks by investors.

Remark 2 (Extended Wishart specification of the market price of risk)**.** In a reduced-form Wishart factor model, one could easily construct a richer market price of risk à la Cheridito, Filipovic, and Kimmel (2007):

$$\Lambda_t = {\Sigma_t}^{-1/2}\Lambda_0 + {\Sigma_t}^{1/2}\Lambda_1,$$

for some $n \times n$ constant matrices Λ_0 and Λ_1. This specification preserves the affine property of the process (3.3) under the risk neutral measure, but modifies both the constant and the mean

reversion terms in its drift. In order to maintain the Wishart distribution under the risk neutral measure, one can set $\Lambda_0 = vQ'$ for $v \in \mathbb{R}_+$ such that $k - 2v > n - 1$.[9] $\square$

3.1.B The Term Structure of Interest Rates

Given expression (3.4) for the short rate, the price at time t of a zero-coupon bond maturing at time T is:

$$P(t,T) = E_t^* \left(e^{-\int_t^T r_s ds} \right) = E_t^* \left(e^{-Tr\left[(D - I_n)\int_t^T \Sigma_s ds\right]} \right), \tag{3.8}$$

where $E_t^*(\cdot)$ denotes the conditional expectation under the risk neutral measure. To move from physical dynamics of the Wishart state in equation (3.3) to risk neutral dynamics, we can apply the standard change of drift technique. Due to the completely affine market price of risk specification, the risk neutral drift adjustment of $d\Sigma_t$ takes on a very simple form: $\Phi_\Sigma = \Sigma Q + Q'\Sigma$.[10]

Remark 3. This form of the drift adjustment is straightforward to justify in a reduced form setting by defining the Radon-Nikodym derivative for the transformation from the physical measure $\tilde{\mathbb{Q}}$ to the risk neutral measure $\mathbb{Q}^*$:

$$\frac{d\mathbb{Q}^*}{d\tilde{\mathbb{Q}}}|_{\mathcal{F}_t} := e^{Tr\left[-\int_0^t \Lambda_u' dB_u - \frac{1}{2}\int_0^t \Lambda_u'\Lambda_u du\right]}, \tag{3.9}$$

where $\Lambda_t = \sqrt{\Sigma_t}$ is our completely affine market price of risk. It follows from the Girsanov's theorem that the process $B_t^* = B_t + \int_0^t \Lambda_u du$ is an $n \times n$ matrix of standard Brownian motions under $\mathbb{Q}^*$. Therefore, the SDE

$$d\Sigma_t = \left(\Omega\Omega' + (M - Q')\Sigma_t + \Sigma_t(M' - Q)\right) dt + \sqrt{\Sigma_t} dB_t^* Q + Q' dB_t^{*\prime} \sqrt{\Sigma_t}$$

represents the risk neutral dynamics of the Wishart process. $\square$

Given the change of drift Φ_Σ, the term structure pricing equation follows by applying the discounted Feynman-Kač formula to expectation (3.8).

[9] For practical applications, this form of Λ_t requires a careful analysis of conditions ensuring no arbitrage. In the classic affine case, Cheridito, Filipovic, and Kimmel (2007) show that the extended affine market price of risk does not admit arbitrage provided that under both measures the state variables cannot achieve their boundary values. In the Wishart factor setting, this condition holds true if the Q matrix is invertible, and $k - 2v > 1$.

[10] As in the scalar case, the elements of Φ_Σ can be interpreted as the expected excess returns on securities constructed so that they perfectly reflect the risk embedded in the corresponding elements of the state matrix Σ.

Proposition 2 (The pricing PDE)**.** *Under Assumptions 1–3, the price at time t of a contingent claim F maturing at time $T > t$, whose value is independent of wealth, satisfies the partial differential equation:*[11]

$$Tr\{\left[\Omega\Omega' + (M - Q')\Sigma + \Sigma(M' - Q)\right]\mathcal{R}F\} + 2Tr\left[\Sigma\mathcal{R}(Q'Q\mathcal{R}F)\right] + \frac{\partial F}{\partial t} - Tr\left[(D - I_n)\Sigma\right]F = 0, \tag{3.10}$$

with the boundary condition:

$$F(\Sigma, T, T) = \Psi(\Sigma, T), \tag{3.11}$$

where $\mathcal{R}$ is a matrix differential operators with $ij-th$ component equal to $\frac{\partial}{\partial\Sigma_{ij}}$.

In expression (3.8), we recognize the Laplace transform of the integrated Wishart process. By the affine property of the process, the solution to the PDE in equation (3.10) is an exponentially affine function of risk factors. The next proposition states this result in terms of prices of zero-coupon bonds.

Proposition 3 (Bond prices)**.** *If there exists a unique continuous solution to the Wishart equation, then under the model dynamics (3.2)–(3.3), the price at time t of a zero-coupon bond P with maturity $T > t$, is of the exponentially affine form:*

$$P(\Sigma, t, T) = e^{b(t,T) + Tr[A(t,T)\Sigma]}, \tag{3.12}$$

for a state independent scalar $b(t, T)$, and a symmetric matrix $A(t, T)$ solving the system of matrix Riccati equations:

$$-\frac{db(t,T)}{dt} = Tr\left[\Omega\Omega' A(t,T)\right] \tag{3.13}$$

$$-\frac{dA(t,T)}{dt} = A(t,T)(M - Q') + (M' - Q)A(t,T) + 2A(t,T)Q'QA(t,T) - (D - I_n) \tag{3.14}$$

with terminal conditions $A(T, T) = 0$ and $b(T, T) = 0$. Letting $\tau = T - t$, the closed-form solution to (3.14) is given by:

$$A(\tau) = C_{22}(\tau)^{-1}C_{21}(\tau),$$

where $C_{12}(\tau)$ and $C_{22}(\tau)$ are $n \times n$ blocks of the following matrix exponential:

[11] See Bru (1991) equation (5.12) for the infinitesimal generator of the Wishart process.

$$\begin{pmatrix} C_{11}(\tau) & C_{12}(\tau) \\ C_{21}(\tau) & C_{22}(\tau) \end{pmatrix} := \exp\left[\tau \begin{pmatrix} M - Q' & -2Q'Q \\ -(D - I_n) & -(M' - Q) \end{pmatrix}\right].$$

Given the solution for $A(\tau)$, the coefficient $b(\tau)$ is obtained directly by integration:

$$b(\tau) = Tr\left(\Omega\Omega' \int_0^\tau A(s)\right) ds.$$

Proof: See Appendix C.1.B. ■

Under the assumed factor dynamics, bond prices are given in closed form. The solution for $A(\tau)$ implied by the matrix Riccati ODE (3.14) has been known since the work of Radon in 1929. The general (non-symmetric) case has been discussed by Levin (1959). The matrix form of the coefficients facilitates the characterization of the definiteness and monotonicity of the solution, given in the next corollary.

Corollary 1 (Definiteness and monotonicity of the solution)**.** *If matrix $A(\tau)$ is the solution to the matrix Riccati equation (3.14), then $A(\tau)$ is negative definite and monotonically decreasing for all $\tau \in [0, T]$, i.e. $A(\tau) < 0$ and $A(\tau_2) < A(\tau_1)$ for $\tau_2 > \tau_1$, if and only if $D - I_n$ is positive definite, $D - I_n > 0$.*

Proof: See Appendix C.1.B. ■

3.1.C Yields

From equation (3.12), we obtain the yield of a zero-bond maturing in $\tau = T - t$ periods:

$$y_t^\tau = -\frac{1}{\tau}\left[b(\tau) + Tr\left(A(\tau)\Sigma_t\right)\right]. \tag{3.15}$$

The affine property of yields in the elements of Σ_t allows the relation (3.15) to be uniquely inverted for the factors. In particular, the symmetric $n \times n$ state matrix can be identified from $\frac{n(n+1)}{2}$ yields (see Appendix C.2.A for details on the inversion). Moreover, to ensure positive yields, the $A(\tau)$ matrix needs to be negative definite, for which it is both necessary and sufficient that $D - I_n > 0$. This simple positivity condition together with unrestricted factor correlations is different from the traditional ATSMs, in which a positive short rate is not ensured in the presence of unrestricted and correlated Gaussian factors. In the classification of Dai and Singleton (2000), $A_N(N)$ is the only subfamily of standard ATSMs that guarantees the positivity of yields. In $A_N(N)$ models, all state variables determine the volatility structure of factors, and thus remain instantaneously

uncorrelated. At the same time, the restrictions imposed on the mean reversion matrix require the unconditional correlations among state variables to be non-negative.[12]

Finally, the modeling of covariances between yields is a nontrivial issue in applications such as bond portfolio selection. Therefore, it is important to understand their properties arising in the Wishart setting. The instantaneous covariance of the changes in yields with different (but fixed) time to maturity (τ_1, τ_2) becomes:

$$\frac{1}{dt} Cov_t \left[dy_t^{\tau_1}, dy_t^{\tau_2} \right] = \frac{4}{\tau_1 \tau_2} Tr \left[A(\tau_1) \Sigma_t A(\tau_2) Q' Q \right]. \tag{3.16}$$

The model implies that yields of different maturities co-vary in a non-deterministic and multivariate way, evident in the presence of Σ_t in the above equation. Given the indefiniteness of matrix $A(\tau_1)Q'QA(\tau_2)$, the correlations of yields can stochastically change sign over time. The secular decline in long yields during 2004/05 tightening period (see Figure 3.1) provides one example of an interest rate environment, in which such feature can be important. While new empirical evidence emphasizes the presence of multiple factors in second moments of yields (Andersen and Benzoni, 2008), this property cannot be captured by the univariate volatility $A_1(N)$ specification. We investigate the consequences of the multifactor volatility structure in the Wishart setting in Section 3.4.B.

Remark 4 (Relationship to quadratic term structure models)**.** As pointed out by Gourieroux and Sufana (2003), in a special case the yield curve in equation (3.15) collapses to an n-factor quadratic term structure model (QTSM) in the spirit of Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2002). With one degree of freedom ($k = 1$) in the factor dynamics (3.3), the state becomes a singular matrix of rank one, $\Sigma_t = X_t^1 X_t^{1\prime}$, where X_t^1 is an n-dimensional Ornstein-Uhlenbeck (OU) process (see Appendix C.2 for details). Using the fact that $Tr\left[A(\tau)X_t^1 X_t^{1\prime}\right] = X_t^{1\prime} A(\tau) X_t^1$, a purely quadratic expression for yields emerges:

$$y_t^\tau = -\frac{1}{\tau}\left[b(\tau) + X_t^{1\prime} A(\tau) X_t^1 \right].$$

[12] Feldhütter (2007) shows that the negativity of yields in ATSMs can indeed become a concern. For instance, in the Gaussian $A_0(3)$ model, which offers a maximal flexibility in modeling correlations of factors, the probability of negative 1-year (5-year) yields amounts to non-negligible 5.98 (3.91) percent. The history of US nominal yields makes this probability look high in comparison. Indeed, until the 2008 credit crisis negative nominal yields in the US remained very much a theoretical concept.

Apart from the state degeneracy, we show that this special case places important limitations on the form of yield correlations. Specifically when $k = 1$, correlations between the diagonal elements $\Sigma_{ii,t}, \Sigma_{jj,t}$, $i \neq j$, turn out to be piecewise constant and take plus/minus the same value (see Result 2 in Appendix C.4). To see this, note that:

$$Corr_t\left(\Sigma_{ii}, \Sigma_{jj}\right) = \frac{\langle \Sigma_{ii}, \Sigma_{jj} \rangle_t}{\sqrt{\langle \Sigma_{ii} \rangle_t}\sqrt{\langle \Sigma_{jj} \rangle_t}} = \text{sgn}(X^1_{i,t} X^1_{j,t}) \frac{Q^{i\prime} Q^j}{\sqrt{Q^{i\prime} Q^i}\sqrt{Q^{j\prime} Q^j}},$$

where $X^1_{i,t}, X^1_{j,t}$ are scalar OU processes given by the elements of vector X^1_t, Q^i, Q^j denote the i-th and j-th column of the Q matrix, respectively, and sgn is the sign function. By studying the more general non-degenerate case we achieve two main goals.[13] First, the out-of-diagonal elements of the state matrix become non-trivial factors with unrestricted sign. This feature allows to capture the predictability of bond returns. Second, since factor correlations are allowed to take any value between -1 and 1, they provide additional flexibility in modeling the stochastic second moments of yields. □

3.1.D Excess Bond Returns

The price dynamics of a zero-coupon bond follow from the application of Ito's Lemma to $P(\Sigma_t, \tau)$:

$$\frac{dP(\Sigma_t, \tau)}{P(\Sigma_t, \tau)} = (r_t + e^\tau_t)dt + Tr\left[\left(\sqrt{\Sigma_t} dB_t Q + Q' dB_t' \sqrt{\Sigma_t}\right) A(\tau)\right], \tag{3.17}$$

where e^τ_t is the term premium (instantaneous expected excess return) to holding a τ-period bond (see Appendix C.1.C). The functional form of the expected excess return can be inferred from the fundamental pricing PDE (3.10), and is represented by a linear combination of the Wishart factors:

$$e^\tau_t = Tr\left[(A(\tau)Q' + QA(\tau))\Sigma\right] = 2Tr\left[\Sigma_t Q A(\tau)\right]. \tag{3.18}$$

The instantaneous variance of bond returns can be written as:

$$v^\tau_t = 4Tr\left[A(\tau)\Sigma_t A(\tau) Q' Q\right]. \tag{3.19}$$

[13] In fact, when we estimate the model with $k = 1$ its overall fit deteriorates considerably compared to the non-degenerate case.

A stylized empirical observation is that excess returns on bonds are on average close to zero, but vary broadly taking both positive and negative values.[14] This means that the mean ratio $\frac{e_t^\tau}{\sqrt{v_t^\tau}}$ is low for all maturities τ. The essentially and extended affine models of Duffee (2002) and Cheridito, Filipovic, and Kimmel (2007) are able to replicate this empirical regularity by assigning to non-volatility factors a market price of risk that can change sign. We can generate excess returns that have a switching sign if the symmetric matrix $A(\tau)Q' + QA(\tau)$ premultiplying Σ in equation (3.18) is indefinite. The set of matrices (and model parameters) satisfying this condition is large, giving us the latitude to capture the combination of low expected excess returns on bonds with their high volatilities. Using estimation results, in Sections 3.3.A and 3.3.B we study the properties of model-implied excess bond returns.

3.1.E Forward Interest Rate

Let $f(t,T)$ be the instantaneous forward interest rate at time t for a contract beginning at time $T = t + \tau$. The instantaneous forward rate is defined as $f(t,T) := -\frac{\partial \ln P(t,T)}{\partial T}$. Taking the derivative of the log-bond price in equation (3.12), we have:

$$f(t,T) = -\frac{\partial b(\tau)}{\partial \tau} - Tr\left(\frac{\partial A(\tau)}{\partial \tau}\Sigma_t\right),$$

where $\partial A(\tau)/\partial\tau$ denotes the derivative with respect to the elements of the matrix $A(\tau)$ given in equation (3.14). The dynamics of the forward rate are given by (see Appendix C.1.D):

$$df(t,T) = -\frac{\partial f(t,T)}{\partial \tau}dt - Tr\left(\frac{\partial A(\tau)}{\partial \tau}d\Sigma_t\right). \tag{3.20}$$

De Jong, Driessen, and Pessler (2004), for example, argue that a humped shape in the volatility term structure of the instantaneous forward rate leads to possible large humps in the implied volatility curves of caplets and caps that are typically observed in the market. We examine the magnitude and sources of the hump in the model-implied volatility of the forward interest rate in Section 3.3.D.

[14] See e.g. Figure 5 in Piazzesi (2003).

3.1.F Interest Rate Derivatives

Our framework allows us to derive convenient expressions for the prices of simple interest rate derivatives. The price of a call option with strike K and maturity S written on a zero bond maturing at $T \geq S$ is:

$$\begin{aligned} \text{ZBC}(t, \Sigma_t; S, T, K) &= E_t^* \left[e^{-\int_t^S r_u du} (P(S,T) - K)^+ \right] \\ &= P(t,T) Pr_t^T \{P(S,T) > K\} - K P(t,S) Pr_t^S \{P(S,T) > K\}, \end{aligned}$$

where $Pr_t^T\{X\}$ denotes the conditional probability of the event X (exercise of the option) based on the forward measure related to the T-maturity bond. We can take the logarithm to obtain:

$$Pr_t^T \{P(S,T) > K\} = Pr_t^T \{b(S,T) + Tr\left(A(S,T)\Sigma_S\right) > \ln K\}.$$

To solve for the option price, we need to determine the conditional distribution of the log bond price under the S- and T-forward measures. In our framework, the characteristic function of the log bond price—due to the affine property in Σ—is available in closed form. Thus, we can readily apply the techniques developed in Heston (1993), Duffie, Pan, and Singleton (2000), and Chacko and Das (2002). The pricing of the option amounts to performing two one-dimensional Fourier inversions under the two forward measures.

Proposition 4 (Zero-coupon bond call option price)**.** *Under Assumptions 1–3, the time-t price of an option with strike K, expiring at time S, written on a zero-bond with maturity $T \geq S$ can be computed by Fourier inversion according to:*

$$\begin{aligned} ZBC(t,S,T,K) = P(t,T) &\left\{ \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re \frac{e^{-iz[\log K - b(S,T)]} \Psi_t^T(iz)}{iz} dz \right\} \\ - K P(t,S) &\left\{ \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re \frac{e^{-iz[\log K - b(S,T)]} \Psi_t^S(iz)}{iz} dz \right\}, \end{aligned}$$

where $\Psi_t^j(iz)$, $j = S, T$, are characteristic functions of $Tr\left[A(S,T)\Sigma_S\right]$ under the S- and T-forward measure, respectively. Details and closed-form expressions for the characteristic function are provided in Appendix C.1.E. ■

The price of the corresponding put bond option can be obtained by the following put call parity relation:

$$\text{ZBP}(t,S,T,K) - \text{ZBC}(t,S,T,K) = KP(t,S) - P(t,T).$$

With these results at hand, we can price interest rate caps and floors, which are respectively portfolios of put and call options on zero-bonds.

3.2 The Model-Implied Factor Dynamics

In this and the following sections, we are guided by the criteria laid down by Dai and Singleton (2003) and study how the completely affine Wishart yield curve model corresponds to the historical behavior of the term structure of interest rates. The model is scrutinized for its ability to match: (i) the predictability of yields, (ii) the persistence of conditional volatilities of yields, (iii) the correlations between different segments of the yield curve, and (iv) the behavior of interest rate derivatives.

To study the ability of the Wishart model to match the criteria (i)–(iv), we begin with the most simple 2×2 specification ($k = 3$) of the state matrix (3.3) and the completely affine market price of risk (3.5). Effectively, we work in a three-factor setting, with two positive factors Σ_{11}, Σ_{22}, and one factor Σ_{12} that can change sign. We ask the model to simultaneously match both conditional and unconditional properties of yields. Due to the choice of a low dimensional state matrix, after imposing the identification restrictions, the estimated 2×2 model has 9 parameters to perform tasks (i)–(iv) mentioned above (compared to 12 parameters of the completely affine CIR with three independent factors).

Below, we outline our econometric approach based on closed-form moments of yields, discuss the model identification, and review the features of the underlying state space implied by the estimated parameters. Proofs are delegated to Appendix C.3.

3.2.A Empirical approach

We use end-of-month data on zero-coupon US Treasury bonds for the period from January 1952 through June 2005. The sample includes the following maturities: 3 and 6 months, 1, 2, 3, 5, 7 and 10 years. Yields for the period from January 1952 through December 1969 are from McCulloch and Kwon data set; yields from January 1970 through December 1999 are from the

Fama and Bliss CRSP tapes. For the last period from January 2000 through June 2005, we use yields compiled by Gürkaynak, Sack, and Wright (2006).[15]

Moments of yields

The estimation of the parameter vector θ comprising the elements of matrices M, Q and D can be posed as the method of moments by setting:

$$\hat{\theta}_T = \arg\min_{\theta} \|\hat{\mu}_T - \mu(\theta)\|,$$

where $\hat{\mu}_T$ represents a vector-valued function of empirical moments based on the historical yields with different maturities, and $\mu(\theta)$ is its theoretical counterpart obtained from the model. The function μ involves the first and second (cross)-moments of yields with different maturities. We estimate the parameter vector θ by using moment conditions that provide both stationary and dynamic information about the term structure. The moments that provide a stationary description of the term structure comprise means, standard deviations and correlations of yields. The conditional information is introduced by augmenting the set of moment conditions with Campbell-Shiller regression coefficients. More specifically, in the 2×2 case we use the unconditional moments of yields with maturities 6 months, 2 years, and 10 years, and the Campbell-Shiller regression coefficients for the 2- and 10-year yields. In estimating the 3×3 specification, we further expand this set with correlations of yield changes and forward rate volatilities, and by adding the 5-year yield. This leaves us with 11 and 20 moment restrictions for the 2×2 and 3×3 model, respectively.

Affine expressions for the term structure facilitate the computation of the theoretical moments of yields as a function of the moments of the Wishart state variable. For brevity, we only provide the unconditional mean and covariance:

$$E(y_t^\tau) = -\frac{1}{\tau}\left\{b_\tau + Tr\left[A_\tau E\left(\Sigma_t\right)\right]\right\}$$
$$Cov(y_{t+s}^{\tau_1}, y_t^{\tau_2}) = \frac{1}{\tau_1\tau_2}\text{vec}\left(\Phi_s' A_{\tau_1}\Phi_s\right)\left[Cov(\text{vec}\Sigma_t)\right]\text{vec}(A_{\tau_2}),$$

where $Cov(\text{vec}\Sigma_t) = E\left[\text{vec}\Sigma_t(\text{vec}\Sigma_t)'\right] - \text{vec}E\Sigma_t\left(\text{vec}E\Sigma_t\right)'$.

[15] The sample is an extension of the one used in Duffee (2002). The three sources use different filtering procedures, thus yields they report for the overlapping period do not match exactly. However, the descriptive statistics for yields of Gürkaynak, Sack, and Wright (2006) are consistent with the Fama-Bliss data set for the overlapping part of both samples.

In that the moments of the Wishart process, stated in the next lemma, are particularly simple and available in fully closed form, the computation of the moments of the term structure becomes a straightforward task.

Lemma 2 (Conditional moments of the Wishart process)**.** *Given the Wishart process* (3.3) *of dimension* n *with* k *degrees of freedom, the first and the second conditional moments of* $\Sigma_{t+\tau}|\Sigma_t$ *are of the form, respectively:*

$$E(\Sigma_{t+\tau}|\Sigma_t) = \Phi_\tau \Sigma_t \Phi_\tau' + kV_\tau, \tag{3.21}$$

and

$$\begin{aligned} E[\mathrm{vec}\Sigma_{t+\tau}\left(\mathrm{vec}\Sigma_{t+\tau}\right)'|\Sigma_t] = {} & \mathrm{vec}(\Phi_\tau \Sigma_t \Phi_\tau' + kV_\tau)\left[\mathrm{vec}(\Phi_\tau \Sigma_t \Phi_\tau' + kV_\tau)\right]' \\ & + (I_{n^2} + K_{n,n})\left[\Phi_\tau \Sigma_t \Phi_\tau' \otimes V_\tau + k(V_\tau \otimes V_\tau) + V_\tau \otimes \Phi_\tau \Sigma_t \Phi_\tau'\right], \end{aligned} \tag{3.22}$$

where I_{n^2} *is the* $n^2 \times n^2$ *identity matrix,* $K_{n,n}$ *denotes the* $n^2 \times n^2$ *commutation matrix,* $\otimes$ *is the Kronecker product, and* vec *denotes vectorization. Matrices* Φ_τ *and* V_τ *are given as:*

$$\begin{aligned} \Phi_\tau &= e^{M\tau} \\ V_\tau &= \int_0^\tau \Phi_s Q'Q\Phi_s' ds. \end{aligned} \tag{3.23}$$

Proof and the closed-form expression for the integral (3.23) *are detailed in Appendix C.3.* ■

The stationarity of the state variables requires the M matrix to be negative definite, in which case the unconditional moments of Σ_t are readily available:[16]

$$\begin{aligned} E\left(\Sigma_t\right) &= kV_\infty \\ E\left[\mathrm{vec}\Sigma_t\left(\mathrm{vec}\Sigma_t\right)'\right] &= k^2 \mathrm{vec}V_\infty\left(\mathrm{vec}V_\infty\right)' + k(I_{n^2} + K_{n,n})\left(V_\infty \otimes V_\infty\right), \end{aligned}$$

where V_∞ can be efficiently computed as:

$$\mathrm{vec}V_\infty = \mathrm{vec}\left(\lim_{\tau\to\infty}\int_0^\tau \Phi(s)Q'Q\Phi'(s)ds\right) = -\left[(I_n \otimes M) + (M \otimes I_n)\right]^{-1}\mathrm{vec}(Q'Q)$$

by exploiting the link between the matrix integral and the Lyapunov equation $MX + XM' = Q'Q$ (see Appendix C.3.A).

[16] Clearly, the negative definiteness of M matrix ensures that $\lim_{\tau\to\infty}\Phi(\tau) = 0$.

Model identification

The unobservability of the state spells out the possibility that two distinct models lead to distributionally equivalent yields. Therefore, the identification of the parameters becomes a concern for empirical applications. A simple way to characterize the identification restrictions in our setting is to exploit its link to QTSMs. In the general case of an integer $k > 1$, the Wishart factor model shows analogy to an $(n \cdot k)$-factor "super-quadratic" model without a linear term and with repeated parameters. To see this, stack k n-dimensional OU processes X and the corresponding Brownian motions W as $Z_t = (X_t^{1\prime}, X_t^{2\prime}, \dots X_t^{k\prime})'$ and $W_t = (W_t^{1\prime}, W_t^{2\prime}, \dots W_t^{k\prime})'$. Then, the factors, the short rate and the market price of risk can be recast in a vector form as:

$$\begin{aligned} dZ_t &= (I_k \otimes M) Z_t dt + (I_k \otimes Q') dW_t \\ r_t &= Z_t' \left[I_k \otimes (D - I_n) \right] Z_t \\ \Lambda_t &= Z_t, \end{aligned}$$

where $\otimes$ denotes the Kronecker product. This model implies a risk neutral dynamics for the state variables of the form $dZ_t = \left[I_k \otimes (M - Q') \right] Z_t dt + (I_k \otimes Q') dW_t$. For instance, the 3×3 Wishart model with $k = 3$ degrees of freedom has a super-quadratic nine-factor interpretation.

With this equivalence, the parameter identification is ensured under conditions similar to those of Ahn, Dittmar, and Gallant (2002) or Leippold and Wu (2002). Since our completely affine market price of risk does not contain additional parameters, we can identify $I_k \otimes (M - Q')$, $I_k \otimes M$ and $I_k \otimes Q'$ from the Gaussian distribution of the state variables, both under physical and risk neutral probabilities. To pin down the parameters with respect to invertible linear transformations, it is enough to require matrices $I_k \otimes (M - Q')$ and $I_k \otimes M$ to be lower triangular. This condition is equivalent to imposing that matrices M and Q' are both lower triangular.

In addition to identification restrictions, the model implies two mild conditions on the parameter matrices: (i) negative definiteness of matrix M, which guarantees the stationarity of factors, and (ii) the invertibility of matrix Q, which ensures that Σ_t is reflected towards the domain of positive definite matrices when the boundary of the state space is reached. We can additionally require the $D - I_n$ matrix to be positive definite, thus ensuring positive yields. Appendix C.5 provides details on the estimation procedure along with the estimated parameter values.

3.2.B Properties of Risk Factors

The Wishart process gives freedom in modeling the conditional dependence between positive factors. The current section investigates this property using the parameters of the estimated 2×2 model.

Factor correlations. To demonstrate how the time variation in correlations comes up in our setting, we consider an example of the instantaneous covariance between the positive elements of a 2×2 matrix of factors:

$$Corr_t(\Sigma_{11}, \Sigma_{22}) = \frac{\langle \Sigma_{11}, \Sigma_{22} \rangle_t}{\sqrt{\langle \Sigma_{11} \rangle_t} \sqrt{\langle \Sigma_{22} \rangle_t}}. \tag{3.24}$$

Instantaneous variances and covariance of the elements Σ_{11} and Σ_{22} are straightforward to compute (see Result 1 in Appendix C.4):

$$\begin{aligned} d\langle \Sigma_{11} \rangle_t &= 4\Sigma_{11}(Q_{11}^2 + Q_{21}^2)dt, \\ d\langle \Sigma_{22} \rangle_t &= 4\Sigma_{22}(Q_{12}^2 + Q_{22}^2)dt, \\ d\langle \Sigma_{11}, \Sigma_{22} \rangle_t &= 4\Sigma_{12}(Q_{11}Q_{12} + Q_{21}Q_{22})dt, \end{aligned} \tag{3.25}$$

where Q_{ij} is the ij-th element of matrix Q.

The conditional second moments are linear in the elements of the factor matrix. The covariance between positive factors is determined by the out-of-diagonal element Σ_{12}, which is either positive or negative. As a result, the instantaneous correlation of Σ_{11} and Σ_{22} is time-varying, unrestricted in sign, and depends on the elements of Σ in a non-linear way.[17] The negative model-implied conditional correlation of positive factors is a peculiarity in the context of ATSMs, in which positive (volatility) factors can be at best unconditionally positively correlated. In fitting the observed yields, however, the possibility of negative correlations plays a crucial role. For instance, Dai and Singleton (2000) report that in a CIR setting with two independent factors, studied earlier in Duffie and Singleton (1997), the correlation between the state variables backed out from yields is approximately -0.5, instead of zero. Their estimation results for the completely affine $A_1(3)$ subfamily give further support to the importance of negative conditional correlations among (conditionally Gaussian) factors.[18]

[17] When the elements of the Wishart matrix admit an interpretation as a variance-covariance matrix of multiple assets, Buraschi, Porchia, and Trojani (2009) show that the correlation diffusion process of $\rho = \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$ is non-linear, with the instantaneous drift and the conditional variance being quadratic and cubic in ρ, respectively. The non-linearity of the correlation process arises despite the affine structure of the covariance process itself.

[18] See Dai and Singleton (2000), Table II and III.

The degrees of freedom parameter k. The properties of the conditional factor correlations in our model are controlled by the degrees of freedom parameter k. Recall that k integer fixes the number of OU processes used to construct the state dynamics in equation (3.3) (see also Remark 4 and Appendix C.2 for a related discussion). As such, it drives the non-singularity of Σ_t. By going beyond unitary degrees of freedom, but keeping them integer, we obtain several features. First, we introduce a time-variation in conditional correlations of positive factors, and thus break the link between the model and the n-factor QTSMs. Second, since the diagonal factors are non-central $\chi^2(k)$ distributed, k influences different moments of yields. To illustrate this, we plot the instantaneous correlations of diagonal factors (Figure 3.2) as well as the distribution of the 5-year yield implied by the estimated 2×2 model for different k's (Figure 3.3). In the special case where $k = 1$, the conditional correlations of positive factors are piecewise constant. Moreover, the restrictive $\chi^2(1)$ factor distribution translates into too high a skewness of yields, as compared to the historical distribution. A higher k tends to mitigate the misfit to the higher moments of yields by making the distribution closer to the Gaussian. To compare the model's performance with the affine class, Figure 3.4 superposes the 5-year US yield against the densities implied by the 2×2 model (panel a) and three benchmark ATSMs estimated by Duffee (2002) (panel b). The $A_1(3)$, $A_2(3)$ ATSMs face problems in matching the higher order moments of yields, and the purely Gaussian $A_0(3)$ implies a non-negligible probability of negative yields. The Wishart model ($k = 3, 7$) appears free of these shortcomings.[19] Intuitively, the multiple roles played by the degrees of freedom parameter explain the relative flexibility of the model.

[Insert Figure 3.2, 3.3 and 3.4 here]

Factor volatilities. From the dynamics of the Wishart process in equation (3.3), all three state variables feature stochastic volatility. This is also visible in the significant GARCH coefficient which we compute for the simulated sample of factors.[20] Ahn, Dittmar, and Gallant (2002) note that the goodness-of-fit of the standard ATSMs may be weakened precisely in settings, in which state variables have pronounced conditional volatility and are simultaneously strongly negatively correlated. The ease of introducing correlations and stochastic volatilities in the Wishart model is one of its useful characteristics (Section 3.3.C and 3.4.B).

[19] Feldhütter (2007) provides extensive evidence of the different abilities of essentially affine, extended affine, and semi-affine models to match the higher-order moments of yields. The general conclusion from his work about the poor performance of the essentially affine family is confirmed in our Figure 3.4.

[20] For the sake of brevity, we do not report the coefficients here, but just remark that for all state variables the GARCH(1,1) coefficient is above 0.85.

$A_m(N)$-type interpretation of factors. It is informative to look at the Wishart setting from the perspective of the $A_m(N)$ taxonomy developed by Dai and Singleton (2000). For instance, the 2×2 Wishart framework combines several features of the previous models: (*i*) the number of unrestricted versus positive factors of the essentially affine $A_2(3)$ specification; (*ii*) the number of stochastic volatility factors of the completely affine $A_3(3)$ specification; (*iii*) the unrestricted (positive and negative) correlations among factors of the $A_0(3)$ specification; and it does not have a counterpart within the $A_m(3)$ family in terms of stochastic correlations among factors. We find the out-of-diagonal element of the Wishart matrix, Σ_{12}, to be negative in more than 90 percent of the simulated sample.[21] This result conforms with the affine literature, which provides evidence for the superior performance of the $A_{m<N}(N)$ class, with some factors having unrestricted signs, over the multifactor CIR models.

3.2.C Factors in Yields

In order to verify the properties of the state dynamics in our model, we study what yields can tell us about the factors.

Factors and principal components. As a basic check whether the historical yields could have been generated by the assumed factor dynamics, we apply the standard principal component analysis to the data and to the yields simulated with the model. It is well-documented that three principal components explain over 99 percent of the total variation in yields (Litterman and Scheinkman, 1991; Piazzesi, 2003). Since we use a 2×2 specification of the state matrix, the model-generated yields are spanned by at most three factors. We find that the portions of yield variation explained by the first two principal components in the model largely coincide with the empirical evidence. Moreover, the traditional factor labels are evident in Figure 3.5 with weights on the first three principal components virtually overlapping with those computed from the data.

[Insert Figure 3.5 here]

Shifting number of factors. Although the three-factor structure of US yields seems robust across different data frequencies and types of interest rates, recent research points to a time variation in the number of common factors underlying the bond market. Pérignon and Villa (2006) reject the hypothesis that the covariance matrix of US yields is constant over time, and

[21] The simulated sample comprises 72000 monthly observations from the model. Whenever subsequently we refer to simulation results, we always use this sample length, unless otherwise stated.

document that both factor weights and the percentage of variance explained by each factor change concurrently with changes in monetary policy under various FED chairmen.

The behavior of instantaneous correlations between the Wishart factors in Figure 3.2 leads us to investigate whether the model can help explain this seemingly changing risk structure. We sort yields according to the level of instantaneous correlations between state variables, and for each group we compute the principal components. This exercise shows that the percentage of yield variation explained by the consecutive principal components changes considerably and in a systematic way across different subsamples. For instance, the loadings on the first and second principal components can vary from over 99 to 95 percent and from 5 to almost zero percent, respectively, depending on the level of instantaneous correlations. Such variability is consistent with the decompositions of the US yield levels in different subsamples. In Figure 3.6 we plot, as a function of instantaneous factor correlation, the portions of yield variation explained by each principal component.

[Insert Figure 3.6 here]

The closed-form expression for the conditional covariance of yields (3.16) allows us to perform a dynamic principal component analysis (Figure 3.7). This decomposition leads to a much higher variation in factors explaining yields than what could be expected from the previous decomposition based on crudely defined subsamples. This discrepancy indicates that an empirical finding of some changeability in the yield factor structure across subsamples may significantly understate the true conditional variability. It stands to reason that, depending on the state of the economy, the relative impact of different macroeconomic variables (e.g., inflation expectations, real activity) onto the yield curve can vary considerably over time. This intuition finds support in the historical behavior of US yields, which seemed to be dominated by the variation of inflation expectations in the 1970s, and by the variation of real rates in the 1990s. Motivated by this evidence, Kim (2007a) highlights the relevance of structural instabilities and changing conditional correlations of macro quantities for explaining the term structure variation in the last 40 years.

[Insert Figure 3.7 here]

3.3 Yield Curve Puzzles

We assess the model in terms of the goodness-of-fit criteria discussed in the introduction to Section 3.2. The analysis based on the 2×2 specification and one set of estimated parameters (given in

Appendix C.5.A) indicates that this simple framework is able to replicate several features of the spot and derivative bond markets.

3.3.A Excess Returns on Bonds

The implication of the essentially affine market price of risk proposed by Duffee (2002) is that excess bond returns have unrestricted sign. This feature is crucial for matching their empirical properties. The completely affine Wishart model shares a similar property. Figures 3.8 and 3.9 present excess bond returns obtained from the 2×2 model. We observe a switching sign of the model-implied risk premia, both instantaneous (Figure 3.8) and those computed from discrete realizations of the bond price process (Figure 3.9). Excess returns are highly volatile, with the conditional ratio $e_t^\tau / \sqrt{v_t^\tau}$ having a large probability mass between ± 1. Third, the above properties hold true across bonds with different maturities.

[Insert Figure 3.8 and 3.9 here]

The model-implied returns match the magnitudes and the distributional properties of the US bond return dynamics. Empirically, the expected excess returns on long bonds are on average higher and more volatile than on short bonds due to the duration effect. In line with this evidence, the model produces expected excess returns and volatilities that rise as a function of maturity. Moreover, the ratio of mean excess returns to their volatility is well below one (0.17 on average) across all maturities. These features play an important role in the model's ability to replicate the failure of the expectations hypothesis, which we discuss next.

3.3.B The Failure of the Expectations Hypothesis

The expectations hypothesis states that yields are a constant plus expected values of the current and average future short rates. Thus, bond returns are unpredictable. This can be tested in a linear projection of the change in yields onto the (weighted) slope of the yield curve, known as the Campbell and Shiller (1991) regression:

$$y_{t+m}^{n-m} - y_t^n = \beta_0 + \beta_1 \frac{m}{n-m} \left(y_t^n - y_t^m \right) + \varepsilon_t, \tag{3.26}$$

where y_t^n is the yield at time t of a bond maturing in n periods, and n, m are given in months. While the expectations hypothesis implies the β_1 coefficient of unity[22] for all maturities n, a number of empirical studies point to its rejection. Moreover, there is a clear pattern to the way the expectations hypothesis is violated: In the data β_1 is found to be negative and increasing in absolute value with maturity. This means that an increase in the slope of the term structure is associated with a decrease in the long term yields. Rephrased in terms of returns, the expected excess returns on bonds are high when the slope of the yield curve is steeper than usually.

To study whether expected returns in the 2×2 model vary in the "right" way with the term structure, we compute the model-implied theoretical coefficients of Campbell-Shiller regressions,[23] and benchmark them against their empirical counterparts (see panels *a* in Table 3.1 and in Figure 3.10). For comparison, we perform an analogous exercise for three preferred affine specifications of Duffee (2002) at his parameter estimates (see panels *b* in Table 3.1 and in Figure 3.10). In Duffee's convention, preferred models drop the parameters that contribute little to the models' QML values. This gives rise to two essentially affine models $A_0(3)$ and $A_1(3)$ and one completely affine model $A_2(3)$.

The results indicate that the model can accommodate the predictability of the yield changes by the term structure slope. While the two Campbell-Shiller coefficients used as moment conditions are matched almost perfectly (see the shading in Table 3.1, panel *a*), the model turns out to do a good job also in fitting other parameters not used in the estimation. Panel *a* of Figure 3.10 shows that all model-implied coefficients lie within the 80 percent confidence bounds computed from the historical sample.[24] The results for the ATSMs concur with the previous literature.

[22] To see this, note that m-month return on n-maturity bond is $r_{t,t+m}^n = \ln P_{t+m}^{n-m} / P_t^n = -(n-m) y_{t+m}^{n-m} + n y_t^n$. Then, the monthly excess return over the risk free return y_t^m is:

$$rx_{t,t+m}^n = \frac{1}{m} r_{t,t+m}^n - y_t^m = -\frac{m}{n-m} \left(y_{t+m}^{n-m} - y_t^n \right) + \left(y_t^n - y_t^m \right).$$

Reformulating and taking expectation yields:

$$E_t \left(y_{t+m}^{n-m} - y_t^n \right) = -\frac{m}{n-m} E_t \left(rx_{t+m}^n \right) + \frac{m}{n-m} \left(y_t^n - y_t^m \right).$$

Under the expectations hypothesis, the first term on the RHS is a constant, and the slope coefficient in a regression based on the above equation is unity.

[23] In computing the theoretical coefficients, we follow Dai and Singleton (2002) who claim that matching the population coefficients to the historical estimates is a much more demanding task than matching the coefficients implied by yields fitted to an ATSM.

[24] Note that the 80 percent bound—however lax for the data—is a more rigid gauge for the model's performance than a wider (e.g. 90 percent) bound. In fact, the coefficients obtained from $A_1(3)$ and $A_2(3)$ still fall outside the 95 percent bound.

Table 3.1: Regressions of the yield changes onto the slope of the term structure

The table presents the parameters of the Campbell-Shiller regression in equation (3.26). The maturities n are quoted in months. The value of m is taken to be six months, for all n. Panel a, the first row presents historical coefficients based on US yields in the period 1952:01–2005:06. The third row shows the model-implied theoretical coefficients along with population t-statistics below. The shading indicates the coefficients used as moment conditions in estimation. The fifth row shows small sample results obtained from the model by Monte Carlo. Panel b shows analogous results for the preferred affine specifications of Duffee (2002) at his parameter estimates. The historical coefficients for the period 1952:01–1994:12 concur with the sample used in estimation. All model-implied t-statistics are computed using Newey-West adjustment of the covariance matrix. Due to unobservability of yields with a half-year spacing of maturity, we follow Campbell and Shiller (1991) (their Table I, p. 502), and approximate y_{t+m}^{n-m} by y_{t+m}^{n}. This approximation is used consistently for all model-implied and historical data.

a. Wishart 2×2 factor model

Maturity (n months)	12	24	36	60	84	120
Data β_1 (1952–2005)	−0.174	−0.615	−0.852	−1.250	−1.660	−2.244
t-stat	−0.4	−1.1	−1.4	−1.8	−2.1	−2.4
Model β_1 (popul.)	−0.070	−0.614	−1.070	−1.713	−2.120	−2.244
t-stat	−6.4	−9.2	−10.8	−12.2	−12.5	−11.5
Model β_1 (648 obs.)*	0.008	−0.521	−0.960	−1.568	−1.922	−2.065
t-stat	0.03	−1.5	−2.2	−2.4	−2.3	−1.9

b. ATSMs

Maturity (n months)	12	24	36	60	84	120
Data β_1 (1952–1994)	−0.392	−0.696	−0.890	−1.291	−1.738	−2.451
t-stat	−0.8	−1.2	−1.4	−1.7	−2.0	−2.3
$A_0(3)$ (essentially)*	−0.037	−0.401	−0.597	−0.986	−1.462	−2.248
t-stat	−0.8	−7.0	−9.0	−12.3	−15.4	−18.8
$A_1(3)$ (essentially)	0.522	0.445	0.545	0.653	0.620	0.472
t-stat	7.1	4.6	4.8	4.6	3.5	2.0
$A_2(3)$ (completely)	1.354	1.416	1.454	1.369	1.226	1.007
t-stat	18.6	14.2	12.6	10.3	8.3	5.6

*) The coefficients and t-statistics are the median of 1000 estimates based on the simulated sample of 648 observations. The simulated path reflects the length of the sample used to estimate the WTSM. To conserve space, for ATSMs we only provide the population results.

The essentially affine Gaussian model $A_0(3)$ conforms with the empirical evidence, whereas

both $A_1(3)$—notwithstanding its essentially affine market price of risk—and $A_2(3)$ models have counterfactual predictability implications.

[Insert Figure 3.10 here]

We have also considered two additional regressions, which are independent from the estimation procedure, but reflect the same reasons for the failure of the expectations hypothesis as the Campbell-Shiller regressions. First, following Duffee (2002) we study projections of the monthly excess constant maturity bond returns on the lagged slope of the term structure defined as the difference between the 5-year and the 3-month yield. Even though this regression merely restates the information conveyed by Campbell-Shiller coefficients, it provides a robustness check of the previous results, because the yields which construct the slope are not used directly in estimating the model. As a second check, we replicate the regressions of Fama and Bliss (1987) projecting the excess one-year holding period bond return on the spot-forward spread. For brevity, we only state the main results without reporting the details. Consistent with the empirical evidence, the model-implied coefficients in both regressions increase with the maturity of the bond used as the dependent variable. A steep slope of the term structure forecasts high excess returns during the next month. Similarly, a positive spot-forward spread is a predictor of higher holding period returns.

Within the essentially affine $A_m(3)$ family, Gaussian models dominate other subfamilies in terms of prediction because they allow for correlated factors as well as the changing sign of the market price of risk. The above exercise suggests that similar features can be obtained within a completely affine setting under the Wishart factor structure.

3.3.C Second Moments of Yields

Two issues that occupy the term structure research agenda are (*i*) the time variation and persistence of the conditional second moments of yields, (*ii*) the humped term structure of unconditional yield volatilities.

Persistence of conditional volatility of yields. In the Wishart model, stochastic volatilities of factors are a consequence of the definition of the process. We now explore how they translate into the conditional second moments of yields. Is the degree of time variation and persistence in yield volatility commensurate with historical evidence? To answer this question, we follow Dai and Singleton (2003), and estimate a GARCH(1,1) model for the 5-year yield (see Table 3.2). The

choice of the 5-year yield is motivated by the fact that this maturity is not involved in estimation of the 2×2 model. Therefore, its conditional and unconditional properties can be traced back to the intrinsic structure of the model. In panel *a* of Table 3.2, we report the coefficients for our model and compare them with the historical estimates. To be able to infer the relative significance of the two sets of parameters, we compute the median GARCH estimate based on 1000 simulated samples with 54 years of monthly observations each. We take the same approach to assess the volatility implications of the preferred $A_1(3)$ and $A_2(3)$ models of Duffee (2002) (see Table 3.2, panel *b*). Due to its constant conditional volatility assumption, the Gaussian $A_0(3)$ model is not taken into consideration.

The results confirm that the degree of volatility persistence implied by the Wishart factor model aligns with the historical figures. For example, the median model-implied GARCH coefficient is 0.847 as compared with 0.820, which is found empirically. Furthermore, the model is able to reproduce the positive link between yield levels and their conditional volatilities observed in the long data sample. We find the correlation between the level of the 5-year yield and its GARCH(1,1) conditional volatility to be 0.72 in the simulated sample.

The volatility persistence in the benchmark affine models is typically too low. As in Dai and Singleton (2003), we document that the essentially affine $A_1(3)$ specification exhibits conditional volatility that is roughly in line with the historical evidence. And yet, as shown is Section 3.3.B, it also largely fails in explaining the conditional first moments of yields. Although the $A_2(3)$ specification allows for two CIR-type factors, the volatility persistence it implies is even lower than in the $A_1(3)$ case. Note, however, that the preferred $A_2(3)$ model of Duffee (2002) is equivalent to the completely affine formulation, because the parameters of the essentially affine market price of risk turn out to be insignificant in estimation. This outcome reinforces the observation made by Dai and Singleton (2003) that the essentially affine market price of risk is the key to modeling the persistence in the conditional second moments of yields in ATSMs. Finally, the small sample confidence intervals for the GARCH estimates convey information about the proximity of the different models and the true process driving the volatility of yields. Out of the three models considered, the GARCH coefficient in the Wishart factor setting is on average closest to the historical number and also the least dispersed one.

Humped term structure of unconditional volatilities. The term structure of unconditional volatilities of yields (and yield changes) is another recurring aspect in the yield curve debate. Its shape, which varies across different subsamples, has aroused increased interest with the appearance of a hump at around 2-year maturity during the Greenspan era (Piazzesi, 2001). Dai and Singleton

Table 3.2: GARCH(1,1) parameters for the model-implied and historical 5-year yield

The table presents the estimates of a GARCH(1,1) model: $\sigma_t^2 = \bar{\sigma} + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, where ε_t is the innovation from the AR(1) representation of the level of the 5-year yield. Panel *a* shows the ML estimates for the Wishart factor model, and compares them to the historical coefficients based on the sample period 1952:01–2005:06. Panel *b* displays estimates for the preferred affine models of Duffee (2002): the essentially affine $A_1(3)$ and the completely affine $A_2(3)$ and compares them to the historical coefficients. Accordingly, the simulation of the ATSMs uses the estimates from Duffee (2002) for the sample period 1952:01–1994:12. The population values are based on 72000 observations.

a. Wishart 2×2 factor model

	α	β	$\bar{\sigma}$
Data (1952–2005)	0.180	0.820	0.000
t-stat	7.6	39.0	3.3
Model (popul.)	0.116	0.870	0.000
Model (648 obs.)*	0.123	0.847 [0.71, 0.96]	0.000
t-stat	4.4	27.6	2.4

b. ATSMs

	α	β	$\bar{\sigma}$
Data (1952–1994)	0.243	0.757	0.003
t-stat	6.8	23.3	3.7
$A_1(3)$ (popul.)	0.257	0.670	0.000
$A_1(3)$ (516 obs.)*	0.153	0.707 [0.50, 0.86]	0.000
t-stat	3.0	8.5	2.7
$A_2(3)$ (popul.)	0.409	0.590	0.000
$A_2(3)$ (516 obs.)*	0.370	0.547 [0.33, 0.84]	0.000
t-stat	5.2	9.7	4.4

*) The coefficients and t-statistics are the median of 1000 estimates based on the simulated sample of 648 and 516 months, respectively. The simulated path reflects the length of the sample used to estimate the different models. The numbers in square brackets show the small sample 99 percent confidence intervals based on Monte Carlo.

(2000) conclude that the key to modeling the hump in ATSMs lies either in correlations between the state variables or in the respective factor loadings in the yield equation. Consistent with this interpretation, our model allows for non-monotonic behavior of yield volatilities. To uncover the mechanism that leads to this non-monotonicity, we decompose the unconditional variance of yields into contributions of factor variances and covariances scaled by the respective loadings

(not reported). The decomposition reveals that the hump in the volatility curve is predominantly driven by the (weighted) variance of the Σ_{12} factor. This result fits within the interpretation of Dai and Singleton (2000) in that Σ_{12} also determines the correlation between the positive factors. The forward yield volatilities and cap implied volatilities are discussed in greater detail next.

3.3.D Aspects in Derivative Pricing

The term structure of forward rate volatilities. Similar to the unconditional second moments of yields, the term structure of forward rate volatilities tends to be hump-shaped for shorter maturities, as reported in Amin and Morton (1994) and Moraleda and Vorst (1997), among others. In WTSM, the instantaneous forward rate is given in closed form in expression (3.20). Thus, we can obtain the whole term structure of instantaneous forward rate volatilities. Consistent with the empirical evidence, at the estimated parameters a pronounced hump becomes visible. The decomposition of the model-implied forward rate variance reveals that the non-monotonicity is due to two elements: the variances of Σ_{12} and Σ_{11}, scaled by the respective functions of the elements of matrix $A(\tau)$.[25] Thus, yields and forward rates share the same source of a hump in volatility curves. Figure 3.11 (panel *a*) plots the instantaneous volatility curves for several dates in the simulated WTSM. In that in reality the hump emerges for discretely spaced (in contrast to instantaneous) forward rates, we also compute the theoretical standard deviations of one-year forward rates, and plot them against maturities in Figure 3.11 (panel *b*). To put the results into perspective with ATSMs, the hump is absent from the affine specifications at the parameters estimated by Duffee (2002). In the $A_0(3)$ model, the forward volatility curve is monotonically decreasing. The mixed models $A_1(3)$ and $A_2(3)$, instead, imply its increase for longer maturities—an implication which is not valid empirically.

[Insert Figure 3.11 here]

Implied volatilities of interest rate caps. The empirical properties of the conditional second moments of yields can be inferred from the implied volatility quotes for the interest rate derivatives, such as caps. Also in this case, the evidence of a hump is ubiquitous (e.g., Leippold and Wu (2003); De Jong, Driessen, and Pessler (2004)). In a yield curve model, the hump of cap volatilities can be induced via forward rates. The "transmission" mechanism follows from the fact

[25] The decomposition refers to the variance of the forward rate $f(t;S,T)$ prevailing at time t for the expiry at time $S > t$, and maturity $T > S$, defined as $f(t;S,T) = \frac{1}{T-S}\ln\frac{P(t,S)}{P(t,T)} = \frac{1}{T-S}\{b(t,S) - b(t,T) + Tr\left[(A(t,S) - A(t,T))\Sigma_t\right]\}$.

that the volatility of a caplet is the integrated instantaneous volatility of the forward rate (see e.g., Brigo and Mercurio, 2006). Thus, those models able to display a hump in the instantaneous forward rate volatility should also perform well in the pricing of interest rate caps.

To verify this statement, we use the model-implied prices and the corresponding Black (1976) volatilities for caps with maturities from one to 15 years. A cap struck at rate $\overline{C}$ starting at T_0 and making equidistant payments at times T_i, $i = 1, \ldots, n$, based on the simply compounded floating rate $L(T_{i-1}, T_i)$, can be priced according to:

$$\begin{aligned} \text{Cap}_t &= \sum_{i=2}^{n} E_t^* \left[e^{-\int_t^{T_i} r_s ds} \delta \left(L(T_{i-1}, T_i) - \overline{C} \right)^+ \right] \\ &= (1 + \delta \overline{C}) \sum_{i=2}^{n} E_t^{T_{i-1}} \left[\left(\frac{1}{1 + \delta \overline{C}} - P(T_{i-1}, T_i) \right)^+ \right], \end{aligned} \tag{3.27}$$

where $T_i - T_{i-1} = \delta$. To get the last expression note that $\delta L(T_{i-1}, T_i) = \frac{1}{P(T_{i-1}, T_i)} - 1$. The last payment date T_n determines the maturity of the cap. $E_t^{T_{i-1}}$ specifies the conditional expectation under the forward measure induced by the zero bond maturing in $T_{i-1} - t$ periods. Thus, using Proposition 4, we can value caps as portfolios of put options on zero bonds. By market convention, we focus on at-the-money (ATM) contracts, for which the strike rate $\overline{C}$ of the T_n-maturity cap is set equal to the current T_n-year swap rate:

$$\overline{C} = \frac{1}{\delta} \frac{P(0, T_1) - P(0, T_n)}{\sum_{i=2}^{n} P(0, T_i)}. \tag{3.28}$$

The mapping from cap prices to Black volatilities assumes an identical volatility for each caplet constituting the contract. Figure 3.12 presents the results as a function of the contract's maturity for several dates in the simulation. Indeed, our 2×2 factor specification is able to adapt an empirically plausible behavior of implied volatilities, which are lower for short maturity caps (one-year), increase in the intermediate range and decline smoothly for longer maturities.

[Insert Figure 3.12 here]

The implied cap volatilities carry information about the conditional features of the model. The investigation of derivatives helps assess the risk-adjusted properties of the state dynamics. The appearance of a hump in volatilities of forward rates and caps reflects the properties of the model under both physical and risk-adjusted measures. The research into the pricing of caps and swaptions points to a link between correlations of different yields/forward rates and the humped

term structure of their volatilities (Collin-Dufresne and Goldstein, 2001; Han, 2007). Interrelations of factors are source of the conditional hump in the quadratic term structure models (Leippold and Wu, 2003). A similar link can be retrieved in the Wishart setting from the role played by the out-of-diagonal factors. In the 2×2 example, Σ_{12} determines the stochastic dependence among different elements of the state matrix, and contributes to the non-monotonicity of the volatility curves. The shape of volatility curves is related to the underlying stochastic factor correlations that are equally present under the physical and risk-adjusted measures. In this sense the model differs from the standard affine class, in which factor correlations, possibly restricted under the risk-adjusted measure, are introduced through the market price of risk.

3.4 Extensions

By construction, any three-factor model is restricted at least in three respects. First, with the low dimension of the state space, the correlation of factors plays multiple (possibly conflicting) roles, steering simultaneously the time-series dynamics of the yield curve (e.g. predictability) along with its cross-sectional characteristics (e.g. humped term structure of yield volatilities). We note, for instance, that the inclusion of Campbell-Shiller coefficients in the set of moment conditions tends to somewhat worsen the model's fit to the unconditional volatility curve. Second, the number of risk factors restricts the number of non-collinear variables (e.g. forward rates or yields) that can be used in a forecasting regression of excess returns. Finally, with three factors necessary to explain the spot interest rates, there is no room for the so-called unspanned factors manifested in the prices of interest rate derivatives. These tensions can be alleviated by enlarging the state space to a higher dimension. In this section, we estimate the 3×3 model along the lines of Section 3.2.A (see Appendix C.5.B for parameters). This extension takes us to a six-factor setting with three positive and three unrestricted factors and 18 parameters. We show that the completely affine six-factor WTSM is able to accommodate further characteristics of fixed income markets in addition to those already captured in the 2×2 case.

3.4.A Forward-rate factor

Cochrane and Piazzesi (2005) strengthen the case against the expectations hypothesis by showing that excess returns across bonds of different maturities can be predicted with a single factor—a linear combination of five forward rates. Notably, the coefficients of a projection of one-year holding period bond returns on a constant and five one-year forward rates exhibit a systematic (tent-like)

pattern. The implications of different models for the Cochrane-Piazzesi-type predictability have been studied by Bansal, Tauchen, and Zhou (2003) and Dai, Singleton, and Yang (2004). These works have focussed on the case with at most three factors. However, to replicate the evidence in full (in an affine framework), at least five factors are required to avoid collinearity. We work with six. In our model, log bond prices are linear in the Wishart factors; hence the predictability due to a single linear combination of forward rates is equivalent to the predictability due to a single linear combination of the elements of Σ. We scrutinize the 3×3 model for the presence and the form of the single return-forecasting factor.

While the single common factor appears to be an established feature of the data, its specific shape could be an artifact of a smoothing method used in constructing the zero-curve. Dai, Singleton, and Yang (2004) document that the pattern can turn into wave-like if, instead of the *unsmoothed* Fama-Bliss (UFB) yields used by Cochrane and Piazzesi (2005), the *smoothed* Fama-Bliss (SFB) data set is employed.[26] Some caveats are in order: First, yields generated from the model naturally lead to a "smooth" zero curve. Second, in our model-generated term structure, as in reality, four factors are sufficient to explain the total variation of yields. Thus, in absence of any cross-sectional measurement error, the collinearity of regressors becomes a concern; the oscillating shape of the forecasting factor found by Dai, Singleton, and Yang (2004) in the smoothed data is a likely signal thereof.[27] Figure 3.13 plots the slope coefficients in regressions of individual one-year excess bond returns on the set of five one-year forward rates. The pattern of factor loadings in the model (Figure 3.13, panel *a*) resembles closely the one found in the SFB data (Figure 3.13, panel *b*).

[Insert Figure 3.13 here]

[26] Both data sets are constructed from the same underlying coupon bond prices. The method used to extract the UFB yields assumes the forward rate to be a piecewise linear (step) function of maturity, whereas the SFB data is computed by smoothing the UFB rates with a Nelson-Siegel exponential spline. See Dai, Singleton, and Yang (2004) and Singleton (2006) for a discussion.

[27] Bansal, Tauchen, and Zhou (2003) suggest that in the presence of three latent factors driving the historical yields, the use of five regressors creates near perfect co-linearity problem, up to cross-sectional measurement errors that mask the singularity. As a standard remedy to collinear regressors, we add a very small amount of i.i.d. noise to model-generated variables. This has virtually no impact on the first four moments of the interest rates distribution. The noise is generated from $N(0, 4.5 \times 10^{-6})$; a different distribution, e.g. t_3-student, does not change the results. Especially, the level of the R^2 statistics remains largely unchanged. This finding is in line with the argument of Cochrane and Piazzesi (2005) that the predictability is not driven by measurement errors. We note that running the regression without noise gives qualitatively identical results, but leads to unreasonably high coefficients, which is again a diagnostic of the collinearity problem.

Judging by the regularity of the slope coefficients, the key intuition for the single factor seems to be supported by the model. Moreover, by including only three forward rates ($f_t^{0\to1}, f_t^{2\to3}, f_t^{4\to5}$) in the regression, the upward pointed triangular shape of the loadings (a "restricted tent") becomes apparent (not reported). The important question, however, is whether the 3×3 model can recover the high degree of predictability due to the *single* forecasting factor, rather than whether it can recover the particular shape. In Table 3.3, we report the univariate second stage regressions of excess returns on the linear combination of forward rates, in Cochrane-Piazzesi notation:

$$\frac{1}{4}\sum_{\tau=2}^{5} rx_{t+1}^{(\tau)} = \gamma'\mathbf{f_t} + \bar{\varepsilon}_{t+1} \qquad \text{(1st stage)} \tag{3.29}$$

$$rx_{t+1}^{(\tau)} = b^{(\tau)}\left(\gamma'\mathbf{f_t}\right) + \varepsilon_{t+1}^{(\tau)} \qquad \text{(2nd stage)} \tag{3.30}$$

where $rx_{t+1}^{(\tau)} = hpr_{t+1}^{\tau\to\tau-1} - y_t^{1Y}$ is the return on holding the τ-maturity bond in excess of one-year yield, and $\mathbf{f_t} = \left[1, f_t^{0\to1}\text{ (spot)}, \ldots, f_t^{4\to5}\right]$ is the vector of forward rates. With γ coefficients in equation (3.30) fixed at the values from the first stage regression, $\gamma'\mathbf{f_t}$ represents the return-forecasting factor.

Note that the moments used to estimate the model do not provide direct information about the above regressions. Nevertheless, the model can replicate the empirical evidence. We study its population as well as small sample implications and juxtapose them with the UFB and SFB data sets. The main conclusion of Cochrane and Piazzesi (2005)—that a single factor accounts for a large portion of time-variation in excess returns—is confirmed by the high R^2 values in Table 3.3. The model tends to generate R^2's that are very much in line with the empirical figures in panel *a*. In support of the single factor hypothesis, the $b^{(\tau)}$ coefficients are all significant and increase smoothly with bond's maturity τ. Such behavior persists across all data sets, irrespective of the underlying pattern of the γ's in the first stage regressions.

An important empirical property of the return-forecasting factor is that it carries information beyond what is captured in the level, slope (typically considered to be the return predicting variable) and curvature. The evidence in Cochrane and Piazzesi (2005) suggests that $\gamma'\mathbf{f_t}$ is related to the fourth principal component of yields, which in turn has only a weak impact on the yields themselves. In our setting, this outcome has two different shades. Despite the great stability of the model-implied finite sample estimates in Table 3.3, the relationship between $\gamma'\mathbf{f_t}$ and the yield factors turns out highly susceptible to the small sample biases. The Monte Carlo analysis (based on 360 observations from the model) indicates that the portions of $\gamma'\mathbf{f_t}$ variance explained

Table 3.3: Single forward rate factor regressions

The table reports the coefficients $b^{(\tau)}$, R^2, and t-statistics for the restricted Cochrane-Piazzesi regressions in equation (3.30). τ in the first row refers to the maturity of the bond whose excess return is forecasted. Panel *a* displays the results for two data sets: smoothed Fama-Bliss (SFB) and unsmoothed Fama-Bliss (UFB) yields. The yields are monthly and span the period 1970:01–2000:12. Panel *b* presents the model-implied estimates in the population (72000 observations), and in a small sample of length 360 months. All reported t-statistics use the Newey-West adjustment of the covariance matrix with 15 lags.

τ	2	3	4	5	2	3	4	5
	a. Data (1970-2000)							
	SFB data				UFB data			
$b^{(\tau)}$	0.46	0.85	1.19	1.50	0.46	0.87	1.22	1.45
t-stat	5.63	5.44	5.18	4.95	7.99	7.60	7.54	7.05
R^2	0.30	0.31	0.31	0.32	0.36	0.37	0.39	0.36
	b. Wishart 3×3 factor model							
	Population				Small sample* (360 obs.)			
$b^{(\tau)}$	0.46	0.86	1.20	1.48	0.46	0.86	1.20	1.48
t-stat	59.48	60.99	58.63	55.42	6.46	6.52	6.32	6.05
R^2	0.31	0.33	0.31	0.29	0.41	0.43	0.41	0.40

*) The coefficients and t-statistics are the median of 1000 estimates, each based on the sample of 360 realizations from the model.

by the respective principal components are highly uncertain quantities: They can range from 4 to 43 percent for the slope factor, and from 2 to 28 percent for the fourth factor (as measured by the upper and lower decile). This changes as the length of the sample becomes large. Then, the slope accounts for as little as 12 percent of the $\gamma' \mathbf{f_t}$ variance, while the total contribution of the fourth and fifth factor approaches 40 percent.

3.4.B Conditional Hedge Ratio

The evidence in the literature suggests that low dimensional models have difficulties in capturing the right dynamics of volatilities and correlations of different segments of the yield curve (Bansal, Tauchen, and Zhou, 2003; Dai and Singleton, 2003). It comes as no surprise that the 2×2 model cannot fully reflect the cross-sectional dynamics of yields. Therefore, we investigate a six-factor setting.

Table 3.4: Properties of the conditional hedge ratio

The table reports the statistics for the conditional hedge ratio HR_t, the ratio of conditional volatilities of yield changes $\sigma_{10}(t)/\sigma_2(t)$, and the conditional correlation of yield changes $\rho_{2,10}(t)$. The conditional volatilities are estimated with a GARCH(1,1) model, the conditional correlations—with the DCC model of Engle (2002). We report the means and volatilities of the estimated conditional quantities. For the model, we provide the population statistics along with their small sample 99 percent confidence intervals (in brackets underneath) based on Monte Carlo with 1000 repetitions of 54 years of monthly data each.

a. Data (1952–2005)

	HR_t	$\sigma_{10}(t)/\sigma_2(t)$	$\rho_{2,10}(t)$
Mean	0.61	0.76	0.80
Volatility	0.16	0.17	0.06
$\text{Corr}(\sigma_2(t), \sigma_{10}(t))$	0.86		

b. Wishart 3×3 factor model

	HR_t	$\sigma_{10}(t)/\sigma_2(t)$	$\rho_{2,10}(t)$
Mean	0.51	0.61	0.83
conf. bound	[0.41, 0.64]	[0.50, 0.74]	[0.80, 0.87]
Volatility	0.13	0.14	0.03
conf. bound	[0.05, 0.15]	[0.06, 0.17]	[0.01, 0.06]
$\text{Corr}(\sigma_2(t), \sigma_{10}(t))$	0.76		
conf. bound	[0.40, 0.91]		

We focus on the conditional hedge ratio between the 10- and 2-year bonds, i.e. $HR(t) = \frac{\sigma_{10}(t)}{\sigma_2(t)}\rho_{2,10}(t)$. $\sigma_2(t)$ and $\sigma_{10}(t)$ are the conditional volatilities of yield changes obtained with a univariate GARCH(1,1); $\rho_{2,10}(t)$ is the conditional correlation of the yield changes estimated with the dynamic conditional correlation (DCC) model of Engle (2002). The model replicates the general properties of HR_t relatively well (see Table 3.4). The model-implied volatility of both the conditional correlation process $\rho_{2,10}(t)$ and of the volatility ratio $\sigma_{10}(t)/\sigma_2(t)$ is close to the empirical one. Overall, the implications for the conditional hedge ratio seem realistic as measured by the small sample confidence intervals. We reach a similar conclusion for the correlation of volatilities of the two bonds.

The conditional behavior of second moments of yields has attracted considerable attention in the latest term structure literature (e.g., Collin-Dufresne, Goldstein, and Jones, 2006; Joslin, 2007). This research tends to agree in that low dimensional affine models face difficulties in

capturing the conditional yield volatility across maturities. Not surprisingly, those models in which more factors have stochastic volatility seem to perform better. Based on the examination of $A_m(3)$ models, Jacobs and Karoui (2006), for instance, suggest the $A_3(4)$ or $A_4(4)$ class as the best potential candidate for modeling volatility. At the same time, they recognize that its heavy parametrization may frustrate the estimation effort.

3.4.C Unspanned Factors

With six state variables at hand, but only three factors needed to explain the variation of yields, the 3×3 model lends itself to exploring the presence of factors unspanned by the spot market.

The discovery of unspanned factors follows from the poor performance of bond portfolios in hedging interest rate derivatives. Recent research reports a considerable variation in cap and swaption prices that appears to be weakly related to the underlying bonds. For instance, Heidari and Wu (2003) document that three common factors in Libor and swap rates can explain little over 50 percent in swaption implied volatility. Li and Zhao (2006) arrive at a similar conclusion for at-the-money (ATM) difference caps. Even though the yield factors can explain around 90 percent of the variation in the short maturity cap returns, their explanatory power deteriorates dramatically at longer maturities approaching just 30 percent for the 10-year maturity.

Two observations lead us to expect a similar phenomenon to arise in our estimated model. First, while the Σ_t matrix includes six factors,[28] the principal component decomposition of the model-generated yields reveals that three factors already explain nearly their total variation. Second, the estimated loading matrix $A(\tau)$ in the yield equation (3.15) is almost reduced rank across all maturities (see Table 3.5, panel *a*), which indicates the possibility that the dimension of the state space generated by the yields is strictly smaller than the dimension of the state space generated by the Wishart factors.

In the absence of a theory for unspanning in the Wishart setting, we take an empirical backing out approach to test the presence of unspanned volatility features in the model. We obtain the model prices of ATM caps for maturities from one to 10 years, and perform three different checks. First, we regress caps against three factors (level, slope and curvature) that span the spot yield curve in the model. Interestingly, even without imposing any structural restrictions, we do find evidence of bond market incompleteness, with an average R^2 of just 53 percent (see Table 3.5,

[28] This can be seen from Σ_t being full rank. Equivalently, a principal component exercise performed on the unconditional covariance matrix of the state variables (i.e. covariance between the elements of Σ_t) also detects six factors.

Table 3.5: Factors in the derivatives market

The table reports the level of spanning of cap prices by the spot yields in the 3×3 model. Panel *a* reports the eigenvalues of the $A(\tau)$ coefficients in the yield equation (3.15). We consider cap contracts with maturities of one to 10 years. The first line in panel *b* contains the R^2 values in regressions of caps on the first three principal components of yields (level, slope and curvature). The second line presents the R^2 values in regressions of cap prices on the (maximal set of) six principal components. Panel *c* provides the principal component decomposition of the covariance matrix of residuals from the first regressions of different maturity caps on the three yield PC's. Panel *d* reports the R^2 values in regressions of spot yields on six principal components retrieved from the covariance matrix of cap prices.

a. Eigenvalues of $A(\tau)$ coefficients

τ	1	2	3	4	5	6	7	8	9	10
λ_1	−0.12	−0.23	−0.36	−0.50	−0.65	−0.81	−0.98	−1.15	−1.32	−1.48
λ_2	−0.04	−0.06	−0.07	−0.08	−0.08	−0.08	−0.09	−0.09	−0.09	−0.09
λ_3	2.7e-6	−2.3e-5	−3.6e-5	−4.4e-5	−4.8e-5	−5.2e-5	−5.5e-5	−5.7e-5	−5.8e-5	−6.0e-5

b. Regressions of ATM caps on yield PC's

Cap maturity	1	2	3	4	5	6	7	8	9	10	Mean
3 PC's*, R^2	55.5	73.0	74.6	70.8	64.8	57.4	48.7	39.0	29.1	19.6	53.2
6 PC's, R^2	91.5	88.9	84.2	79.0	73.2	66.6	59.0	50.5	41.6	32.9	66.7

* Note that regressions of each of the yields on the first three PC's give the R^2 of 1.

c. Decomposition of the covariance matrix of residuals

Eigenvector	1	2	3	4	5	6	7	8	9	10
% Explained	97.94	1.81	0.22	0.03	0.00	0.00	0.00	0.00	0.00	0.00

d. Regressions of yields on cap PC's

Yield maturity	1	2	3	4	5	6	7	8	9	10	Mean
6 cap PC's, R^2	93.4	93.8	94.0	94.1	94.2	94.2	94.3	94.3	94.4	94.4	94.1

panel *b*). By using just three factors we remove some information, which—while irrelevant for spot yields—could possibly contribute to the variation of caps. To check this possibility, we subsequently include the remaining principal components in the regressions. Although the R^2 increases, the general conclusion persists, and is roughly consistent with the findings of Li and Zhao

(2006) for ATM difference caps or Heidari and Wu (2003) for swaptions. Second, we decompose the covariance matrix of residuals from the first set of regressions (panel *c*). The decomposition exposes at least three additional factors influencing cap prices. These factors reflect the residual degrees of freedom in the model which have not been exploited in its estimation with the spot yield data. Finally, to asses whether the results are just a spurious effect of applying a linear regression to a nonlinear problem, we reverse the first exercise, and regress yields on factors obtained from caps. The ability of caps to hedge the interest rate risk is evident in the high R^2 values (on average 94 percent) reported in panel *d*.

In ATSMs, the theoretical conditions for the existence of unspanned factors restrict the coefficient loadings in the yield equation to be linearly dependent for all maturities, as first noted by Collin-Dufresne and Goldstein (2002). Under such constraints, not all state variables are revealed through the yield curve dynamics alone. This pioneering analysis has been recently expanded by Joslin (2006) who provides general conditions for the incomplete bond markets in affine $\mathbb{R}_+^m \times \mathbb{R}^{n-m}$ models. His results, however, do not carry over to the Wishart factor framework due to the different form of the state space. In a companion paper, Joslin (2007) shows that a four-factor ATSM with an unspanned volatility restriction is soundly rejected by the data. An important—and so far unexplored—issue is whether such a conclusion would also persist in a model of larger dimension.[29] Our findings seem to indicate that the completely affine enlarged WTSM provides a way to reconcile unspanned factors with the remaining stylized facts of the yield curve. The characterization of the theoretical conditions for bond market incompleteness in the Wishart yield curve setting is an interesting topic for future research.

3.5 Conclusions

In this article, we study the implications of a term structure model with stochastically correlated risk factors driven by a matrix-valued Wishart process. Under this class, we resurrect the completely affine market price of risk specification and document that the setting provides an explanation for several term structure puzzles.

The model is endowed with three characteristics: (*i*) the market price of risk can take both positive and negative values, (*ii*) the correlation structure of factors is stochastic and unrestricted

[29] Heidari and Wu (2003), for instance, point to a necessity of a Gaussian 3+3 factor model to explain the joint behavior of yields and interest rate derivatives.

in sign, and (*iii*) all factors display stochastic volatility implying multivariate dynamics in second moments of yields. With these elements we investigate the following issues:

First, we are able to replicate the distributional properties of yields and the dynamic behavior of expected bond returns. The predictability of returns in the Wishart framework violates the expectations hypothesis in line with historical evidence. The model-implied population coefficients in the Campbell and Shiller (1991) regressions are negative and increase in absolute value with time to maturity. Similarly, a steeper slope of the term structure and a larger spot-forward spread forecast higher excess bond returns in the future.

Second, the model-implied conditional yield volatilities match the data in terms of GARCH estimates. We document the superior performance of the model in replicating the empirically relevant degree of volatility persistence as compared to the preferred affine specifications estimated by Duffee (2002).

Third, we find that the term structure of the forward rate volatilities in the model is marked by a hump around the two-year maturity. The result is preserved both instantaneously, and for the unconditional volatilities of discretely spaced one-year forward rates generated by the model. The conditional hump is further confirmed in Black implied volatilities of caps. Implicit in the volatility curves are the correlations between the state variables, which unlike in the standard affine class are equally present under physical and risk-neutral measures.

Several additional facts are worth highlighting. First, to illustrate its basic properties, we use the most parsimonious formulation of the model. The choice of a 2×2 state matrix puts us in a three-factor framework, with two positive and one unrestricted factor, and the simple completely affine market price of risk specification. In this form, the model has only 9 parameters to perform the tasks listed above. Second, using a single set of parameters, the setting reconciles several properties of model-implied yields with their historical counterparts. The factor structure of the model permits to reproduce both the unconditional and conditional features of the data, such as the persistent conditional volatilities and humped term structure of cap implied volatilities.

It is also useful to recognize the analogies of the Wishart setting with the standard ATSMs. The argument we apply to derive the market price of risk is the one that stands behind the completely affine class. Likewise, for an arbitrary dimension of the state matrix, we benefit from the analytical tractability comparable to a multifactor CIR model. In spite of these similarities, the theoretical properties of the state space set the Wishart approach apart from ATSMs.

The presented framework allows for an easy extension beyond three factors. With the 3×3 dimension of the state space, the model has six factors, three of which are restricted in sign. While

this gain in flexibility is attractive, in terms of the number of parameters (18) the model remains tractable. The enlarged model performs well across the dimensions mentioned above, and also has the scope to tackle more complex dynamics of the fixed income data. As a consequence, we are able to address several issues exposed by the recent yield curve literature: In the enlarged framework, the predictability of excess bond returns is supported by the single forecasting factor of Cochrane and Piazzesi (2005). The model produces realistic behavior of conditional hedge ratios between bonds. Finally, some state variables that load weakly on yields have an economically significant impact on the prices of interest rate caps in line with the notion of unspanned factors.

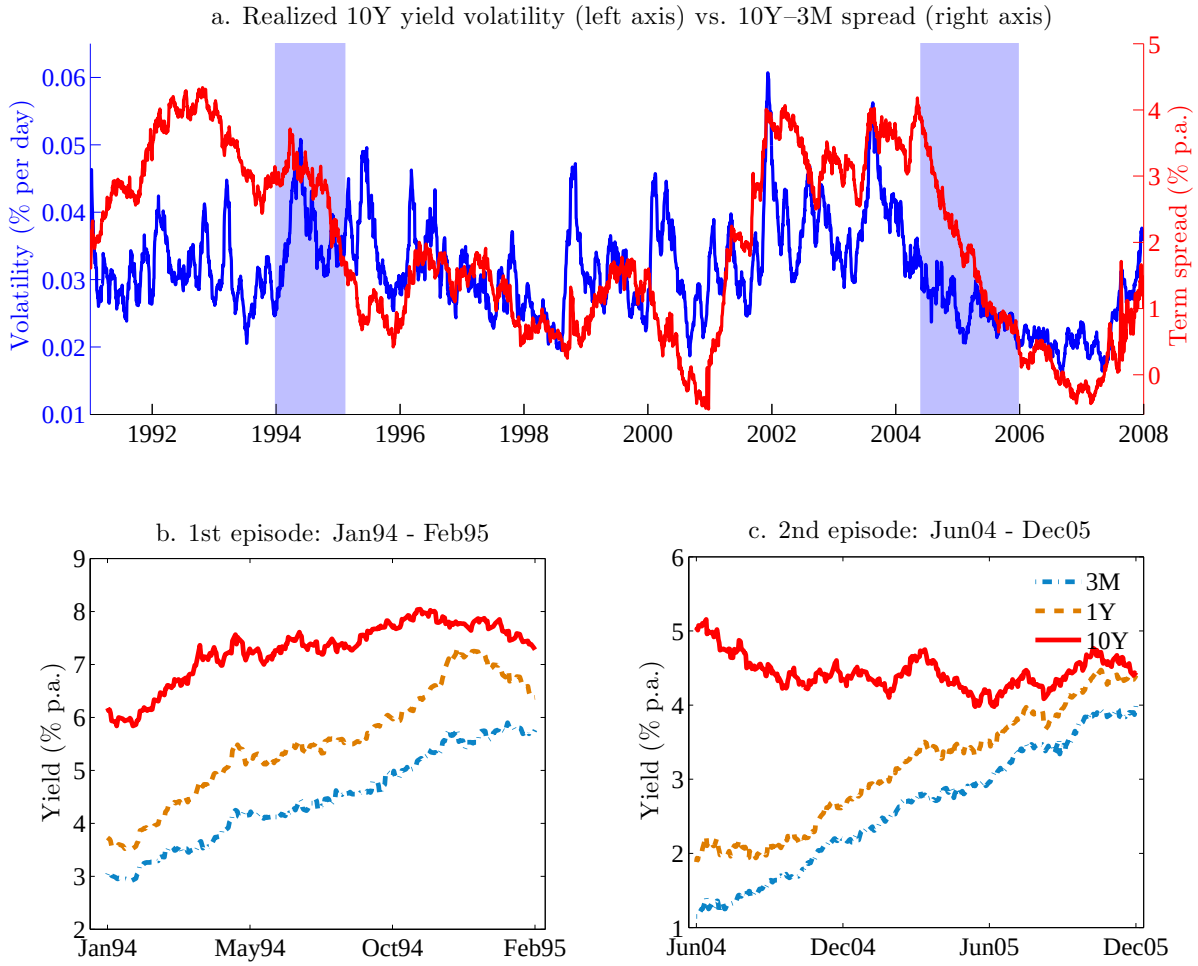


Figure 3.1: Term spreads and yield volatilities during 1994/95 and 2004/05 tightenings

Panel *a* plots interest rate volatilities (left axis) and yield spread (right axis) covering the period from 1991:01 to 2007:12. The shadings mark the 1994/95 and 2004/05 tightenings. The yield volatility is the 22-day moving average of realized daily volatilities ($v_\tau(t+h), h = 1$ day) obtained from high-frequency returns on 10-year Treasury note futures, where $v_\tau^2(t+h) = \frac{1}{\tau^2}\sum_{i=1}^{n} r^2(t + \frac{ih}{n}), n = 40$, and $r(t + \frac{ih}{n})$ is the 10-minute log return×100 on the futures contract. The spread is computed as the difference between the 10-year yield minus the 3-month T-bill rate. Panels *b* and *c* give a zoomed view on the dynamics of the 3-month, 1-year and 10-year yields during 1994/95 and 2004/05 episodes, respectively. Data sources: Tick-by-tick interest futures prices are from TickData.com, zero yields—from Gürkaynak, Sack, and Wright (2006) files, and 3-month T-bill rate—from Fed's H.15 files.

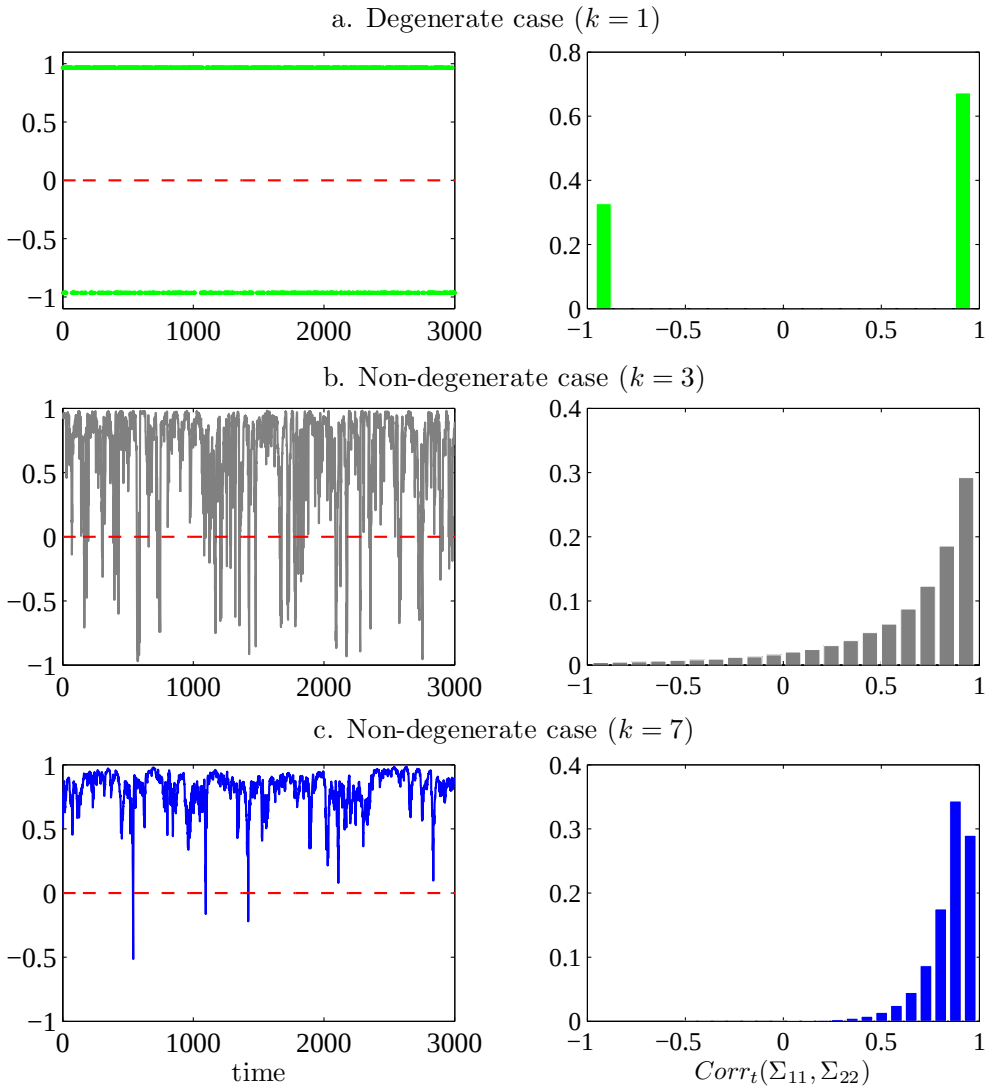


Figure 3.2: Instantaneous correlations of positive factors for different k's in the 2×2 WTSM

Panels a, b, and c plot the instantaneous correlations between the diagonal factors Σ_{11} and Σ_{22} in the 2×2 WTSM with $k = 1, 3$ and 7 degrees of freedom, respectively. The panels on the left show the consecutive realizations of the instantaneous correlations calculated according to equation (3.24), while the panels on the right display their respective histograms. In the degenerate case of $k = 1$, the WTSM narrows down to a QTSM.

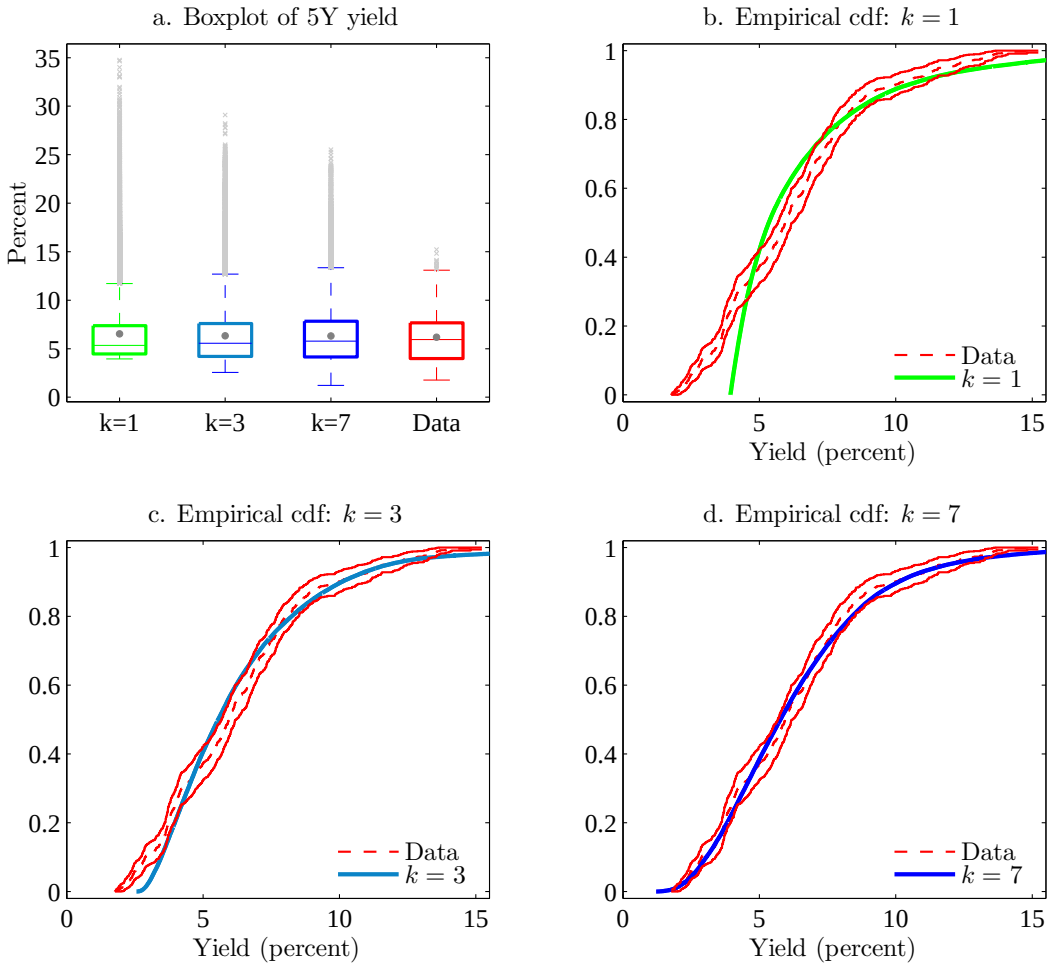


Figure 3.3: Unconditional distribution of the 5-year yield for different k in the 2×2 WTSM

Panel a presents boxplots of the 5-year yield in the 2×2 model with $k = 1, 3$ and 7 degrees of freedom, respectively. The results are juxtaposed with the US 5-year yield (1952:01–2005:06). Panels b, c and d illustrate the related empirical cumulative distribution functions both for the model (thick line) and for the data (dashed thin line). The solid thin lines mark the 99% upper and lower confidence bounds computed from the data with the standard Greenwood's formula.

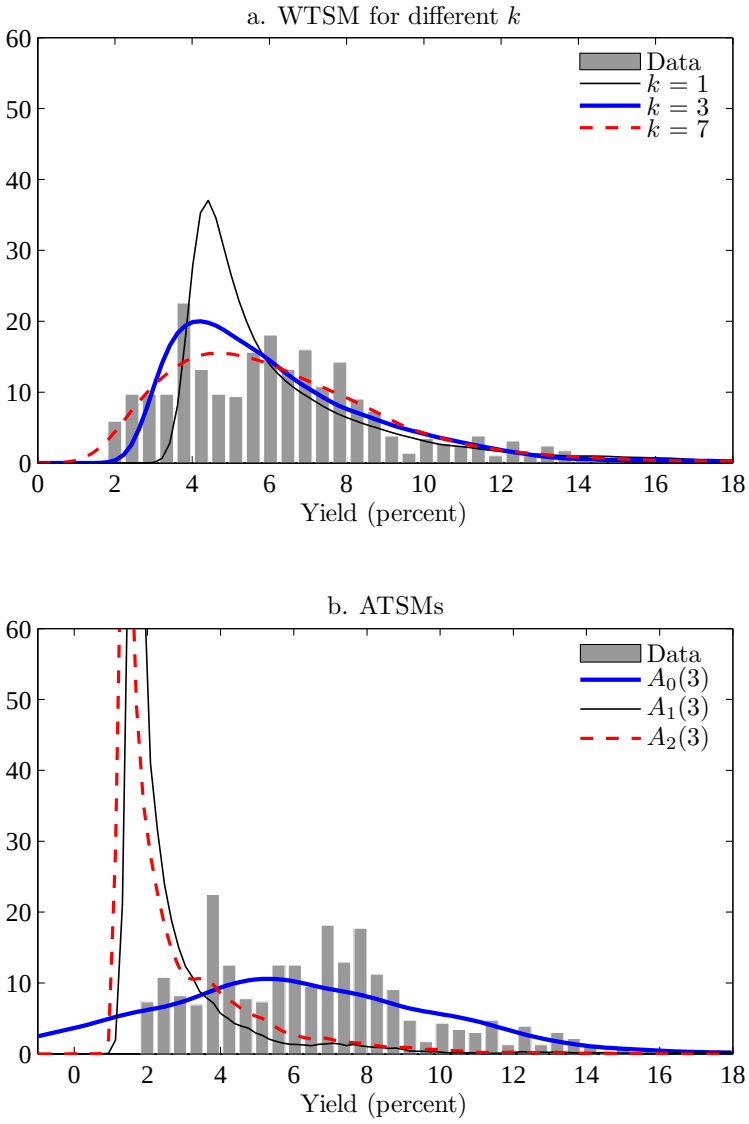


Figure 3.4: Actual and model-implied unconditional distributions of the 5-year yield

Panel *a* presents the densities of the 5-year yield obtained from the 2×2 WTSM with different degrees of freedom k. The results are superposed with the histogram of the 5-year US yield (1952:01–2005:06). Panel *b* shows the preferred ATSMs estimated by Duffee (2002): Gaussian $A_0(3)$, mixed models $A_1(3)$ and $A_2(3)$, and compares them with the histogram of the 5-year US yields (Duffee's sample 1952:01–1994:12).

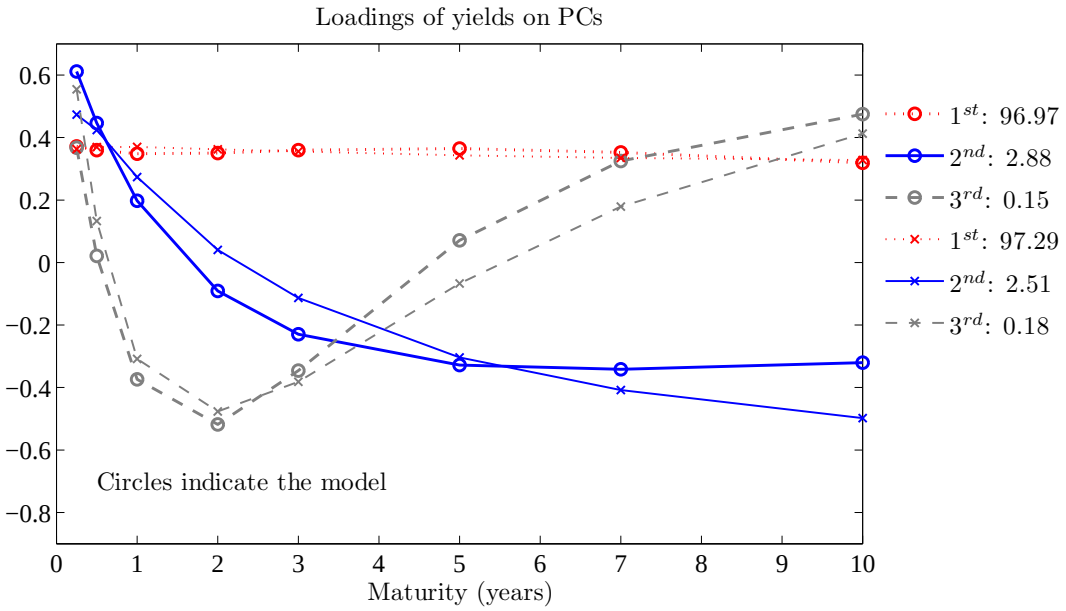


Figure 3.5: Loadings of yields on principal components: 2×2 WTSM vs. data

The covariance matrix of yields is decomposed as $U\Lambda U'$, where U is the matrix of eigenvectors normalized to have unit lengths, and Λ is the diagonal matrix of associated eigenvalues. The figure shows columns (factor loadings) of U associated with the three largest eigenvalues. Thicker circled lines indicate loadings of yields obtained from the 2×2 WTSM ($k = 3$). Finer lines are loadings obtained from the sample of US yields, 1952:01–2005:06. The legend gives the corresponding percentages of yield variance explained by the first three principal components.

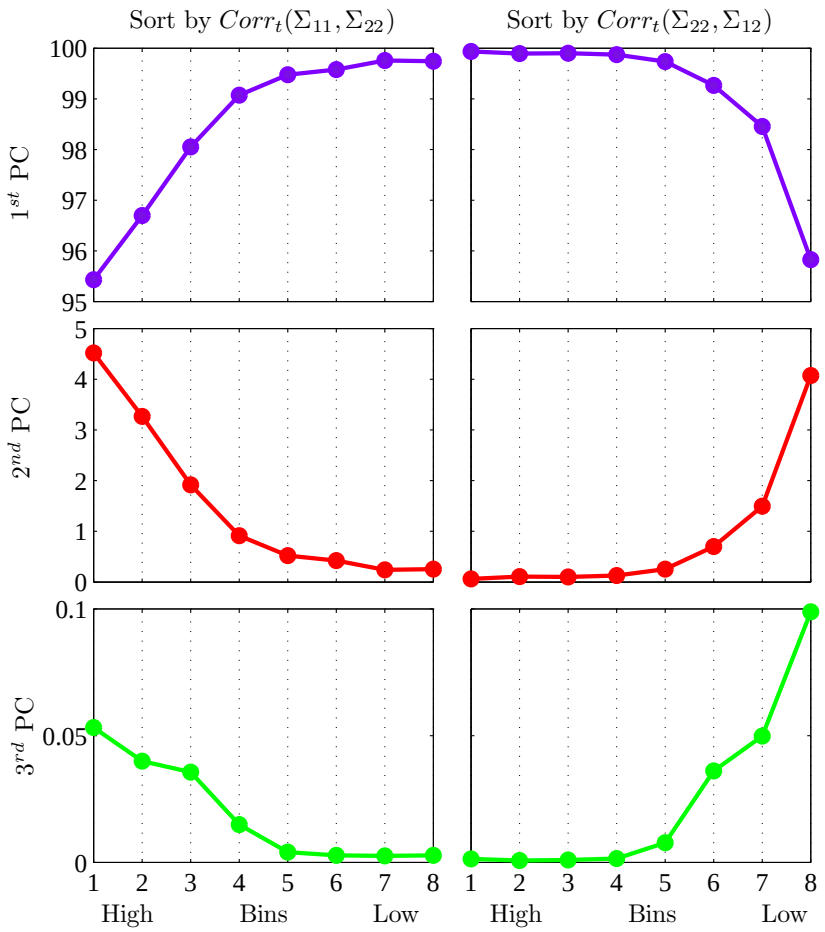


Figure 3.6: Variance explained by principal components, conditional on factor correlation

The figure shows the portions of yield variance explained by each principal component. The principal components of yields are computed conditional on the level of instantaneous correlation of factors in the 2×2 WTSM ($k = 3$). We form eight correlation bins, and number them from 1 (highest) to 8 (lowest). The bins are in descending order: $(1, .9)$, $(.9, .8)$, $(.8, .5)$, $(.5, 0)$, $(0, -.5)$, $(-.5, -.8)$, $(-.8, -.9)$, $(-.9, -1)$. Two sort criteria are used: the left-hand panel sorts by the level of $Corr_t(\Sigma_{11}, \Sigma_{22})$, the right-hand panel—by the level of $Corr_t(\Sigma_{22}, \Sigma_{12})$.

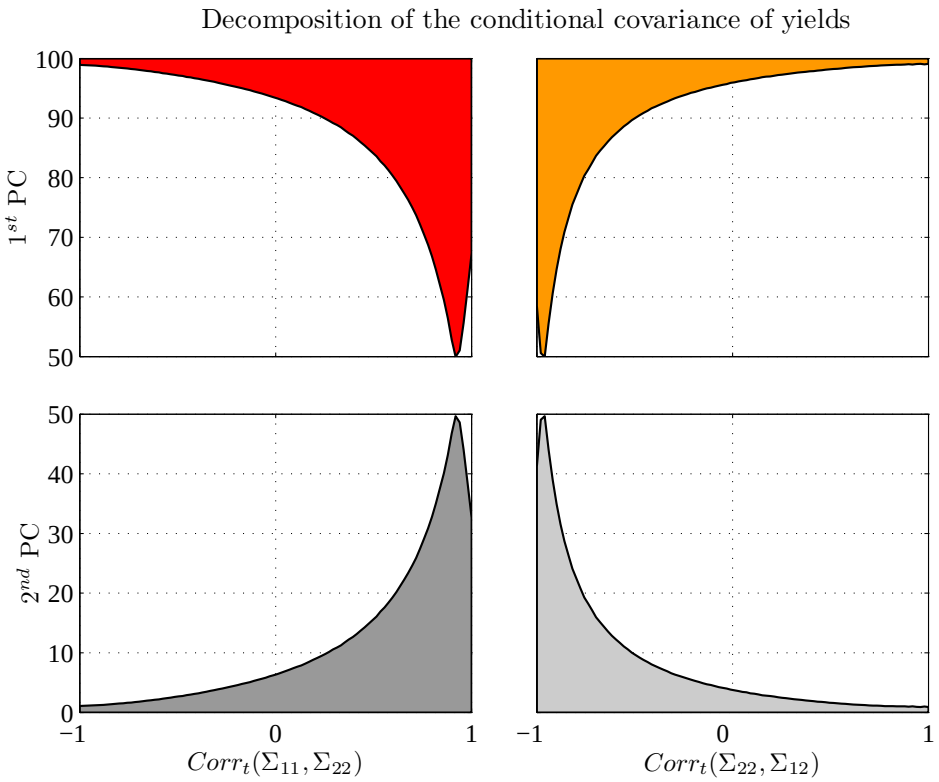


Figure 3.7: Conditional principal components of yields in the 2×2 WTSM

The figure displays the principal component decomposition of the instantaneous covariance of yields conditional on the level of the instantaneous correlation of factors in the 2×2 WTSM ($k = 3$). The shaded areas show the percentage of the conditional variance explained by the first principal component (upper panel) and the second principal component (bottom panel). The impact of the third factor ranges from 0 to 0.5 percent, and thus is not presented. The instantaneous conditional covariance of yields is given in expression (3.16). We consider 13 model-implied yields with maturities from 3 months to 10 years. The shaded areas are obtained as contours of scatter plots based on 72000 observations simulated from the model.

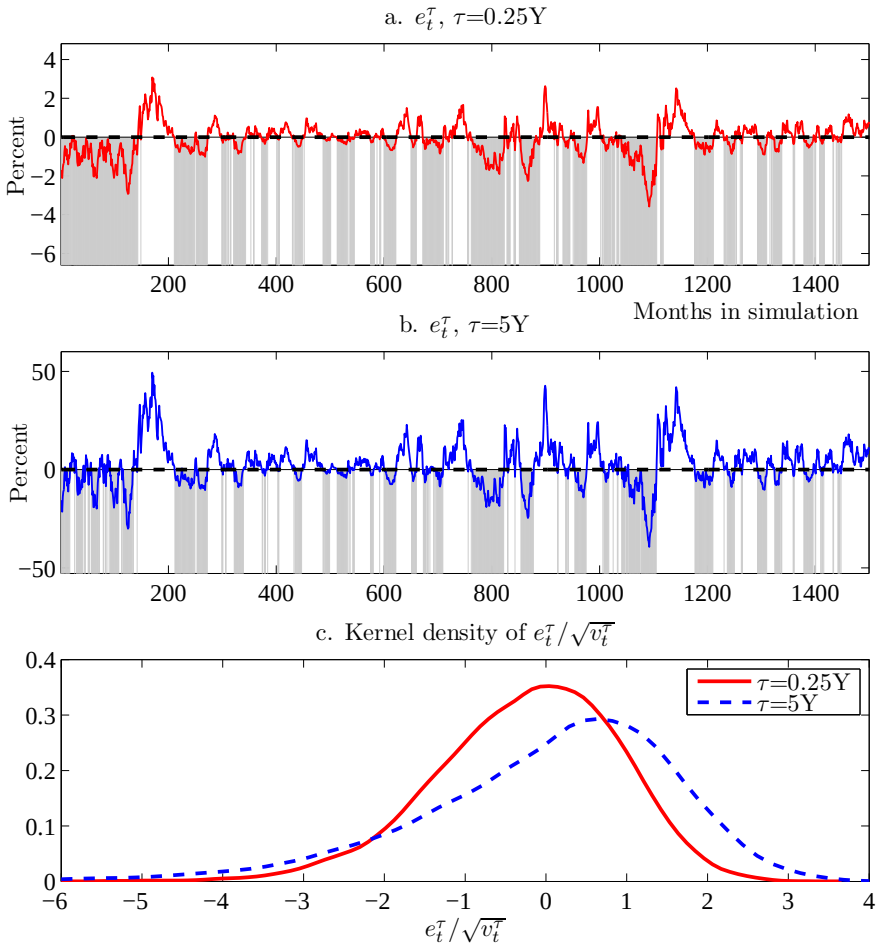


Figure 3.8: Properties of expected excess bond returns in the 2×2 WTSM

Panels a and b display the instantaneous expected excess returns on a 3-month and 5-year bond, respectively, as implied by the 2×2 Wishart factor model ($k = 3$). The expected excess returns are computed according to equation (3.18). Panel c plots the kernel density of the ratio of the instantaneous expected excess returns to their instantaneous volatility, $e_t^\tau/\sqrt{v_t^\tau}$, obtained by simulating 72000 observations from the model.

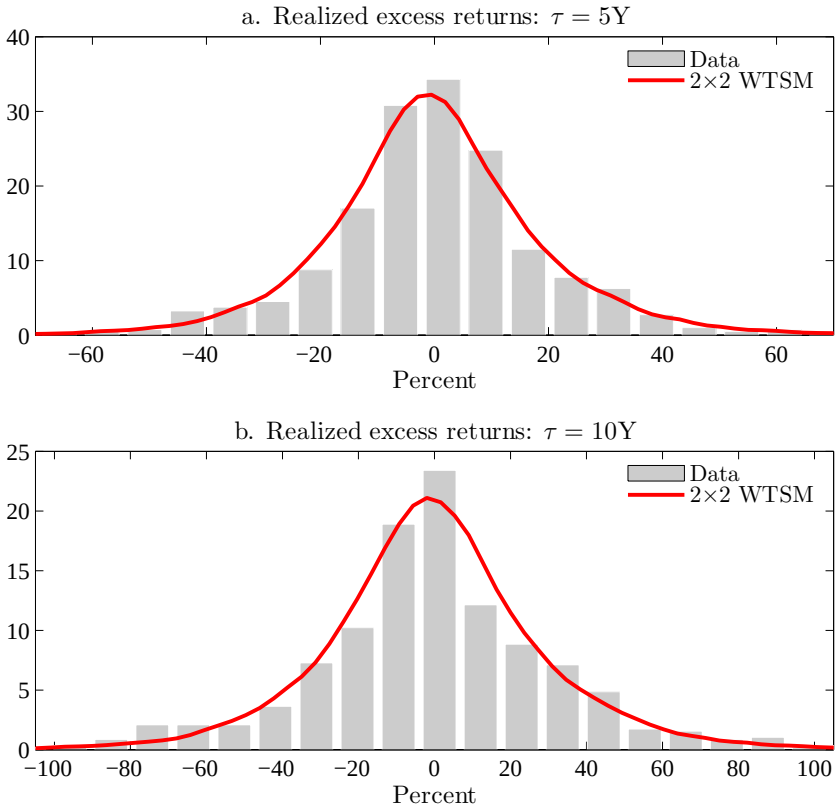


Figure 3.9: Properties of realized excess bond returns

Panels a and b display the distribution of realized monthly excess returns on 5-year and 10-year bonds, respectively. The realized excess returns implied by the 2×2 Wishart factor model ($k = 3$) are superposed with the histograms of realized excess returns on the corresponding US zero bonds (1952:01–2005:06). In both panels, the realized excess return is computed as the return on the long bond over the 3-month bond. All returns are annualized by multiplying with a factor 1200.

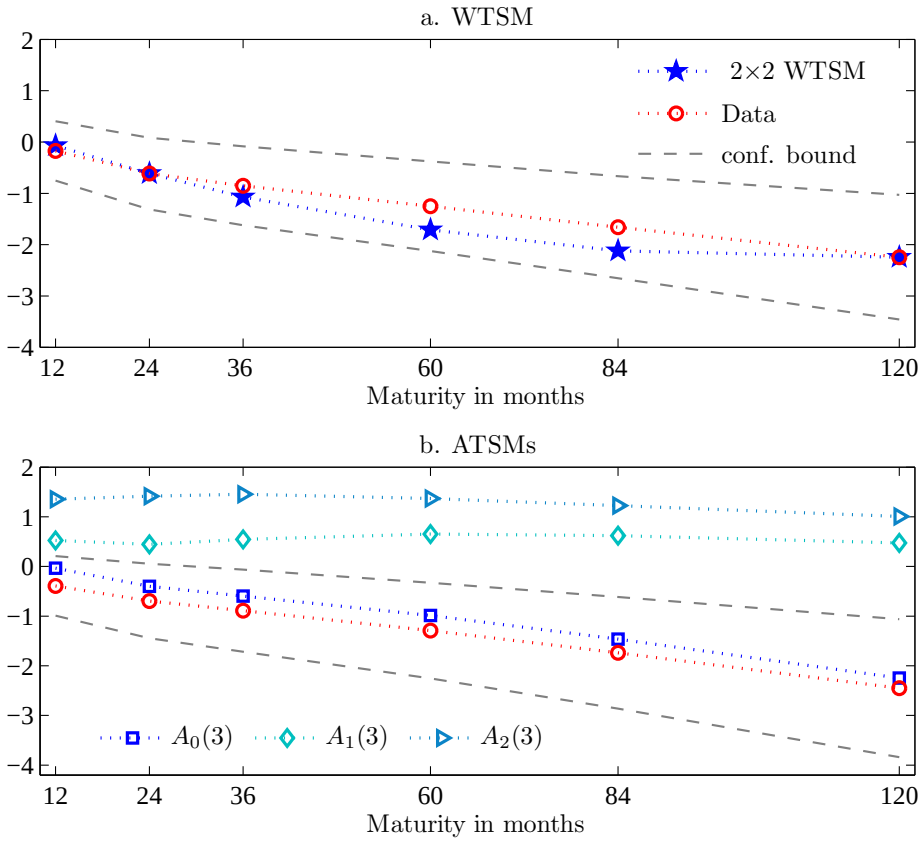


Figure 3.10: Campbell-Shiller regression coefficients

The figure plots—as a function of maturity—the parameters of Campbell and Shiller (1991) regression in equation (3.26). Panel *a* displays the coefficients obtained from the US yields in the sample period 1952:01–2005:06 and the theoretical coefficients implied by the 2×2 Wishart factor model ($k = 3$). Panel *b* performs the same exercise for the preferred affine models estimated by Duffee (2002), and compares them to the empirical coefficients for the relevant sample period 1952:01–1994:12. The dashed lines plot the 80 percent confidence bounds for the historical estimates based on the Newey-West covariance matrix. The 80 percent bound is lax for the data, but rigid for the model: Clearly, a more conservative choice (e.g. 90 percent) results in a still broader bound for the data, and thus is easier to match for the model. Further remarks from Table 3.1 apply.

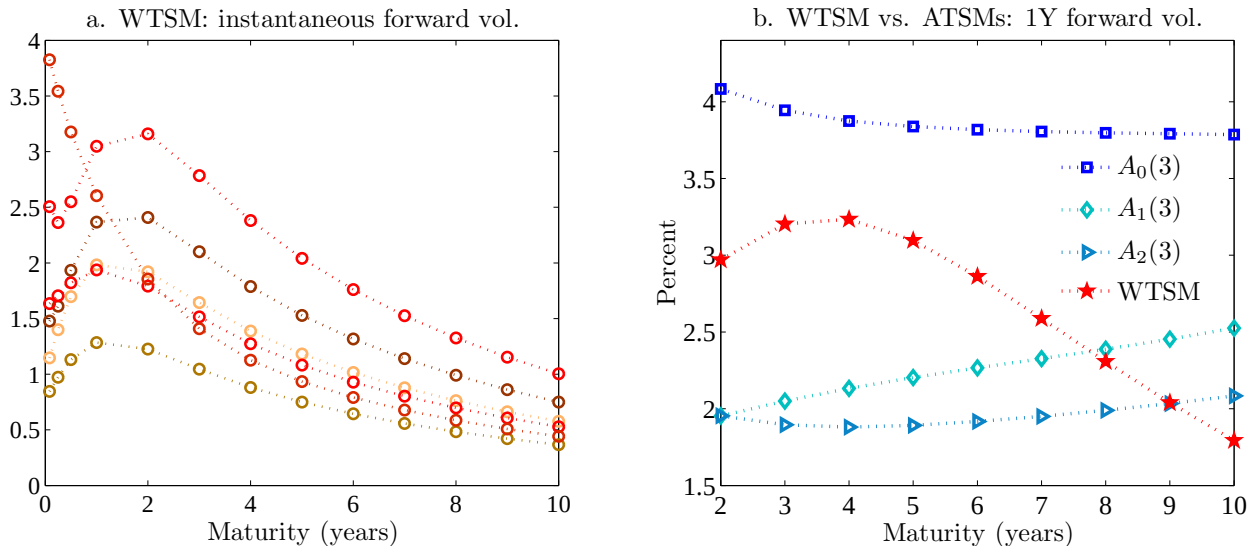


Figure 3.11: Term structure of forward interest rate volatilities

Panel *a* presents the term structure of the instantaneous volatility of the (instantaneous) forward rate given in equation (3.20), as implied by the 2×2 Wishart setting ($k = 3$). The instantaneous volatility is computed as $v^f(t,\tau) = 4Tr[\frac{dA(\tau)}{d\tau}\Sigma_t\frac{dA(\tau)}{d\tau}Q'Q]$, where $\frac{dA(\tau)}{d\tau}$ is given in closed form in equation (3.14). Panel *b* shows the theoretical unconditional volatility of the one-year forward rate from the same WTSM, and compares it with the preferred affine models of Duffee (2002): $A_0(3)$, $A_1(3)$ and $A_2(3)$. The x axis gives the maturity τ of the forward rate, $f_t^{\tau-1\to\tau} = \ln(P_t^{\tau-1}/P_t^{\tau})$.

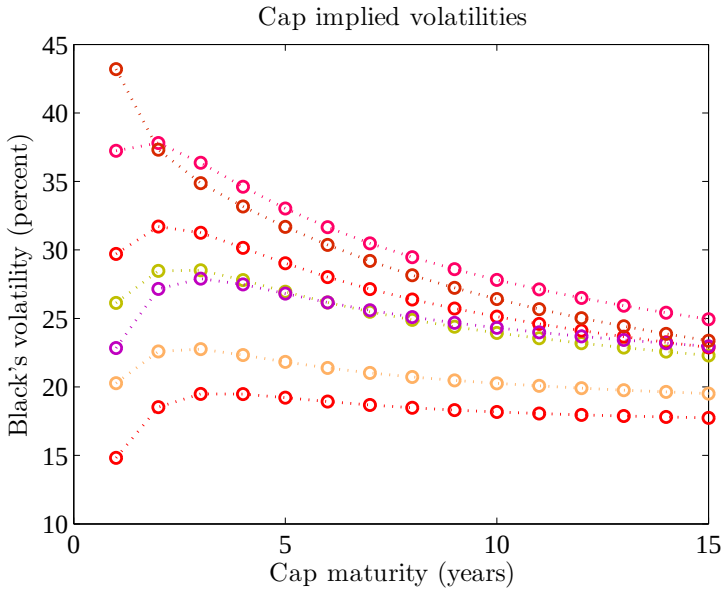


Figure 3.12: Term structure of cap implied volatilities

The figure exhibits several term structures of cap implied volatilities in the 2×2 Wishart model ($k = 3$), conditional on different values of the state matrix. The 3-month interest rate is the basis for each cap.

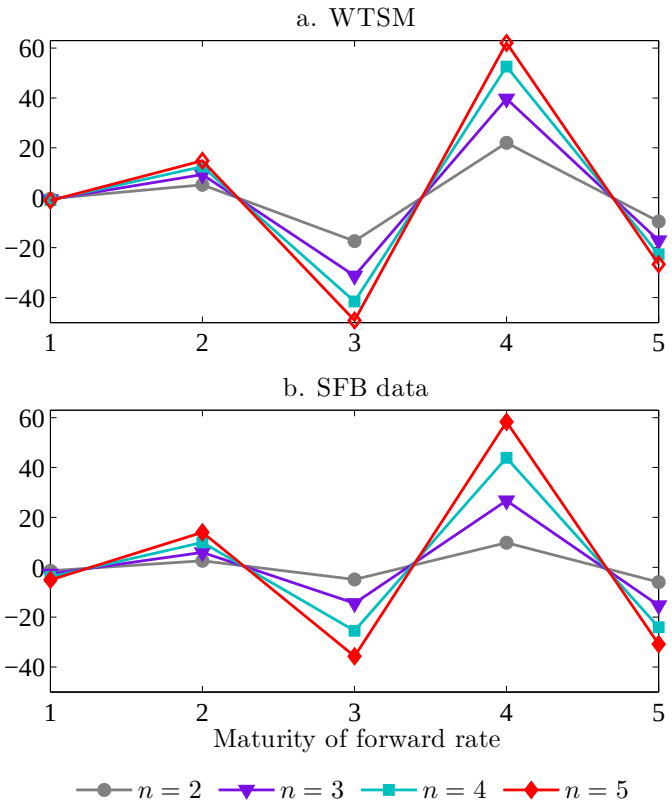


Figure 3.13: Cochrane-Piazzesi projection coefficients

The figure presents the slope coefficients in regressions of individual one-year excess bond returns on the set of one-year forward rates $f^{\tau-1\to\tau}, \tau = 1,2,3,4,5$, as a function of maturity τ. The legend $n = 2,3,4,5$ refers to the maturity of the bond whose excess return is forecast. Panel a displays coefficients in the simulated 3×3 Wishart economy ($k = 3$). The regressions are run on a sample of 72000 monthly observations from the model. Panel b shows the loadings in the smooth Fama-Bliss (SBF) data set used by Dai, Singleton, and Yang (2004). The sample is monthly and spans the period 1970:01–2000:12.

Appendix C

Throughout this Appendix the notation $A > B$, for two conformable matrices A and B, should be understood as their difference $A - B > 0$ being a positive definite matrix.

C.1 Proofs: Term Structure of Interest Rates

C.1.A Second Moments of the Short Interest Rate

Let $D - I_n = C$, a symmetric matrix. In a first step, we derive the expression for the instantaneous variance of the interest rate. Applying Ito's Lemma to the interest rate, we have that $dr = Tr\left(Cd\Sigma\right)$. By Result 3 in Appendix C.4, we obtain the expression (3.6):

$$\begin{aligned} Var_t\left[Tr(Cd\Sigma)\right] &= 4Tr\left(C\Sigma CQ'Q\right)dt = 4Tr\left[(D-I_n)\Sigma(D-I_n)Q'Q\right]dt \\ &= 4Tr\left[(D-I_n)Q'Q(D-I_n)\Sigma\right]dt > 0. \end{aligned}$$

With similar arguments, the expression for the covariance between the changes in the level and the variance of interest rate follows. Let $(D - I_n)Q'Q(D - I_n) = P$, a symmetric matrix, and apply Ito's Lemma to V_t:

$$\begin{aligned} Cov_t\left(dr, dV\right) &= Cov_t\left[Tr\left(Cd\Sigma\right), Tr\left(Pd\Sigma\right)\right] = 4Tr\left[P\Sigma CQ'Q\right]dt \\ &= 4Tr\left[\left(D-I_n\right)Q'Q\left(D-I_n\right)Q'Q\left(D-I_n\right)\Sigma\right]dt. \end{aligned}$$

Note that the multiplier of Σ, i.e. $H = \left(D - I_n\right)Q'Q\left(D - I_n\right)Q'Q\left(D - Id_n\right)$, is again a symmetric matrix. ■

C.1.B Proof of Proposition 3: Solution for the Term Structure of Interest Rates

The coefficients $A(t,T)$ and $b(t,T)$ in the bond price expression are identified by inserting function (3.12) into the pricing PDE (3.10) and solving the resulting matrix Riccati equation. Note that $\mathcal{R}P = A(t,T)P$, and

$$\frac{\partial P}{\partial t} = P\left[\frac{d}{dt}b(t,T) + Tr\left(\frac{d}{dt}A(t,T)\Sigma\right)\right], \tag{C.1}$$

where for brevity P denotes the price at time t of a bond maturing at time T. The pricing PDE (3.10) can be expressed as:

$$Tr\left[(\Omega\Omega' + (M - Q')\Sigma + \Sigma(M' - Q))\,A + 2\Sigma AQ'QA\right] + \frac{db}{dt} + Tr\left(\frac{dA}{dt}\Sigma\right) - Tr\left[(D - I_n)\Sigma\right] = 0.$$

Matrix Riccati equation. The above equation holds for all t, T and Σ. By the matching principle, we get a system of ODEs in A and b:

$$-\frac{db}{dt} = Tr\left(\Omega\Omega' A\right) \tag{C.2}$$

$$-\frac{dA}{dt} = A(M - Q') + (M' - Q)A + 2AQ'QA - (D - I_n). \tag{C.3}$$

with the respective terminal conditions $b(T,T) = 0$ and $A(T,T) = 0$. For convenience, we consider $A(\cdot)$ and $b(\cdot)$ as parametrized by the time to maturity $\tau = T - t$. Clearly, this reparametrization merely requires the LHS of the above system to be multiplied by -1:

$$\frac{db}{d\tau} = Tr\left(\Omega\Omega' A\right) \tag{C.4}$$

$$\frac{dA}{d\tau} = A(M - Q') + (M' - Q)A + 2AQ'QA - (D - I_n), \tag{C.5}$$

with boundary conditions $A(0) = 0$ and $b(0) = 0$. We note that the instantaneous interest rate is:

$$r_t = \lim_{\tau \to 0} -\frac{1}{\tau}\log P(t,\tau) = -\frac{db(0)}{d\tau} - Tr\left(\frac{dA(0)}{d\tau}\Sigma_t\right) = Tr\left[(D - I_n)\Sigma_t\right].$$

Closed-form Solution to the Matrix Riccati Equation

The closed-form solution to the matrix Riccati equation (C.5) is obtained, via Radon's lemma, by linearizing the flow of the differential equation. For completeness, we give it in this Appendix. We express $A(\tau)$ as:

$$A(\tau) = H(\tau)^{-1}G(\tau), \tag{C.6}$$

for $H(\tau)$ invertible and $G(\tau)$ being a square matrix. Differentiating (C.6), we have:

$$\begin{aligned}\frac{d}{d\tau}\left[H(\tau)A(\tau)\right] &= \frac{dG(\tau)}{d\tau} \\ \frac{d}{d\tau}\left[H(\tau)A(\tau)\right] &= \frac{dH(\tau)}{d\tau}A(\tau) + H(\tau)\frac{dA(\tau)}{d\tau}.\end{aligned}$$

Premultiplying (C.5) by $H(\tau)$ gives:

$$H\frac{dA}{d\tau} = HA(M - Q') + H(M' - Q)A + 2HAQ'QA - H(D - I_n).$$

This is equivalent to:

$$\frac{dG}{d\tau} - \frac{dH}{d\tau}A = G(M - Q') + H(M' - Q)A + 2GQ'QA - H(D - I_n),$$

where for brevity we suppress the argument τ of $A(\cdot), H(\cdot)$ and $G(\cdot)$. After collecting coefficients of A in the last equation, we obtain the following matrix-valued system of ODEs:

$$\begin{aligned}\frac{dG(\tau)}{d\tau} &= G(M - Q') - H(D - I_n) \\ \frac{dH(\tau)}{d\tau} &= -2GQ'Q - H(M' - Q),\end{aligned}$$

or written compactly:

$$\frac{d}{d\tau}\begin{pmatrix} G(\tau) & H(\tau) \end{pmatrix} = \begin{pmatrix} G(\tau) & H(\tau) \end{pmatrix}\begin{pmatrix} M - Q' & -2Q'Q \\ -(D - I_n) & -(M' - Q) \end{pmatrix}.$$

The solution to the above ODE is obtained by exponentiation:

$$\begin{aligned}\begin{pmatrix} G(\tau) & H(\tau) \end{pmatrix} &= \begin{pmatrix} G(0) & H(0) \end{pmatrix}\exp\left[\tau\begin{pmatrix} M - Q' & -2Q'Q \\ -(D - I_n) & -(M' - Q) \end{pmatrix}\right] \\ &= \begin{pmatrix} A(0) & I_n \end{pmatrix}\exp\left[\tau\begin{pmatrix} M - Q' & -2Q'Q \\ -(D - I_n) & -(M' - Q) \end{pmatrix}\right] \\ &= \begin{pmatrix} A(0)C_{11}(\tau) + C_{21}(\tau) & A(0)C_{12}(\tau) + C_{22}(\tau) \end{pmatrix} \\ &= \begin{pmatrix} C_{21}(\tau) & C_{22}(\tau) \end{pmatrix},\end{aligned}$$

where we use the fact that $A(0) = 0$, and

$$\begin{pmatrix} C_{11}(\tau) & C_{12}(\tau) \\ C_{21}(\tau) & C_{22}(\tau) \end{pmatrix} := \exp\left[\tau\begin{pmatrix} M - Q' & -2Q'Q \\ -(D - I_n) & -(M' - Q) \end{pmatrix}\right].$$

From equation (C.6), the closed-form solution to (C.5) is given by:

$$A(\tau) = C_{22}(\tau)^{-1} C_{21}(\tau),$$

whenever it exists. Given the solution for $A(\tau)$, the coefficient $b(\tau)$ is obtained directly by integration, and admits the following closed form (da Fonseca, Grasselli, and Tebaldi, 2006):

$$b(\tau) = Tr\left(\Omega\Omega' \int_0^\tau A(s)\right) ds = -\frac{k}{2} Tr\left[\ln C_{22}(\tau) + \tau(M' - Q)\right].$$

■

Characterization of the Solution to the Riccati Equation

In this section, we discuss the definiteness and monotonic properties of the solution to the matrix Riccati equation (3.14) in Proposition 3. For convenience, we stick to the notation in equation (C.5) of the Appendix. Since both the equation (C.5) and the terminal condition $A(0)$ are real and symmetric, the solution must be real and symmetric on the whole interval $[0, \tau]$.

Negative definiteness of the solution. To discuss the definiteness of the solution, let us rewrite the equation:

$$\frac{dA(\tau)}{d\tau} = A(\tau)\widetilde{M} + \widetilde{M}'A(\tau) + 2A(\tau)Q'QA(\tau) + C, \tag{C.7}$$

$$A(\tau_0) = 0, \tag{C.8}$$

where $\tau_0 = 0$ and for brevity $C = -(D - I_n)$ and $\widetilde{M} = M - Q'$, as the time-varying *linear* ODE of the form:

$$\frac{dA(\tau)}{d\tau} = W'(\tau)A(\tau) + A(\tau)W(\tau) + C,$$

where $W(\tau) = \widetilde{M} + Q'QA(\tau)$. A well-known result from the control theory (see e.g. Brockett, 1970, p.59, p.162) allows us to state the solution to this equation as:[1]

$$A(\tau) = \Phi(\tau, \tau_0)A(\tau_0)\Phi'(\tau, \tau_0) + \int_{\tau_0}^{\tau} \Phi(s, \tau_0)C\Phi'(s, \tau_0)ds = \int_{\tau_0}^{\tau} \Phi(s, \tau_0)C\Phi'(s, \tau_0)ds, \tag{C.9}$$

where $\Phi(\cdot, \cdot)$ is the state transition matrix of the system matrix $W(\tau)$ solving $\dot{\Phi}(\tau, \tau_0) = W'(\tau)\Phi(\tau, \tau_0)$, $\Phi(\tau_0, \tau_0) = I_n$. The RHS of equation (C.9) represents a congruent transformation of matrix C. Congruent transformations may change the eigenvalues of a matrix but they cannot change the signs of the eigenvalues (Sylvester's law of inertia). Thus, given $A(\tau_0) = 0$, the necessary and sufficient condition for $A(\tau)$ to be negative definite is that $C < 0$, therefore $D - I_n > 0$.

[1] To be exact, the statement (C.9) should express the solution in terms of some $\Pi(\tau)$ (rather than $A(\tau)$), which solves the matrix Riccati ODE (C.7). However, to keep the notation simple, with some abuse of notation, we state the solution in terms of $A(\tau)$. This has no impact on the argument which ensues.

Monotonic properties of the solution. To prove the monotonicity of the solution, let us differentiate the equation (C.7) with respect to time to maturity, τ:

$$\begin{aligned}\ddot{A}(\tau) &= \dot{A}(\tau)\widetilde{M} + \widetilde{M}'\dot{A}(\tau) + 2\dot{A}(\tau)Q'QA(\tau) + 2A(\tau)Q'Q\dot{A}(\tau) \\ &= V'(\tau)\dot{A}(\tau) + \dot{A}(\tau)V(\tau),\end{aligned}$$

where $V(\tau) = \widetilde{M} + 2Q'QA(\tau)$, and for convenience we use dot and double-dot notation for the first and second derivative, respectively. The solution to this equation is given as:

$$\dot{A}(\tau) = \Phi(\tau, \tau_0)\dot{A}(\tau_0)\Phi'(\tau, \tau_0),$$

with the state transition matrix $\Phi(\tau, \tau_0)$ of the system matrix $V(\tau)$ solving $\dot{\Phi}(\tau, \tau_0) = V'(\tau)\Phi(\tau, \tau_0)$, $\Phi(\tau_0, \tau_0) = I_n$. By plugging the terminal condition $A(\tau_0) = 0$ into equation (C.7), we have:

$$\dot{A}(\tau_0) = C = -(D - I_n).$$

Therefore,

$$\dot{A}(\tau) = \Phi(\tau, \tau_0)C\Phi'(\tau, \tau_0),$$

which is negative definite if $C < 0$. By integrating the above expression on the interval $(s, t), s < t$ for some $s, t \in [\tau_0, \tau_1]$, we have:

$$A(t) - A(s) = \int_s^t \dot{A}(u)du < 0.$$

Hence, $A(\tau)$ declines with the time to maturity τ of the bond. ■

C.1.C Bond Returns

By Ito's Lemma, for a smooth function $\phi(\Sigma, t)$ we have:

$$d\phi = \left(\frac{\partial \phi}{\partial t} + \mathcal{L}_\Sigma \phi\right) dt + Tr\left[(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma})\mathcal{R}\phi\right], \tag{C.10}$$

where $\mathcal{L}_\Sigma$ denotes the infinitesimal generator of the Wishart process. Using this result, the drift of the bond price $P(\Sigma, t, T)$ can be written as:

$$\begin{aligned}\frac{1}{dt}E_t\left(dP\right) &= \frac{\partial P}{\partial t} + \mathcal{L}_\Sigma P \\ &= \frac{\partial P}{\partial t} + Tr\left[\left(\Omega\Omega' + M\Sigma + \Sigma'M\right)\mathcal{R}P + 2\Sigma\mathcal{R}(Q'Q\mathcal{R}P)\right].\end{aligned}$$

From the fundamental PDE (3.10), we note that at equilibrium the drift must satisfy:

$$\frac{1}{dt}E_t\left(dP\right) - Tr(\Phi_\Sigma \mathcal{R}P) = rP.$$

By taking derivatives of the bond price with respect to the Wishart matrix, $\mathcal{R}P = A(\tau)P$, it follows that the expected excess bond return (over the short rate) is given by:

$$\begin{aligned} e_t^\tau &= Tr\left[(A(\tau)Q' + QA(\tau))\Sigma_t\right] \\ &= 2Tr\left[QA(\tau)\Sigma_t\right]. \end{aligned} \tag{C.11}$$

For completeness, we also provide the expression for the instantaneous variance of the bond return. From equation (C.10), the diffusion part of the bond dynamics dP is given by $Tr\left[\left(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma}\right)A(\tau)P\right]$. Using Result 3 in Appendix C.4, the instantaneous variance of the bond returns is:

$$\begin{aligned} Var_t\left(\frac{dP}{P}\right) &= Var_t\left[Tr\left(\left(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma}\right)A(\tau)\right)\right] \\ &= 4Tr\left[A(\tau)\Sigma A(\tau)Q'Q\right]dt. \end{aligned}$$

■

C.1.D Dynamics of the Forward Rate

From the expression for the instantaneous forward rate:

$$f(t,\tau) = -\frac{\partial \log P(t,t+\tau)}{\partial \tau} = -\frac{\partial b(\tau)}{\partial \tau} - Tr\left[\frac{\partial A(\tau)}{\partial \tau}\Sigma_t\right],$$

the dynamics of $f(t,\tau)$ can be computed from:

$$df(t,\tau) = -\frac{\partial d\log P(t,t+\tau)}{\partial \tau}.$$

By Ito's Lemma, we first obtain the dynamics of the logarithm of the bond price from equation (3.12):

$$d\log P = \left[-\frac{\partial b}{\partial \tau} - Tr\left(\frac{\partial A}{\partial \tau}\Sigma\right) + Tr\left[(\Omega\Omega' + M\Sigma + \Sigma M')A\right]\right]dt + Tr\left[(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma})A\right].$$

By noting that the first two terms in the drift of $d\log P$ equal the forward rate $f(t,\tau)$, we arrive at the instantaneous forward rate dynamics:

$$df(t,\tau) = -\left(\frac{\partial f}{\partial \tau} + Tr\left[(\Omega\Omega' + M\Sigma + \Sigma M')\frac{\partial A}{\partial \tau}\right]\right)dt - Tr\left[\left(\sqrt{\Sigma}dBQ + Q'dB'\sqrt{\Sigma}\right)\frac{\partial A}{\partial \tau}\right].$$

■

C.1.E Pricing of Zero-Bond Options

Let $\text{ZBC}(t, \Sigma_t; S, T, K)$ denote the price of a European option with expiry date S and exercise price K, written on a zero-bond maturing at time $T \geq S$:

$$\text{ZBC}(t, \Sigma_t; S, T, K) = P(t,T)\,\text{Pr}_t^T\{P(S,T) > K\} - KP(t,S)\,\text{Pr}_t^S\{P(S,T) > K\}.$$

Change of Drift for the Wishart Factors: The Forward Measure

To evaluate the two probabilities Pr_t^T and Pr_t^S in the above expression, we need to obtain the dynamics of the Wishart process under the two forward measures associated with bonds maturing at time S and T, respectively.

The risk-neutral dynamics of a S-maturity zero-bond $P(t,S)$ are:

$$\frac{dP(t,S)}{P(t,S)} = r_t dt + Tr(\Theta'(t,S)dB_t^*) + Tr(\Theta(t,S)dB_t^{*\prime}), \tag{C.12}$$

where $\Theta(t,S) = \sqrt{\Sigma_t}A(t,S)Q'$, $A(t,S)$ and $\sqrt{\Sigma_t}$ are symmetric, and $A(t,S)$ solves the matrix Riccati equation (C.3). The transformation from the risk neutral measure $\mathbb{Q}^*$ to the forward measure $\mathbb{Q}^S$ is given by:

$$\frac{d\mathbb{Q}^S}{d\mathbb{Q}^*}|_{\mathcal{F}_S} = e^{Tr\left[\int_0^S \Theta'(u,S)dB_u^* - \frac{1}{2}\int_0^S \Theta'(u,S)\Theta(u,S)du\right]},$$

where we use the fact that $Tr(\Theta'(t,S)dB_t^*) = (\text{vec}\Theta(t,S))'\text{vec}(dB_t^*)$. By Girsanov's theorem it follows:

$$dB_t^* = dB_t^S + \sqrt{\Sigma_t}A(t,S)Q'dt, \tag{C.13}$$

where dB_t^S is a $n \times n$ matrix of standard Brownian motions under $\mathbb{Q}^S$. Arguing similarly, we have:

$$dB_t^* = dB_t^T + \sqrt{\Sigma_t}A(t,T)Q'dt, \tag{C.14}$$

where dB_t^T is a $n \times n$ matrix of standard Brownian motions under $\mathbb{Q}^T$.

Remark 5. The measure transformations presented here are standard, but for the matrix-trace notation. Equivalently, we could use the vector notation for the T-maturity bond dynamics:

$$\frac{dP_t}{P_t} = r_t dt + \text{vec}(\Theta)'\text{vec}(dB_t^*) + \text{vec}(\Theta')'\text{vec}(dB_t^{*\prime}) \tag{C.15}$$

Then:

$$\text{vec}(dB_t^*) = \text{vec}(\Theta)dt + \text{vec}(dB_t^T) = \text{vec}(\Theta dt + dB_t^T).$$

Reversing the vec operation, we obtain a matrix of Brownian motions: $dB_t^* = dB_t^T + \Theta dt$. $\square$

Recall that the risk-neutral dynamics of the Wishart process is given by:

$$d\Sigma_t = (\Omega\Omega' + (M - Q')\Sigma_t + \Sigma_t(M' - Q))\,dt + \sqrt{\Sigma_t}dB_t^* Q + Q' dB_t^{*\prime}\sqrt{\Sigma_t}. \tag{C.16}$$

We are now ready to express the dynamics of the process under the S-forward measure:

$$d\Sigma_t \quad = \quad \{\Omega\Omega' + [M - Q'(I_n - QA)]\Sigma_t + \Sigma_t[M' - (I_n - AQ')Q]\}dt + \sqrt{\Sigma_t}dB_t^S Q + Q' dB_t^{S'}\sqrt{\Sigma_t},$$

where for brevity we write A for $A(t, S)$. The dynamics under the T-forward measure $\mathbb{Q}^T$ follows analogously.

Pricing of Zero-Bond Option by Fourier Inversion

Due to the affine property of the Wishart process, the conditional characteristic function of log-bond prices is available in closed form. Thus, the pricing of bond options amounts to performing two *one-dimensional* Fourier inversions under the two forward measures (see e.g., Duffie, Pan, and Singleton (2000)). We note that:

$$Pr_t^j\{P(S,T) > K\} = Pr_t^j\{b(S,T) + Tr\left[A(S,T)\Sigma_S\right] > \ln K\}, \text{ where } j = \{S,T\}.$$

To evaluate this probability by Fourier inversion, we find the characteristic function of the random variable $Tr[A(S,T)\Sigma_S]$ under the S- and T-forward measures. Let $\tau = S - t$, then the conditional characteristic function is:

$$\Psi_t^S(iz;\tau) = E_t^S\left(e^{izTr[A(t+\tau,T)\Sigma_{t+\tau}]}\right), \tag{C.17}$$

where E_t^S denotes the conditional expectation under the S-forward measure, $i = \sqrt{-1}$, and $z \in \mathbb{R}$. In the sequel, we show the argument for the S-forward measure, the argument for the T-forward measure being analogous. By the affine property of Σ_t, the characteristic function is itself of the exponentially affine form in Σ_t:

$$\Psi_t^S(iz;\tau) = e^{Tr[\hat{A}(z,\tau)\Sigma_t] + \hat{b}(z,\tau)}, \tag{C.18}$$

where $\hat{A}(z,\tau)$ and $\hat{b}(z,\tau)$ are, respectively, a symmetric matrix and a scalar with possibly complex coefficients, which solve the system of matrix Riccati equations (C.20)–(C.21) detailed below. With the characteristic functions of $Tr[A(S,T)\Sigma_S]$ for the S- and T-forward measure at hand, we can express the bond option price by the Fourier inversion as:

$$\begin{aligned}\text{ZBC}(t,S,T) &= P(t,T)\left\{\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \text{Re}\frac{e^{-iz[\log K - b(S,T)]}\Psi_t^T(iz;\tau)}{iz}dz\right\}\\ &\quad -KP(t,S)\left\{\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \text{Re}\frac{e^{-iz[\log K - b(S,T)]}\Psi_t^S(iz;\tau)}{iz}dz\right\},\end{aligned}$$

in which the integral can be evaluated by numerical methods.

The coefficients $\hat{A}(z,\tau)$ and $\hat{b}(z,\tau)$ in (C.18) are derived by the same logic as in Appendix C.1.B. By the Feynman-Kač argument applied to (C.17), Ψ_t^S solves the following PDE:

$$\frac{\partial \Psi_t^S}{\partial \tau} = \mathcal{L}_\Sigma \Psi_t^S. \tag{C.19}$$

Then, plugging for Ψ_t^S the expression (C.18), and collecting terms, gives the system of ordinary differential equations:

$$\frac{\partial \hat{b}(z,\tau)}{\partial \tau} = Tr[\Omega\Omega'\hat{A}(z,\tau)] \tag{C.20}$$

$$\frac{\partial \hat{A}(z,\tau)}{\partial \tau} = \hat{A}(z,\tau)M^S + M^{S'}\hat{A}(z,\tau) + 2\hat{A}(z,\tau)Q'Q\hat{A}(z,\tau), \tag{C.21}$$

where $M^S = M - Q'[I_n - QA(t,S)]$ results from the drift adjustment under the S-forward measure (given in Appendix C.1.E). The boundary conditions at $\tau = 0$ are:

$$\begin{aligned}\hat{b}(0) &= 0\\ \hat{A}(0) &= ziA(S,T).\end{aligned}$$

By Radon's lemma, the solution for $\hat{A}(\tau)$ reads:

$$\hat{A}(\tau) = (ziA(S,T)\hat{C}_{12} + \hat{C}_{22})^{-1}(ziA(S,T)\hat{C}_{11} + \hat{C}_{21}), \tag{C.22}$$

with

$$\begin{pmatrix} \hat{C}_{11}(\tau) & \hat{C}_{12}(\tau) \\ \hat{C}_{21}(\tau) & \hat{C}_{22}(\tau) \end{pmatrix} := \exp\left[\tau \begin{pmatrix} M^S & -2Q'Q \\ 0 & -M^{S'} \end{pmatrix}\right].$$

The coefficient $\hat{b}(z,\tau)$ is obtained by integration:

$$\hat{b}(z,\tau) = \int_0^\tau Tr[\Omega\Omega'\hat{A}(u)]du = -\frac{k}{2}Tr\left[\log(ziA(S,T)\hat{C}_{12}(\tau) + \hat{C}_{22}(\tau)) + \tau M^{S'}\right].$$

Remark 6. Let $\mathbf{A} = \tau \begin{pmatrix} M & -2Q'Q \\ 0 & -M' \end{pmatrix}$. Since $\exp(\mathbf{A}) = \sum_{i=0}^{\infty} \frac{\mathbf{A}^i}{i!}$, then using the rules for the product of block matrices, it is easily seen that the blocks of the matrix $\exp(\mathbf{A})$ are of the simple form:

$$\begin{aligned}
\hat{C}_{11}(\tau) &= e^{\tau M} \\
\hat{C}_{12}(\tau) &= 2\sum_{d=1}^{\infty} \frac{1}{d!}\tau^d \sum_{j=1}^{d} (-1)^j M^{d-j} Q'Q (M')^{j-1} \\
\hat{C}_{21}(\tau) &= 0_{n\times n} \\
\hat{C}_{22}(\tau) &= e^{-\tau M'}.
\end{aligned}$$

□

C.2 Relation to the Quadratic Term Structure Models (QTSMs)

For an integer degree of freedom k, the $n \times n$ state matrix Σ_t can be represented as the sum of k outer products of n-dimensional Ornstein-Uhlenbeck (OU) processes: $\Sigma_t = \sum_{i=1}^{k} X_t^i X_t^{i\prime}$. The OU dynamics is given by $dX_t^i = MX_t^i dt + Q'dW_t^i$, where dW_t^i is a n-vector of independent Brownian motions, and dW_t^i, dW_t^j are independent for $i \neq j$. First, we show the equivalence of the dynamics $d\left(\sum_{i=1}^{k} X_t^i X_t^{i\prime}\right)$ and $d\Sigma_t$. Then, we discuss the link to QTSMs.

By the independence of the OU processes, we can write $d(\sum_{i=1}^{k} X_t^i X_t^{i\prime}) = \sum_{i=1}^{k} d(X_t^i X_t^{i\prime})$, where:

$$\begin{aligned}
d(X_t^i X_t^{i\prime}) &= X_t^i dX_t^{i\prime} + dX_t^i X_t^{i\prime} + dX_t^i dX_t^{i\prime} \\
&= X_t^i (MX_t^i dt + Q'dW_t^i)' + (MX_t^i dt + Q'dW_t^i)X_t^{i\prime} + (MX_t^i dt + Q'dW_t^i)(MX_t^i dt + Q'dW_t^i)' \\
&= (Q'Q + MX_t^i X_t^{i\prime} + X_t^i X_t^{i\prime} M')dt + Q'dW_t X_t^{i\prime} + X_t^i dW_t^{i\prime} Q.
\end{aligned}$$

By summing over $i = 1, \ldots, k$, we get:

$$d(\sum_{i=1}^{k} X_t^i X_t^{i\prime}) = (kQ'Q + M\sum_{i=1}^{k} X_t^i X_t^{i\prime} + \sum_{i=1}^{k} X_t^i X_t^{i\prime} M')dt + Q'\sum_{i=1}^{k} dW_t^i X_t^{i\prime} + \sum_{i=1}^{k} X_t^i dW_t^{i\prime} Q. \tag{C.23}$$

Clearly, the drift in expression (C.23) is identical to the drift of $d\Sigma_t$ in equation (3.3). Next, we show the distributional equivalence between the diffusion parts in (3.3) and (C.23). It suffices to consider the instantaneous covariance between the following matrix forms (a, b, c, f are n-vectors):

$$\begin{aligned}
Cov_t[a'd(X_t^i X_t^{i\prime})b, c'd(X_t^i X_t^{i\prime})f] &= E[a'(X_t^i dW_t^{i\prime} Q + Q'dW_t^i X_t^{i\prime})b \;\; c'(X_t^i dW_t^{i\prime} Q + Q'dW_t^i X_t^{i\prime})f] \\
&= (b'Q'Qf \; a'X_t^i X_t^{i\prime} c + b'Q'Qc \, a'X_t^i X_t^{i\prime} f + a'Q'Qf \; b'X_t^i X_t^{i\prime} c + a'Q'Qc \, b'X_t^i X_t^{i\prime} f)dt.
\end{aligned} \tag{C.24}$$

Table C.1: Quadratic versus Wishart factor model

The table presents a mapping between the QTSM of Ahn, Dittmar, and Gallant (2002) and the WTSM. For readability, we preserve the respective notations.

QTSM(n)	WTSM($n \times n$) $k = 1$
State variables	
$dY_t = (\mu + \xi Y_t)dt + \Sigma dW_t$	$dX_t = MX_t dt + Q' dW_t$ $\mu = 0,\ \xi = M,\ \Sigma = Q'$
Short rate	
$r_t = \alpha + \beta' Y_t + Y_t' \Psi Y_t$ $\beta = 0$ (identification)	$r_t = X_t'(D - I_n)X_t$ $\alpha = 0,\ \beta = 0$ $\Psi = D - I_n$
Market price of risk	
$\Lambda_t = \delta_0 + \delta_1 Y_t$	$\Lambda_t = X_t$ $\delta_0 = 0,\ \delta_1 = I_n$

Note that when a, b, c, f are different unit vectors in $\mathbb{R}^n$, expression (C.24) characterizes all second moments of $d(X_t^i X_t^{i\prime})$. By independence of $X_t^i, X_t^j, i \neq j$, the result easily extends to $\sum_{i=1}^k X_t^i X_t^{i\prime}$:

$$
\begin{aligned}
&Cov_t[a' d(\sum_{i=1}^k X_t^i X_t^{i\prime}) b, c' d(\sum_{i=1}^k X_t^i X_t^{i\prime}) f] \\
&= (b'Q'Qf\ a' \sum_{i=1}^k X_t^i X_t^{i\prime} c + b'Q'Qc\ a' \sum_{i=1}^k X_t^i X_t^{i\prime} f + a'Q'Qf\ b' \sum_{i=1}^k X_t^i X_t^{i\prime} c + a'Q'Qcb' \sum_{i=1}^k X_t^i X_t^{i\prime} f) dt.
\end{aligned}
\tag{C.25}
$$

The expression (C.25) is equivalent to the covariation between different elements of Σ_t in Result 1 of Appendix C.4. Thus, the diffusion parts of $d(\sum_{i=1}^k X_t^i X_t^{i\prime})$ and $d\Sigma_t$ are distributionally equivalent.

When $k = 1$, Σ_t becomes singular. We can recast the model in terms of the single OU vector process as:

$$
d(X_t^1 X_t^{1\prime}) = (Q'Q + MX_t^1 X_t^{1\prime} + X_t^1 X_t^{1\prime} M')dt + Q' dW_t X_t^{1\prime} + X_t^1 dW_t^{1\prime} Q, \tag{C.26}
$$

where $dX_t^1 = MX_t^1 dt + Q' dW_t^1$. In this special case, our model has a direct analogy to a $n-$factor quadratic term structure model (QTSM) of Ahn, Dittmar, and Gallant (2002) and Leippold and Wu (2002), with the underlying OU dynamics of X_t^1 that has no intercept. Table C.1 compares the corresponding elements in the two settings. We preserve the notation used by Ahn, Dittmar, and Gallant (2002) for the QTSMs, and map it into the notation used throughout our paper.

C.2.A Unique invertibility of the state

The affine property of yields in the elements of Σ_t represents an advantage of our setting over the QTSMs. When $k \geq n$, the state variables in Σ_t can be *uniquely* backed out from the observed yields. In particular, an $n \times n$ state matrix Σ_t can be identified from $\bar{n} = \frac{n(n+1)}{2}$ yields. Let us stack the yields in a vector using the fact that $Tr\left[A\left(\tau\right)\Sigma_t\right] = \left[\text{vec}A(\tau)\right]' \text{vec}\Sigma_t$:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_{\bar{n}}) \end{pmatrix} = - \begin{pmatrix} \frac{1}{\tau_1} b(\tau_1) \\ \frac{1}{\tau_2} b(\tau_2) \\ \vdots \\ \frac{1}{\tau_n} b(\tau_{\bar{n}}) \end{pmatrix} - \begin{pmatrix} \frac{1}{\tau_1} \text{vec}A(\tau_1)' \\ \frac{1}{\tau_2} \text{vec}A(\tau_2)' \\ \vdots \\ \frac{1}{\tau_n} \text{vec}A(\tau_{\bar{n}})' \end{pmatrix} \text{vec}(\Sigma_t),$$

or in a short-hand vector-matrix notation:

$$\vec{y}_t = -\vec{b} - \mathbf{A}\ \text{vec}(\Sigma_t).$$

To be able to invert the last expression for the unique elements of Σ_t, we prune the non-unique elements of $\mathbf{A}$ and $\text{vec}(\Sigma_t)$ by using the half-vectorization:

$$\vec{y}_t = -\vec{b} - \mathbf{A}S_n\ \text{vech}\left(\Sigma_t\right),$$

where S_n is a duplication matrix of dimension $n^2 \times \frac{n(n+1)}{2}$ such that $S_n \text{vech}(\Sigma_t) = \text{vec}(\Sigma_t)$. It follows that the state is identified from yields as:

$$\text{vech}(\Sigma_t) = -(\mathbf{A}S_n)^{-1}(\vec{y}_t + \vec{b}).$$

C.3 Moments of the factors and yields

To provide a general formulation for the moments of the Wishart process, we proceed via the conditional Laplace transform. The derivation, which starts from the Laplace transform of the discrete time process, holds true also for non-integer degrees of freedom k, and thus does not require the restrictive interpretation of Σ_t as the sum of outer products of OU processes. The assumption of an integer k is implicit only via the mapping between the discrete and continuous time parameters of the process, which we discuss next.

C.3.A Moments of the discrete time Wishart process

The Wishart process allows for an exact discretization, i.e. there exists an explicit mapping between the discrete and continuous time parameters of the process:

$$\Phi_\Delta = e^{M\Delta} \tag{C.27}$$

$$V_\Delta = \int_0^\Delta \Phi_s Q' Q \Phi_s' ds, \tag{C.28}$$

where Δ denotes the discretization horizon. These expressions are the well-known conditional moments of the underlying multivariate OU process and hence are stated without proof.[2] For the tractability of subsequent derivations, we use the above mapping along with the conditional Laplace transform of the *discrete* time process to compute the moments of the *continuous* time process.

Laplace transform of the Wishart process

Let $\Sigma_\Delta|\Sigma_0 \sim Wis(k, \Phi, V)$. The conditional Laplace transform of $\Sigma_\Delta|\Sigma_0 = \Sigma$ is given by (see e.g. Muirhead, 1982, p. 442):

$$\begin{aligned}\Psi_\Delta(\Theta) &:= E_0[\exp(Tr(\Gamma\Sigma_\Delta))|\Sigma_0 = \Sigma] \\ &= \exp\left\{Tr[\Phi_\Delta'\Gamma(I_n - 2V_\Delta\Gamma)^{-1}\Phi_\Delta\Sigma] - \frac{k}{2}\log\det(I_n - 2V_\Delta\Gamma)\right\},\end{aligned}$$

where $\Gamma := \Gamma(\Theta)$ and $\Gamma = (\gamma_{ij}), i, j = 1, ..., n$ with $\gamma_{ij} = \frac{1}{2}(1 + \delta_{ij})\theta_{ij}$, where δ_{ij} is the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j. \end{cases}$$

The cumulant generating function is:

$$\mathcal{K}_\Delta(\Theta) := \log\Psi_\Delta(\Theta) = \left\{Tr[\Phi_\Delta'\Gamma(I_n - 2V_\Delta\Gamma)^{-1}\Phi_\Delta\Sigma] - \frac{k}{2}\log\det(I_n - 2V_\Delta\Gamma)\right\}.$$

In the sequel, we use the shorthand notation Γ to be understood as $\Gamma(\Theta)$. For brevity, the subscript Δ at Φ and V denoting the discretization horizon is neglected.

Moments of the Wishart process

The moments of the process are obtained by evaluating the derivatives of the Laplace transform (cumulant generating function) at $\Gamma = 0$.

We apply the following definition of the derivative of some function F (possibly matrix-valued) with respect to the matrix argument (Magnus and Neudecker, 1988, p. 173):

$$\mathcal{D}F(\Theta) := \frac{d\,\mathrm{vec}F(\Theta)}{d(\mathrm{vec}\Theta)'}.$$

[2] See e.g. Fisher and Gilles (1996b) for a detailed derivation of the conditional moments of a general affine process.

Lemma 3. *The closed-form expression for the first order derivative of the conditional cumulant generating function $\mathcal{K}(\Theta)$:*

$$\mathcal{D}\mathcal{K}(\Theta) = \frac{d\,\mathcal{K}(\Theta)}{d\,(\text{vec}\Theta)'} = P_1(\Theta) + P_2(\Theta) \tag{C.29}$$

where

$$P_1(\Theta) = \text{vec}[(I_n - 2V\Gamma)^{-1}\Phi\Sigma\Phi'(I_n - 2\Gamma V)^{-1}]'$$
$$P_2(\Theta) = k\ \text{vec}[V(I_n - 2\Gamma V)^{-1}]'.$$

Proof. The proof is an application of matrix calculus rules to the cumulant generating function:

$$\begin{aligned}\mathcal{D}K(\Theta) &= \frac{d\,K(\Theta)}{d\,(\text{vec}\Theta)'} \\ &= \underbrace{\frac{d\,Tr[\Phi'\Gamma(I_n - 2V\Gamma)^{-1}\Phi\Sigma]}{d\,(\text{vec}\Theta)'}}_{P_1(\Theta)} - \underbrace{\frac{k}{2}\frac{d\,(\log\det(I_n - 2V\Gamma))}{d\,(\text{vec}\Theta)'}}_{P_2(\Theta)}\end{aligned}$$

□

Corollary 4 (First moment of the Wishart process)**.** *From the results in Lemma 3, the expression for the first conditional moment of $\Sigma_\Delta|\Sigma_0 = \Sigma$ follows immediately:*

$$E((\text{vec}\Sigma_\Delta)'|\Sigma_0 = \Sigma) = \frac{dK(0)}{d(\text{vec}\Theta)'} = P_1(0) + P_2(0) = (\text{vec}(\Phi\Sigma\Phi' + kV))'.$$

Equivalently, in matrix notation:

$$E(\Sigma_\Delta|\Sigma_0 = \Sigma) = \Phi\Sigma\Phi' + kV.$$

□

Remark 7. The first moment of the Wishart process is straightforward to obtain for the *integer* degree of freedom k. Let x_t^i denote the OU process with the discrete time dynamics:

$$x_t^i = \Phi x_{t-\Delta}^i + \epsilon_t^i, \quad \text{where } \epsilon_t^i \sim N(0, V).$$

For an integer k, the Wishart is constructed as $\Sigma_t = \sum_i x_t^i x_t^{i\prime}$, and we have:

$$\sum_i x_t^i x_t^{i\prime} = \sum_i (\Phi x_{t-\Delta}^i + \epsilon_t)(\Phi x_{t-\Delta}^i + \epsilon_t)'$$
$$E_{t-\Delta}(\sum_i x_t^i x_t^{i\prime}) = \Phi\Sigma_t\Phi' + E_{t-\Delta}(\sum_i \epsilon_t^i \epsilon_t^{i\prime}) = \Phi\Sigma_t\Phi' + kV.$$

In contrast, our derivation via the Laplace transform is more generic as it does not rely (at least in the discrete time case) on the link between the Wishart and the OU process. Thus, the assumption of integer degrees of freedom is not required. Moreover, if we relax the restriction that $\Omega\Omega' = kQ'Q$ in the drift of the continuous time Wishart, then the conditional first moment can be further generalized to:

$$E_t\left(\Sigma_{t+\Delta}\right) = e^{\Delta M}\Sigma_t e^{\Delta M'} + \int_0^{\Delta} e^{sM}\Omega\Omega' e^{sM'} ds,$$

where the last expression follows from the Laplace transform of the continuous time process. □

The second derivative of $\mathcal{K}(\Theta)$ is defined as (see Magnus and Neudecker, 1988, p. 188):

$$\mathcal{H}\mathcal{K}(\Theta) = \mathcal{D}\left(\mathcal{D}\mathcal{K}(\Theta)\right)' = \frac{d\text{vec}\left[\mathcal{D}\mathcal{K}(\Theta)\right]'}{d\left(\text{vec}\Theta\right)'}. \tag{C.30}$$

Lemma 5. *The closed-form expression for the second order derivative (C.30) is defined as:*

$$\mathcal{H}K(\Theta) = \left(I_{n^2} + K_n\right)\left\{R_1(\Theta)\otimes R_2(\Theta) + K\left[R_2(\Theta)\otimes R_2(\Theta)\right] + R_2(\Theta)\otimes R_1(\Theta)\right\},$$

where

$$\begin{aligned} R_1(\Theta) &= \left(I_n - 2V\Gamma\right)^{-1}\Phi\Sigma\Phi'\left(I_n - 2\Gamma V\right)^{-1} \\ R_2(\Theta) &= \left(I_n - 2V\Gamma\right)^{-1}V, \end{aligned}$$

where I_{n^2} is an $n^2 \times n^2$ identity matrix, and $K_{n,n}$ is the commutation matrix defined as: $\text{vec}S' = K_{n,n}\text{vec}S$ *for some square matrix S.*

Proof. The above expression follows from taking the following derivatives:

$$\frac{d\,\text{vec}\left[\mathcal{D}\mathcal{K}(\Theta)\right]'}{d\left(\text{vec}\Theta\right)'} = \frac{d}{d(\text{vec}\Theta)'}\left[\frac{d\,\mathcal{K}(\Theta)}{d\left(\text{vec}\Theta\right)'}\right]' = \frac{d\,P_1'(\Theta)}{d(\text{vec}\Theta)'} + \frac{d\,P_2'(\Theta)}{d(\text{vec}\Theta)'}.$$

□

Corollary 6 (Second moment of the Wishart process)**.** *With the results in Lemma 5, the second conditional moments of $\Sigma_\Delta|\Sigma_0 = \Sigma$ follow:*

$$E[\text{vec}(\Sigma_\Delta)\text{vec}(\Sigma_\Delta)'|\Sigma_0 = \Sigma] = \mathcal{H}\Psi(0) = \left[\mathcal{D}\mathcal{K}(0)\right]'\left[\mathcal{D}\mathcal{K}(0)\right] + \mathcal{H}\mathcal{K}(0).$$

Therefore:

$$E[\text{vec}(\Sigma_\Delta)\text{vec}(\Sigma_\Delta)'|\Sigma_0 = \Sigma] = \text{vec}(\Phi\Sigma\Phi' + kV)\text{vec}(\Phi\Sigma\Phi' + kV)' + (I_{n^2} + K_{n,n})[\Phi\Sigma\Phi' \otimes V + k(V \otimes V) + V \otimes \Phi\Sigma\Phi'].$$

Importantly, Result 4 and 6 cover the general case of non-integer degrees of freedom $k > n - 1$. The integer degrees of freedom is implicit only through the mapping between the continuous and discrete time parameters in (C.27)–(C.28).

Closed-form expressions for the matrix integrals

The conditional moments of the Wishart involve the evaluation of:

$$V_\tau = \int_0^\tau \Phi_s Q' Q \Phi_s' ds. \tag{C.31}$$

The closed-form expression for this integral is given as:

$$\int_0^\tau e^{Ms} Q'Q e^{M's} ds = -\frac{1}{2}\hat{C}_{12}(\tau)\hat{C}_{11}'(\tau), \tag{C.32}$$

where $\hat{C}_{11}(\tau)$ and $\hat{C}_{12}(\tau)$ are blocks of the matrix exponential associated with the coefficients of the Laplace transform of the continuous time Wishart process:

$$\exp\left[\tau \begin{pmatrix} M & -2Q'Q \\ 0 & -M' \end{pmatrix}\right] = \begin{pmatrix} \hat{C}_{11}(\tau) & \hat{C}_{12}(\tau) \\ \hat{C}_{21}(\tau) & \hat{C}_{22}(\tau) \end{pmatrix}. \tag{C.33}$$

In applications, the expression (C.31) turns out numerically stable (for finite maturities) and computationally very efficient.

Proof. The elements of the matrix exponential can be expressed as (see Van Loan, 1978, Thm. 1):

$$\begin{aligned} \hat{C}_{11}(\tau) &= e^{M\tau} \\ \hat{C}_{12}(\tau) &= \int_0^\tau e^{M(\tau-s)}(-2Q'Q)e^{-M's} ds \\ \hat{C}_{21}(\tau) &= 0_{n\times n} \\ \hat{C}_{22}(\tau) &= e^{-M'\tau} \end{aligned}$$

Postmultiplying the expression for $\hat{C}_{12}(\tau)$ by $\hat{C}_{11}'(\tau) = e^{M'\tau}$ and applying the change of variable $u = \tau - s$, yields:

$$\hat{C}_{12}(\tau)\hat{C}_{11}'(\tau) = -\int_0^\tau e^{Mu}(2Q'Q)e^{M'u} du.$$

After reformulating, the result follows:

$$\int_0^\tau e^{Mu} Q'Q e^{M'u} du = -\frac{1}{2}\hat{C}_{12}(\tau)\hat{C}'_{11}(\tau).$$

□

A similarly tractable and computationally efficient expression is readily available for the limit of the integral (C.31), which occurs in the unconditional moments of the Wishart process. In vectorized form, we have:

$$\begin{aligned} \text{vec}V_\infty &= \text{vec}\left(\lim_{\tau\to\infty}\int_0^\tau \Phi_s Q'Q\Phi'_s ds\right) \\ &= -\left[(I_n \otimes M) + (M \otimes I_n)\right]^{-1}\text{vec}(Q'Q). \end{aligned} \tag{C.34}$$

Proof. We exploit the relationship between the integral (C.34) and the solution to the following Lyapunov equation:

$$MX + XM' = Q'Q,$$

which can be written as (see e.g. Laub, 2005, p. 145):

$$X = -\int_0^\infty e^{Ms} Q'Q e^{M's} ds.$$

At the same time, the solution for X can be expressed in closed-form using the relationship between the vec operator and the Kronecker product, $\text{vec}(IXM) = (M' \otimes I)\text{vec}X$. This results in:

$$\text{vec}X = \left[(I_n \otimes M) + (M \otimes I_n)\right]^{-1}\text{vec}(Q'Q).$$

Thus, the integral can be efficiently computed as:

$$\text{vec}\left(\int_0^\infty e^{Ms} Q'Q e^{M's} ds\right) = -\left[(I_n \otimes M) + (M \otimes I_n)\right]^{-1}\text{vec}(Q'Q).$$

□

C.3.B Moments of yields

For the unconditional moments of yields to exist, we require $M < 0$. Since the unconditional first moment is straightforward to obtain, we only focus on the second moment. Its general specification comprises a cross moments of yields, one of which possibly lagged:

$$
\begin{aligned}
Cov(y_{t+s}^{\tau_1}, y_t^{\tau_2}) &= \frac{1}{\tau_1\tau_2}\left\{E\left[Tr(A_{\tau_1}\Sigma_{t+s})Tr\left(A_{\tau_2}\Sigma_t\right)\right] - E\left[Tr(A_{\tau_1}\Sigma_{t+s})\right]E\left[Tr\left(A_{\tau_2}\Sigma_t\right)\right]\right\} \\
&= \frac{1}{\tau_1\tau_2}\left\{E\left[\left(\text{vec}A_{\tau_1}\right)'\text{vec}\Sigma_{t+s}\left(\text{vec}\Sigma_t\right)'\text{vec}A_{\tau_2}\right] - \left(\text{vec}A_{\tau_1}\right)'\text{vec}E\Sigma_{t+s}\left(\text{vec}E\Sigma_t\right)'\text{vec}A_{\tau_2}\right\} \\
&= \frac{1}{\tau_1\tau_2}\left\{E\left[\left(\text{vec}A_{\tau_1}\right)'\left[\text{vec}\left(\Phi_s\Sigma_t\Phi_s'\right)\right]\left(\text{vec}\Sigma_t\right)'\left(\text{vec}A_{\tau_2}\right)\right] - \left(\text{vec}A_{\tau_1}\right)'\left[\text{vec}\left(\Phi_sE\Sigma_t\Phi_s'\right)\right]\left(\text{vec}E\Sigma_t\right)'\text{vec}A_{\tau_2}\right\} \\
&= \frac{1}{\tau_1\tau_2}\left(\text{vec}A_{\tau_1}\right)'\left(\Phi_s\otimes\Phi_s\right)\left[E(\text{vec}\Sigma_t(\text{vec}\Sigma_t)') - \left(\text{vec}E\Sigma_t\right)\left(\text{vec}E\Sigma_t\right)'\right]\text{vec}A_{\tau_2} \\
&= \frac{1}{\tau_1\tau_2}\text{vec}\left(\Phi_sA_{\tau_1}\Phi_s'\right)'\left[Cov(\text{vec}\Sigma_t)\right]\text{vec}A_{\tau_2},
\end{aligned}
$$

where when moving from the second to the third line, we have used the law of iterated expectations:

$$
\begin{aligned}
E\left[\text{vec}\Sigma_{t+s}\left(\text{vec}\Sigma_t\right)'\right] &= E\left[E_t\left(\text{vec}\Sigma_{t+s}\left(\text{vec}\Sigma_t\right)'\right)\right] \\
&= E\left[\text{vec}\left(\Phi_s\Sigma_t\Phi_s'\right)\left(\text{vec}\Sigma_t\right)' + k\ \text{vec}V_s\left(\text{vec}\Sigma_t\right)'\right].
\end{aligned}
$$

To obtain a contemporaneous covariance, note that $\Phi_{s=0} = e^0 = I_n$. Therefore:

$$
Cov(y_t^{\tau_1}, y_t^{\tau_2}) = \frac{1}{\tau_1\tau_2}(\text{vec}A_{\tau_1})'\left[Cov(\text{vec}\Sigma_t)\right]\text{vec}A_{\tau_2}.
$$

C.4 Useful Results for the Wishart Process

Result 1. The following result facilitates the computation of the second moments of the Wishart process. Given $n \times n$ Wishart SDE $d\Sigma$ in equation (3.3) and arbitrary n-dimensional vectors a, b, c, f it follows:

$$
Cov_t\left(a'd\Sigma_tb, c'd\Sigma_tf\right) = \left(a'Q'Qfb'\Sigma_tc + a'Q'Qcb'\Sigma_tf + b'Q'Qfa'\Sigma_tc + b'Q'Qca'\Sigma_tf\right)dt.
$$

Covariances between arbitrary quadratic forms of $d\Sigma$ are linear combinations of quadratic forms of Σ. In particular, both drift and instantaneous covariances of the single components of the matrix process Σ are themselves affine functions of Σ.

Using the above results, it is straightforward to compute the (cross-)second moments of factors in the 2×2 case:

$$
\begin{aligned}
d\langle\Sigma_{11}\rangle_t &= 4\Sigma_{11}\left(Q_{11}^2 + Q_{21}^2\right)dt \\
d\langle\Sigma_{22}\rangle_t &= 4\Sigma_{22}\left(Q_{22}^2 + Q_{12}^2\right)dt \\
d\langle\Sigma_{12}\rangle_t &= \left[\Sigma_{11}\left(Q_{12}^2 + Q_{22}^2\right) + \Sigma_{22}\left(Q_{11}^2 + Q_{21}^2\right) + 2\Sigma_{12}\left(Q_{11}Q_{12} + Q_{21}Q_{22}\right)\right]dt \\
d\langle\Sigma_{11}, \Sigma_{22}\rangle_t &= 4\Sigma_{12}\left(Q_{11}Q_{12} + Q_{21}Q_{22}\right)dt \\
d\langle\Sigma_{11}, \Sigma_{12}\rangle_t &= \left[2\Sigma_{11}\left(Q_{11}Q_{12} + Q_{21}Q_{22}\right) + 2\Sigma_{12}\left(Q_{11}^2 + Q_{21}^2\right)\right]dt \\
d\langle\Sigma_{22}, \Sigma_{12}\rangle_t &= \left[2\Sigma_{22}\left(Q_{11}Q_{12} + Q_{21}Q_{22}\right) + 2\Sigma_{12}\left(Q_{22}^2 + Q_{12}^2\right)\right]dt,
\end{aligned}
$$

where Σ_{ij} and Q_{ij} denote the ij-th element of matrix Σ and Q, respectively. More generally, for an arbitrary dimension n of the state matrix, we obtain:

$$\begin{aligned} d\langle \Sigma_{ii} \rangle_t &= 4\Sigma_{ii} Q^{i'} Q^i dt \\ d\langle \Sigma_{ii}, \Sigma_{jj} \rangle_t &= 4\Sigma_{ij} Q^{i'} Q^j dt, \end{aligned}$$

where Q^i, Q^j denote the i-th and j-th column of the Q matrix, respectively.

Result 2. The special case of the above result has been given by Gourieroux (2006):

$$\begin{aligned} &Cov_t \left(\alpha' d\Sigma_t \alpha, \beta' d\Sigma_t \beta \right) \\ &\quad = Cov_t \left[\alpha' \left(\sqrt{\Sigma_t} dB_t Q + Q' dB_t' \sqrt{\Sigma_t} \right) \alpha, \beta' \left(\sqrt{\Sigma_t} dB_t Q + Q' dB_t' \sqrt{\Sigma_t} \right) \beta \right] \\ &\quad = E_t \left[\left(\alpha' \sqrt{\Sigma_t} dB_t Q \alpha + \alpha' Q' dB_t' \sqrt{\Sigma_t} \alpha \right) \left(\beta' \sqrt{\Sigma_t} dB_t Q \beta + \beta' Q' dB_t' \sqrt{\Sigma_t} \beta \right) \right] \\ &\quad = 4 \left(\alpha' \Sigma_t \beta \alpha' Q' Q \beta \right) dt, \end{aligned}$$

where for any n-dimensional vectors u and v it holds that:

$$\begin{aligned} E_t \left(dB_t u v' dB_t \right) &= E_t \left(dB_t' u v' dB_t' \right) = v u' dt \\ E_t \left(dB_t u v' dB_t' \right) &= E_t \left(dB_t' u v' dB_t \right) = v' u I_n dt. \end{aligned}$$

Result 3. Given square matrices A and C, we have:

$$Cov_t \left[Tr(A d\Sigma_t), Tr(C d\Sigma_t) \right] = Tr \left[(A + A') \Sigma_t (C + C') Q' Q \right] dt. \tag{C.35}$$

Moreover, for a square matrix A it holds that:

$$Var_t \left[Tr(A d\Sigma_t) \right] = Tr \left[(A + A') \Sigma_t (A + A') Q' Q \right] dt > 0 \quad \text{iff} \quad \Sigma_t > 0. \tag{C.36}$$

In contrast to Gourieroux and Sufana (2003), in obtaining these results we do not impose definiteness restrictions on A and C.

Proof. To derive the result, we can directly consider the expectation of the product of the two traces:

$$\begin{aligned} &Cov_t \left[Tr(A d\Sigma_t), Tr(C d\Sigma_t) \right] = \\ &E_t \left[Tr \left(A \sqrt{\Sigma_t} dB_t Q + A Q' dB_t' \sqrt{\Sigma_t} \right) Tr \left(C \sqrt{\Sigma_t} dB_t Q + C Q' dB_t' \sqrt{\Sigma_t} \right) \right]. \end{aligned}$$

The above expression can be split into the sum of four expectations. For brevity, we only provide the derivation for one of the terms:

$$
\begin{aligned}
E_t\left[Tr\left(A\sqrt{\Sigma_t}dB_tQ\right)\ Tr\left(CQ'dB_t'\sqrt{\Sigma_t}\right)\right] = \\
&= E_t\left[\left(\text{vec}\left(QA\sqrt{\Sigma_t}\right)'\right)'\left(\text{vec}\ dB_t\right)\left(\text{vec}\left(\sqrt{\Sigma_t}CQ'\right)'\right)'\left(\text{vec}\ dB_t'\right)\right] \\
&= E_t\left[\left(\text{vec}\left(QA\sqrt{\Sigma_t}\right)'\right)'\left(\text{vec}\ dB_t\right)\left(\text{vec}\ dB_t\right)'K_{n,n}\left(\text{vec}\left(\sqrt{\Sigma_t}CQ'\right)'\right)\right] \\
&= \left(\text{vec}\left(QA\sqrt{\Sigma_t}\right)'\right)'\left(\text{vec}\left(\sqrt{\Sigma_t}CQ'\right)\right)dt = Tr\left[CQ'QA\Sigma\right]dt.
\end{aligned}
$$

where $K_{n,n}$ is the commutation matrix (thus $K_{n,n} = K_{n,n}'$), and we apply the following facts:

$$
E_t\left[\left(\text{vec}\ dB_t\right)\left(\text{vec}\ dB_t\right)'\right] = E_t\left[\left(\text{vec}\ dB_t'\right)\left(\text{vec}\ dB_t'\right)'\right] = I_{n^2}dt
$$

and

$$
Tr\left(QA\sqrt{\Sigma_t}dB_t\right) = \left(\text{vec}(\sqrt{\Sigma_t}AQ')'\right)'\text{vec}(dB_t).
$$

To prove the positivity of $Var_t\left[Tr(Ad\Sigma_t)\right]$ in equation (C.36), note that:

$$
Tr\left[(A+A')\Sigma_t(A+A')Q'Q\right] = Tr\left[Q(A+A')\Sigma_t(A+A')Q'\right].
$$

The expression within the trace on the RHS is a congruent transformation of the Wishart matrix Σ, which (by Sylvester's law) can change the values but not the signs of matrix eigenvalues. Thus, provided that $\Sigma_t > 0$, it follows that $Q(A+A')\Sigma_t(A+A')Q' > 0$. Combining this result with the properties of the trace we have:

$$
Tr\left[Q(A+A')\Sigma_t(A+A')Q'\right] = \sum_{i=1}^{n}\lambda_i > 0,
$$

where λ_i denotes the eigenvalue of $Q(A+A')\Sigma_t(A+A')Q'$. □

Result 4. If Σ_t is a Wishart process and C is a positive definite matrix, then the scalar process $Tr(C\Sigma_t)$ is positive (see Gourieroux (2006)).

Proof. By the singular value decomposition, a symmetric (positive or negative) definite $n \times n$ matrix D can be written as $D = \sum_{j=1}^{n}\lambda_j m_j m_j'$, where λ_j and m_j are the eigenvalues and eigenvectors of D, respectively. Let D be positive definite, and we get:

$$
\begin{aligned}
Tr(D\Sigma) &= Tr\left(\sum_{j=1}^{n}\lambda_j m_j m_j'\Sigma\right) = \sum_{j=1}^{n}\lambda_j Tr(m_j m_j'\Sigma) \\
&= \sum_{j=1}^{n}\lambda_j Tr(m_j'\Sigma m_j) = \sum_{j=1}^{n}\lambda_j m_j'\Sigma m_j > 0,
\end{aligned}
$$

where we use the facts that: (i) we can commute within the trace operator, (ii) $\lambda_j > 0$ for all j, and (iii) Σ is positive definite. Note that for a positive definite $n \times n$ matrix $D = \sum_{j=1}^{n} a_j a_j'$, where $a_j = \sqrt{\lambda_j} m_j$. □

C.5 Details on the Estimation Approach

C.5.A The 2×2 Model

The 2×2 model comprises 9 parameters: the elements of matrices M, Q' (both lower triangular) and D (symmetric), plus an integer value of the degrees of freedom parameter k. The estimation is based on 11 moment conditions (see Table C.2). The following steps describe our optimization technique:

Step 1. For M, Q and D, generate $I_{max} = 200$ from the uniform distribution under the condition that: (i) the diagonal elements of M are negative, (ii) the eigenvalues of $D - I_n$ are positive.

Step 2. Select the possible degrees of freedom k on a grid of integers from 1 to 9.

Step 3. Run 9×200 optimizations (for each k and each I_{max}) in order to select 10 parameter sets with the lowest value of the loss function.

Step 4. To determine the final parameter values, improve on the selected parameter sets using a gradient-based optimization routine (Matlab `lsqnonlin`).

This optimization procedure gives rise to $k = 3$ and to the parameters:

$$D = \begin{pmatrix} 1.0281 & 0.0047 \\ 0.0047 & 1.0018 \end{pmatrix},$$

$$M = \begin{pmatrix} -0.1263 & 0 \\ 0.0747 & -0.6289 \end{pmatrix},$$

$$Q = \begin{pmatrix} 0.4326 & -2.9238 \\ 0 & -0.5719 \end{pmatrix}.$$

The matrices satisfy the theoretical requirements: (i) the M matrix is negative definite, i.e. Σ_t does not explode; (ii) the Q matrix is invertible, i.e. Σ_t is reflected towards positivity whenever the boundary of the state space is reached; and (iii) the $D - I_2$ matrix is positive definite with eigenvalues (0.0010, 0.0289), i.e. the positivity of yields is ensured. Table C.2 summarizes the percentage errors for the respective moment conditions and yields used in estimation.

Table C.2: Fitting errors for the 2×2 model

The table presents percentage fitting errors for the moments of the 6-month, 2-year and 10-year yields. The percentage error is the difference between the model-implied and the empirical value of a given moment per unit of its empirical value. All errors are in percent.

	Maturities used		Error (%)	
Average y_t^τ	6M, 2Y, 10Y	-0.27	-0.59	0.86
Volatility y_t^τ	6M, 2Y, 10Y	0.35	-0.19	-0.78
Corr. y^{τ_1}, y^{τ_2}	(6M,2Y), (6M,10Y), (2Y,10Y)	-1.77	1.63	2.44
CS coeff.	n = 2Y, 10Y; m = 6M	-0.15	0.00	–

C.5.B The 3×3 Model

Our 3×3 framework is described by 18 parameters in M, Q and D matrices, plus the degrees of freedom parameter k. The estimation involves 24 moment conditions summarized in Table C.3. The optimization technique follows the same steps as for the 2×2 model. The final set of parameters has $k = 3$ degrees of freedom, and the following M, Q and D matrices:

$$D = \begin{pmatrix} 1.0396 & 0.0224 & -0.0094 \\ 0.0224 & 1.1412 & 0.0117 \\ -0.0094 & 0.0117 & 1.0043 \end{pmatrix},$$

$$M = \begin{pmatrix} -0.8506 & 0 & 0 \\ 0.2249 & -0.0787 & 0 \\ -2.2800 & -2.4125 & -0.9121 \end{pmatrix},$$

$$Q = \begin{pmatrix} 1.3276 & -0.2950 & 5.2410 \\ 0 & -0.0453 & 0.0667 \\ 0 & 0 & -0.6443 \end{pmatrix}.$$

Note from the diagonal elements, the M matrix is negative definite, and the Q matrix is invertible. The matrix $D - I_3$, however, is not positive definite with eigenvalues $(-0.0001, 0.0387, 0.1465)$. Even though there is a theoretical probability that yields become negative, the empirical frequency of such occurrences is zero across all maturities (based on the simulated sample of 72000 monthly observations). The violation of positive definiteness is thus inconsequential.

Table C.3: Fitting errors for the 3×3 model

The table presents percentage fitting errors for the moments of the 6-month, 2-year, 5-year and 10-year yields. The percentage error is the difference between the model-implied and the empirical value of a given moment per unit of its empirical value. All errors are in percent.

Moment	Maturities used	Error (%)			
Average y_t^τ	6M, 2Y, 5Y, 10Y	0.22	-0.40	0.16	-0.03
Volatility y_t^τ	6M, 2Y, 5Y, 10Y	0.85	-2.45	2.06	-0.23
Corr. y^{τ_1}, y^{τ_2}	(6M,2Y), (6M,5Y), (6M,10Y), (2Y,10Y)	-0.39	0.17	0.16	-0.23
Corr. $\Delta y^{\tau_1}, \Delta y^{\tau_2}$	(6M,2Y), (6M,10Y), (2Y,10Y)	-0.01	-0.03	0.02	–
CS coeff.	$n =$ 2Y, 10Y; $m =$ 6M	-0.00	-0.00	–	–
Forward rate volatility	6M→2Y, 2Y→5Y, 5Y→10Y	-3.38	4.99	-2.15	–

Bibliography

Adrian, T., and H. Wu (2009): "The Term Structure of Inflation Expectations," Working Paper, Federal Reserve Bank of New York.

Ahn, D.-H., R. F. Dittmar, and A. R. Gallant (2002): "Quadratic Term Structure Models: Theory and Evidence," *Review of Financial Studies*, 15, 243–288.

Amin, K. I., and A. J. Morton (1994): "Implied Volatility Functions in Arbitrage-Free Term Structure Models," *Journal of Financial Economics*, 35, 141–180.

Andersen, T. G., and L. Benzoni (2008): "Do Bonds Span Volatility Risk in the US Treasury Market? A Specification Test for Affine Term Structure Models," FRB of Chicago Working Paper No. 2006-15.

——— (2010): "Do Bonds Span Volatility Risk in the US Treasury Market? A Specification Test for Affine Term Structure Models," *Journal of Finance*, 65, 603–653.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2007): "Real-Time Price Discovery in Global Stock, Bond and Foreign Exchange Markets," *Journal of International Economics*, 73, 251–277.

Ang, A., and G. Bekaert (2007): "Stock Return Predictability: Is It There?," *Review of Financial Studies*, 20, 651–707.

Ang, A., J. Boivin, S. Dong, and R. Loo-Kung (2009): "Monetary Policy Shifts and the Term Structure," Working paper, Columbia University, HEC Montréal, Ortus Capital Management.

Ang, A., S. Dong, and M. Piazzesi (2007): "No-Arbitrage Taylor Rules," Working paper, Columbia University, University of Chicago, NBER and CEPR.

Ang, A., and M. Piazzesi (2003): "A No-Arbitrage Vector Autoregression of Term Structure with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745–787.

Atkeson, A., and P. J. Kehoe (2008): "On the Need for a New Approach to Analyzing Monetary Policy," *NBER Macroeconomics Annual*, forthcoming.

Audrino, F., and F. Corsi (2007): "Realized Correlation Tick-by-Tick," Working paper, Univerisity of St. Gallen.

——— (2008): "Realized Covariance Tick-by-Tick in Presence of Rounded Time Stamps and General Microstructure Effects," *Discussion paper no. 2008-04*.

Backus, D. K., and J. H. Wright (2007): "Cracking the Conundrum," Working Paper, Stern School of Business and Board of Governors of the Federal Reserve System.

Bandi, F., and J. Russell (2008): "Microstructure Noise, Realized Variance, and Optimal Sampling," *Review of Economic Studies*, 75, 339–369.

Bansal, R., G. Tauchen, and H. Zhou (2003): "Regime-Shifts, Risk Premiums in the Term Structure, and the Business Cycle," Duke University and Federal Reserve Board.

Bansal, R., and H. Zhou (2002): "Term Structure of Interest Rates with Regime Shifts," *Journal of Finance*, 57, 1997–2038.

Barndorff-Nielsen, O. E., and N. Shephard (2004): "Econometric Analysis of Realised Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics," *Econometrica*, 72, 885–925.

Bean, C., M. Paustian, A. Penalver, and T. Taylor (2010): "Monetary Policy after the Fall," Bank of England, Federal Reserve Bank of Kansas City Annual Conference, Jackson Hole, Wyoming.

Bekker, P. A., and K. E. Bouwman (2009): "Risk-free Interest Rates Driven by Capital Market Returns," Working Paper, University of Groningen and Erasmus University Rotterdam.

Bikbov, R., and M. Chernov (2009): "Unspanned Stochastic Volatility in Affine Models: Evidence from Eurodollar Futures and Options," Management Science, *forthcoming*.

Black, F. (1976): "The Pricing of Commodity Contracts," *Journal of Financial Economics*, 3, 167–179.

Branch, W., and G. W. Evans (2006): "A Simple Recursive Forecasting Model," *Economic Letters*, 91, 158–166.

Brandt, M. W., and D. A. Chapman (2002): "Comparing Multifactor Models of the Term Structure," Working paper, University of Pennsylvania and University of Texas.

Brigo, D., and F. Mercurio (2006): *Interest Rate Models: Theory and Practice*. Springer, Berlin, Heidelberg.

Brockett, R. (1970): *Finite Dimensional Linear Systems*. Wiley, New York.

Bru, M.-F. (1991): "Wishart Processes," *Journal of Theoretical Probability*, 4, 725–751.

Buraschi, A., A. Cieslak, and F. Trojani (2008): "Correlation Risk and the Term Structure of Interest Rates," Working Paper, University of Lugano.

Buraschi, A., and A. Jiltsov (2006): "Habit Formation and Macroeconomic Models of the Term Structure of Interest Rates," *Journal of Finance*, 62, 3009–3063.

Buraschi, A., P. Porchia, and F. Trojani (2009): "Correlation Risk and Optimal Portfolio Choice," Journal of Finance, *forthcoming*.

Campbell, J. (1995): "Some Lessons from the Yield Curve," *Journal of Economic Perspectives*, 9, 129–152.

Campbell, J. Y., and P. Perron (1991): "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," *NBER Macroeconomics Annual*, 6, 141–201.

Campbell, J. Y., and R. J. Shiller (1991): "Yield Spreads and Interest Rate Movements: A Bird's Eye View," *Review of Economic Studies*, 58, 495–514.

Campbell, J. Y., A. Sunderam, and L. M. Viceira (2010): "Inflation Bets or Deflation Hedges? The Changing Risk of Nominal Bonds," Working paper, Harvard Business School.

Campbell, J. Y., and S. Thompson (2008): "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?," *Review of Financial Studies*, 21, 1509–1531.

Carceles-Poveda, E., and C. Giannitsarou (2007): "Adaptive Learning in Practice," *Journal of Economic Dynamics and Control*, 31, 2659–2697.

Carlson, J. A. (1977): "A Study of Price Forecasts," *Annals of Economic and Social Measurement, NBER*, 6, 33–63.

Carr, P., and L. Wu (2007): "Stochastic Skew in Currency Options," *Journal of Financial Economics*, 86, 213–247.

Chacko, G., and S. Das (2002): "Pricing Interest Rate Derivatives: A General Approach," *Review of Financial Studies*, 15, 195–241.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1227.

Cheridito, P., D. Filipovic, and R. L. Kimmel (2007): "Market Price of Risk Specifications for Affine Models: Theory and Evidence," *Journal of Financial Economics*, 83, 123–170.

Christoffersen, P., K. Jacobs, L. Karoui, and K. Mimouni (2009): "Non-Linear Filtering in Affine Term Structure Models: Evidence from the Term Structure of Swap Rates," Working paper, McGill University, Goldman Sachs, American University in Dubai.

Clarida, R., J. Galí, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147–180.

Clarida, R. H. (2010): "What Has—and Has Not—Been Learned About Monetary Policy in a Low Inflation Environment? A Review of the 2000s," Speech Delivered to the Boston Federal Reserve Bank Conference.

Clark, T., and M. McCracken (2001): "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85–110.

——— (2005): "Evaluating Direct Multi-Step Forecasts," *Econometric Reviews*, 24, 369–404.

Cochrane, J. H., and M. Piazzesi (2005): "Bond Risk Premia," *American Economic Review*, 95, 138–160.

——— (2008): "Decomposing the Yield Curve," Working paper, University of Chicago.

Collin-Dufresne, P., and R. S. Goldstein (2001): "Stochastic Correlations and the Relative Pricing of Caps and Swaptions in a Generalized-Affine Framework," Working Paper, Carnegie Mellon University and Washington University.

——— (2002): "Do Bonds Span the Fixed Income Markets? Theory and Evidence for the Unspanned Stochastic Volatility," *Journal of Finance*, 58(4), 1685–1730.

Collin-Dufresne, P., R. S. Goldstein, and C. S. Jones (2006): "Can Interest Rate Volatility Be Extracted from the Cross Section of Bond Yields? An Investigation of Unspanned Stochastic Volatility," Working Paper, University of California Berkeley, University of Minnesota, and University of Southern California.

——— (2009): "Can Interest Rate Volatility Be Extracted from the Cross-Section of Bond Yields?," *Journal of Financial Economics*, 94, 47–66.

Cooper, I., and R. Priestley (2009): "Time-Varying Risk Premiums and the Output Gap," *Review of Financial Studies*, 22, 2801–2833.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985a): "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, 53, 363–384.

——— (1985b): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 373–384.

da Fonseca, J., M. Grasselli, and C. Tebaldi (2006): "Option Pricing when Correlations Are Stochastic: An Analytical Framework," Working Paper, esliv, University of Padova and University of Verona.

Dai, Q. (2003): "Term Structure Dynamics in a Model with Stochastic Internal Habit," Working Paper, New York University.

Dai, Q., and K. Singleton (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943–1978.

——— (2002): "Expectation Puzzles, Time-Varying Risk Premia, and Affine Models of the Term Structure," *Journal of Financial Economics*, 63, 415–441.

——— (2003): "Term Structure Dynamics in Theory and Reality," *Review of Financial Studies*, 16, 631–678.

Dai, Q., K. J. Singleton, and W. Yang (2004): "Predictability of Bond Risk Premia and Affine Term Structure Models," Working Paper, New York University and Stanford University.

David, A., and P. Veronesi (2009): "What Ties Return Volatilities to Price Valuations and Fundamentals?," Working paper, University of Calgary and University of Chciago.

De Jong, F., J. Driessen, and A. Pessler (2004): "On the Information in the Interest Rate Term Structure and Option Prices," *Review of Derivatives Research*, 7, 99–127.

Dewachter, H., and L. Iania (2010): "An Extended Macro-Finance Model with Financial Factors," Working paper, Katholieke Universiteit Leuven.

Dewachter, H., and M. Lyrio (2006): "Learning, Macroeconomic Dynamics and the Term Structure of Interest Rates," Working paper, Katholieke Universiteit Leuven, Erasmus University of Rotterdam.

Donati-Martin, C., Y. Doumerc, H. Matsumoto, and M. Yor (2004): "Some Properties of the Wishart Process and a Matrix Extension of the Hartman-Watson Laws," *Publications of the Research Institute for Mathematical Sciences*, 40, 1385–1412, Working Paper, University of P. and M. Curie, University of P. Sabatier, and Nagoya University.

Drost, F. C., and T. E. Nijman (1993): "Temporal Aggregation of Garch Processes," *Econometrica*, 61, 909–927.

Duffee, G. R. (2002): "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, 405–443.

——— (2007): "Are Variations in Term Premia Related to the Macroeconomy?," Working paper, University of California – Berkeley.

——— (2011): "Information in (and Not in) the Term Structure," *Review of Financial Studies*, forthcoming.

Duffie, D., D. Filipovic, and W. Schachermayer (2003): "Affine Processes and Applications in Finance," *Annals of Applied Probability*, 13, 984–1053.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Duffie, D., and K. Singleton (1997): "An Econometric Model of the Term Structure of Interest-Rate Swap Yields," *Journal of Finance*, 52, 1287–1321.

Engle, R. (2002): "Dynamic Conditional Correlation – A Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R., and C. W. Granger (1987): "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

Engle, R. F., and K. Sheppard (2001): "Theoretical and Emprical Properties of Dynamic Conditional Correlation Multivariate GARCH," NBER Working Paper No. 8554.

Evans, G. W., and S. Honkapohja (2009): "Learning and Macroeconomics," *Annual Review of Economics*, 1, 421–449.

Evans, G. W., S. Honkapohja, and N. Williams (2010): "Generalized Stochastic Gradient Learning," *International Economic Review*, 51, 237–262.

Fama, E. (2006): "The Behavior of Interest Rates," *Review of Financial Studies*, 19, 359–379.

Fama, E. F., and R. R. Bliss (1987): "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

Feldhütter, P. (2007): "Can Affine Models Match the Moments in Bond Yields?," Working Paper, Copenhagen Business School.

Fengler, M. R., W. K. Härdle, and C. Villa (2001): "The Dynamics of Implied Volatilities: A Common Principal Components Approach," Working Paper, Humboldt-Universität zu Berlin, University of Rennes.

Fisher, M., and C. Gilles (1996a): "Estimating Exponential-Affine Models of the Term Structure," Working Paper, Federal Reserve Bank of Atlanta.

Fisher, M., and C. Gilles (1996b): "Term Premia in Exponential-Affine Models of the Term Structure," Working Paper, Federal Reserve Bank of Atlanta.

Fisher, M., D. Nychka, and D. Zervos (1994): "Fitting the term structure of interest rates with smoothing splines," Working paper, Federal Reserve and North Carolina State University.

Fleming, M. (2003): "Measuring Treasury Market Liquidity," *FRBNY Economic Policy Review*, September, 84–108.

Fleming, M. J. (1997): "The Round-the-Clock Market for U.S. Treasury Securities," *FRBNY Economic Policy Review*.

Fleming, M. J., and B. Mizrach (2008): "The Microstructure of a U.S. Treasury ECN: The BrokerTec Platform," Working paper, Federal Reserve Bank of New York and Rutgers University.

Fontaine, J.-S., and R. Garcia (2010): "Bond Liquidity Premia," Working paper, University of Montreal, CIREQ and EDHEC Business School.

Gagliardini, P., P. Porchia, and F. Trojani (2007): "Ambiguity Aversion and the Term Structure of Interest Rates," *Review of Financial Studies*, *forthcoming*.

Goodfriend, M., and R. G. King (2009): "The Great Inflation Drift," NBER working paper.

Gourieroux, C. (2006): "Continuous Time Wishart Process for Stochastic Risk," *Econometric Reviews*, 25, 177–217.

Gourieroux, C., J. Jasiak, and R. Sufana (2009): "The Wishart Autoregressive Process of Multivariate Stochastic Volatility," *Journal of Econometrics*, 150, 167–181.

Gourieroux, C., and R. Sufana (2003): "Wshart Quadratic Term Structure Models," Working Paper, CREST, CEPREMAP, and University of Toronto.

——— (2004): "Derivative Pricing with Wishart Multivariate Stochastic Volatility: Application to Credit Risk," Working paper, CREST, CEPREMAP, and University of Toronto.

Goyal, A., and I. Welch (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

Greenspan, A. (2009): "The Fed Didn't Cause the Housing Bubble," *Wall Street Journal*, March 11, Eastern Edition, A15.

Gürkaynak, R. S., B. Sack, and E. Swanson (2005): "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models," *American Economic Review*, 95, p. 425 – 436.

Gürkaynak, R. S., B. Sack, and J. H. Wright (2006): "The U.S. Treasury Yield Curve: 1961 to the Present," Working paper, Federal Reserve Board.

Han, B. (2007): "Stochastic Volatilities and Correlations of Bond Yields," *Journal of Finance*, 62, 1491–1524.

Hatzius, J., P. Hooper, F. Mishkin, K. Schoenholtz, and M. Watson (2010): "Financial Conditions Indexes: A Fresh Look after the Financial Crisis," Working paper, Goldman Sachs, Deutsche Bank, Columbia University, New York University and Princeton University.

Haubrich, J., G. Pennacchi, and P. Ritchken (2008): "Estimating Real and Nominal Term Structures using Treasury Yields, Inflation, Inflation Forecasts, and Inflation Swap Rates," Working Paper, Federal Reserve Bank of Cleveland.

Hautsch, N., and Y. Ou (2008): "Yield Curve Factors, Term Structure Volatility, and Bond Risk Premia," Working paper, Humboldt Univeristy Berlin.

Hayashi, T., and N. Yoshida (2005): "On covariance estimation of non-synchronously observed diffusion processes," *Bernoulli*, 11, 359–379.

Heidari, M., and L. Wu (2003): "Are Interest Rate Derivatives Spanned by the Term Structure of Interest Rates?," *Journal of Fixed Income*, 13, 75–86.

Heston, S. (1993): "A Closed-Form Solution for Options with Stochastic Volatility and Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327–343.

Hodrick, R. J. (1992): "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 357–386.

Hu, X., J. Pan, and J. Wang (2010): "Noise as Information for Illiquidity," Working paper, MIT Sloan.

Huang, J.-Y., and Z. Shi (2010): "Determinants of Bond Risk Premia," Working paper, Penn State University.

Jacobs, K., and L. Karoui (2006): "Affine Term Structure Models, Volatility and the Segmentation Hypothesis," Working Paper, McGill University.

Jacobs, K., and L. Karoui (2009): "Conditional Volatility in Affine Term-Structure Models :Evidence from Treasury and Swap Markets," *Journal of Financial Economics*, 91, 288–318.

Jacod, J. (1994): "Limit of Random Measures Associated with the Increments of a Brownian Semimartingale," Working Paper, Universite Pierre et Marie Curie.

Jardet, C., A. Monfort, and F. Pegoraro (2010): "No-Arbitrage Near-Cointegrated VAR(p) Term Structure Models, Term Premia and GDP Growth," Working paper, Banque de France, CNAM, CREST.

Jensen, G. R., and T. Moorman (2010): "Inter-temporal variation in the illiquidity premium," *Journal of Financial Economics*, 98, 338–358.

Jones, C. M., O. Lamont, and R. L. Lumsdaine (1998): "Macroeconomic News and Bond Market Volatility," *Journal of Financial Economics*, 47, 315–337.

Joslin, S. (2006): "Can Unspanned Stochastic Volatility Models Explain the Cross Section of Bond Volatilities?," Working paper, MIT Sloan School of Management.

——— (2007): "Pricing and Hedging Volatility Risk in Fixed Income Markets," Working Paper, MIT Sloan School of Management.

Joslin, S., M. Priebsch, and K. Singleton (2010): "Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks," Working paper, MIT Sloan School of Management and Stanford University.

Joslin, S., K. J. Singleton, and H. Zhu (2011): "A New Perspective on Gaussian Dynamic Term Structure Models," *Review of Financial Studies*, forthcoming.

Jotikasthira, C., A. Le, and C. Lundblad (2010): "Why Do Term Structures in Different Currencies Comove?," Working paper, University of North Carolina at Chapel Hill.

Julier, S., and J. Uhlmann (1997): "A new extension of the Kalman filter to nonlinear systems," *in Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls*.

Kalman, R. (1960): "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, Vol. 82, 35–45.

Kim, D. H. (2007a): "Challenges in Macro-Finance Modelling," BIS Working Paper, No. 240.

——— (2007b): "Spanned Stochastic Volatility in Bond Markets: A Re-examination of the Relative Pricing between Bonds and Bond Options," BIS Working Paper.

Kim, D. H., and K. Singleton (2010): "Term Structure Models and the Zero Bound: An Empirical Investigation of Japanese Yields," Working paper, Stanford University and Yonsei University.

Koijen, R. S., O. Van Hemert, and S. Van Nieuwerburgh (2009): "Mortgage Timing," *Journal of Financial Economics*, 93, 292–324.

Kozicki, S., and P. Tinsley (1998): "Moving Endpoints and the Internal Consistency of Agents' Ex Ante Forecasts," *Computational Economics*, 11, 21–40.

——— (2001a): "Shifting Endpoints in Term Structure of Interest Rates," *Journal of Monetary Economic*, 47, 613–652.

——— (2001b): "Term Structure Views of Monetary Policy under Alternative Models of Agent Expectations," *Journal of Economic Dynamics & Control*, 25, 149–184.

——— (2005): "Permanent and Transitory Policy Shocks in an Empirical Macro Model with Asymmetric Information," *Journal of Economic Dynamics and Control*, 29, 1985–2015.

——— (2006): "Survey-Based Estimates of the Term Structure of Expected U.S. Inflation," Working paper, Bank of Canada.

Künsch, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

Laub, A. J. (2005): *Matrix Analysis for Scientists and Engineers*. SIAM, Davis, California.

Leippold, M., and L. Wu (2002): "Asset Pricing under the Quadratic Class," *Journal of Financial and Quantitative Analysis*, 37, 271–295.

——— (2003): "Estimation and Design of Quadratic Term Structure Models," *Review of Finance*, 7, 47–73.

Levin, J. (1959): "On the Matrix Riccati Equation," *Proceedings of the American Mathematical Society*, 10, 519–524.

Li, H., and F. Zhao (2006): "Unspanned Stochastic Volatility: Evidence from Hedging Interest Rate Derivatives," *Journal of Finance*, 61, 341–378.

Litterman, R., and J. Scheinkman (1991): "Common Factors Affecting Bond Returns," *Journal of Fixed Income*, 1, 54–61.

Longstaff, F. A., and E. S. Schwartz (1992): "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, 47, 1259–1282.

Ludvigson, S. C., and S. Ng (2009): "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22, 5027–5067.

Magnus, J. R., and H. Neudecker (1979): "The Commutation Matrix: Some Properties and Applications," *Annals of Statistics*, 7, 381–394.

——— (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics, Chichester.

Malmendier, U., and S. Nagel (2009): "Learning from Inflation Experiences," Working paper, UC Berkeley and Stanford University.

Mankiw, N. G. (2001): "U.S. Monetary Policy during the 1990s," NBER working paper.

Mishkin, F. S. (2007): "Housing and the Monetary Transmission Mechanism," Finance and Economics Discussion Series, Federal Reserve Board.

Mizrach, B., and C. J. Neely (2006): "The Transition to Electronic Communications Networks in the Secondary Treasury Market," *Federal Reserve Bank of St. Louis Review*.

Moench, E. (2008): "Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach," *Journal of Econometrics*, 146, 26–43.

Moraleda, J. M., and T. C. Vorst (1997): "Pricing American Interest Rate Claims with Humped Volatility Models," *Journal of Banking and Finance*, 21, 1131–1157.

Muirhead, R. J. (1982): *Aspects of Multivariate Statistical Theory*. Wiley Series in Probability and Mathematical Statistics.

Orphanides, A., and M. Wei (2010): "Evolving Macroeconomic Perceptions and the Term Structure of Interest Rates," Working paper, Board of Governors of the Federal Reserve System.

Pérignon, C., and C. Villa (2006): "Sources of Time Variation in the Covariance Matrix of Interest Rates," *Journal of Business*, 79, 1535–1549.

Phoa, W. (1997): "Can You Derive Market Volatility Forecast from the Observed Yield Curve Convexity Bias?," *Journal of Fixed Income*, pp. 43–53.

Piazzesi, M. (2001): "An Econometric Model of the Yield Curve with Macroeconomic Jump Effects," Working Paper, ucla and nber.

——— (2003): "Affine Term Structure Models," Working Paper, University of Chicago.

Piazzesi, M., and M. Schneider (2011): "Trend and Cycle in Bond Premia," Working paper, Stanford University and NBER.

Price, K. V., R. M. Storn, and J. A. Lampinen (2005): *Differential Evolution: A Practical Approach to Global Optimization*. Springer Berlin.

Roma, A., and W. Torous (1997): "The Cyclical Behavior of Interest Rates," *Journal of Finance*, 52, No. 4, 1519–1542.

Rudebusch, G. D., E. T. Swanson, and T. Wu (2006): "The Bond Yield "Conundrum" from a Macro Finance Perspective," Working Paper, Federal Reserve Banks of San Franciso and Dallas.

Rudebusch, G. D., and T. Wu (2008): "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy," *The Economic Journal*, 118, 906–926.

Singleton, K. J. (2006): *Empirical Dynamic Asset Pricing*. Princeton University Press, Princeton and Oxford.

Stambaugh, R. F. (1988): "The Information in Forward Rates: Implications for Models of the Term Structure," *Journal of Financial Economics*, 21, 41–70.

Stock, J. H. (1987): "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55, 1035–1056.

Stock, J. H., and M. W. Watson (2007): "Why Has U.S. Inflation Become Harder to Forecast?," *Journal of Money, Credit, and Banking*, 39, 3–33.

Thompson, S. (2008): "Identifying Term Structure Volatility from the LIBOR-Swap Curve," *Review of Financial Studies*, 21, 819–854.

Trolle, A. B., and E. S. Schwartz (2009): "A General Stochastic Volatility Model for the Pricing of Interest Rate Derivatives," *Review of Financial Studies*, 22, 2007–2057.

Van Loan, C. F. (1978): "Computing Integrals Involving Matrix Exponential," *IEEE Transactions on Automatic Control*, 23, 395–404.

Vasicek, O. A. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Waggoner, D. F. (1997): "Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices," *FRB of Atlanta working paper*, 97-10.

Wan, E. A., and R. van der Merwe (2001): *Kalman Filtering and Neural Networks*. John Wiley & Sons, Inc.

Wei, M., and J. H. Wright (2010): "Reverse Regressions and Long-Horizon Forecasting," Working paper, Federal Reserve Board and Johns Hopkins University.

Wright, J., and H. Zhou (2009): "Bond Risk Premia and Realized Jump Risk," *Journal of Banking and Finance*, 33, 2333–2345.

Wright, J. H. (2009): "Term Premia and Inflation Uncertainty: Empirical Evidence from an International Panel Dataset," *American Economic Review*, forthcoming.

Zhang, L., P. A. Mykland, and Y. Ait-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.