Measuring and Modelling Realized Volatility: from Tick-by-tick to Long Memory

Fulvio Corsi

Submitted for the degree of Ph.D. in Economics at

Faculty of Economics
University of Lugano
Lugano, Switzerland

Thesis Committee:
Prof. G. Barone-Adesi, advisor, University of Lugano
Prof. T. Bollerlsev, Duke University
Prof. F. Audrino, University of Lugano

7th June 2005

ACKNOWLEDGEMENTS

I would like to gratefully acknowledges Tim Bollerslev, Giovanni Barone-Adesi, Giuseppe Curci, Michel Dacorogna, Ulrich Müller, Gilles Zumbach, Loriano Mancini, Alessandro Palandri and Uta Kretschmer for insightful discussions.

Contents

Introduction

With transaction data becoming more widely available, considerable interest has recently been devoted to the use of high frequency data for measuring and forecasting volatility: the so called *realized volatility approach*. The thesis concerns the developments of new realized volatility measures and models with the aim to better understand and forecast the dynamics of market volatility.

Untill, recently the prevailing approach employed (at most) daily data and considered volatility as an unobservable variable. Modelling the unobserved conditional variance led to the vast ARCH-GARCH and stochastic volatility literature. This work focus instead on a nonparametric volatility approach which fully exploits intraday information to develop observable proxies for the ex-post volatility: the realized volatility measure.

In its standard form realized volatility is nothing more than the square root of the sum of squared high-frequency returns over a given non-vanishing time interval, i.e. the second uncentered sample moment of the high-frequency returns. Under very general conditions the sum of intraday squared returns converges (as the number of intraday return increases) to the relevant notion of volatility of the interval. Thus, realized volatility provides us, in principle, with a consistent nonparametric measure of volatility.

A model-free and error-free estimation of volatility would allow us to treat volatility as an observable variable, rather than a latent one, as in the GARCH(1,1) model for example. This would open the possibility to directly analyze, model, and forecast volatility itself. Therefore, more sophisticated dynamic models can be directly estimated and optimized without having to rely on the complicated estimation procedures needed when volatility is assumed to be unobserved. For forecasting purposes, moreover, a better estimate of the target function allows to better extract the real underlying signal and then improve the forecasting performance.

Unfortunately, because of market microstructures effects, the assumption that log asset prices evolve as a diffusion process becomes less realistic as the time scale reduces. At the tick time scale, the empirical data differ from the frictionless continuous-time price process assumed in the standard theory of realized volatility. Thus, simple realized volatility measures computed with very short time intervals are no longer unbiased and consistent estimators of daily volatilities

The first purpose of this thesis is to develop new realized volatility estimators which, while fully exploiting all the available information contained in very high frequency data, are able to effectively correct for the bias induced by microstructure effects. The second purpose is to develop, by building on such highly accurate realized volatility measures, new conditional volatility models able to provide superior and easy-to-implement volatility forecasts. The thesis consists of four main chapters.

In the first chapter the definition of realized volatility is first introduced. Then the impact of the presence of different sources of microstructure effects on realized volatility estimation is explained and quantified.

The second chapter reviews and proposes realized volatility measures which are able to deal with the challenge posed by the presence of microstructure effects on the basis of a simple structural model for the tick-by-tick price process that seems to well accommodates the empirical properties of the data. First, a simple filter, which attempt to remove the cause of the bias from the raw tick-by-tick time series, is presented. This filter consists in an adaptive exponential moving average (EMA), calibrated on the autocorrelation structure of past tick-by-tick returns. The result is an effective reduction of the realized volatility bias for FX data, particularly for the most liquid currencies. However, this type of EMA filter, is a non-local estimator which adapts only slowly to changes in the properties of the pricing error component. Moreover, the EMA filter corrects only for the bias deriving from the first lag of the return autocorrelation function, while it is sensitive to significant higher lags coefficients. When, as it often happen on real data, a significant autocorrelation at lags length greater than one is present and the autocorrelation structures dynamically change over time, the filtering problem become much more complex.

Then, new realized volatility measures based on Multi-Scale regression and *Discrete Sine Transform* (DST) approaches are presented in order to deal with these more challenging cases. We show that Multi-Scales estimators similar to that recently proposed by Zhang (2004) can be constructed within a simple regression based approach by exploiting the linear relation existing between the market microstructure bias and the realized volatilities computed at different frequencies. These regression based estimators can be further improved and robustified by using the DST approach to prefilter market microstructure noise. The motivation for the DST approach rests on its ability to diagonalize MA type of processes which arise naturally in discrete time models of tick-by-tick returns with market microstructure noise. Hence, the

DST provides a natural orthonormal basis decomposition of observed returns which permits to optimally disentangle the volatility signal of the underlying price process from the market microstructure noise. Robustness of the DST approach with respect to more general dependent structure of the microstructure noise is also analytically shown. Then, the combination of such Multi-Scale regression approach with the DST gives us a Multi-Scales DST realized volatility estimator which is robust against a wide class of noise contaminations and model misspecifications. Thanks to the DST orthogonalization which also allows us to analytically derive closed form expressions for the Cramer-Rao bounds of MA(1) processes, an evaluation of the absolute efficiency of volatility estimators under the i.i.d. noise assumption becomes available, indicating that the Multi-Scales DST estimator possesses a finite sample variance very close to the optimal Cramer-Rao bounds.

Monte Carlo simulations based on a realistic model for microstructure effects and volatility dynamics, show the superiority of DST estimators compared to alternative local volatility proxies. DST estimators result to be the most accurate daily volatility measures for every level of the noise to signal ratio, and highly robust against the presence of significant autocorrelation at lags greater than one in the return process. The empirical analysis based on six years of tick-by-tick data for S&P 500 index-future, FIB 30, and 30 years U.S. Tresaury Bond future seems to confirm Monte Carlo results.

The third chapter deals with realized volatility modelling. We employ the realized volatility measures computed from tick-by-tick data to directly analyze, model and forecast the time series behavior of volatility. The purpose is to reproduce the long memory property of volatility with a simple and parsimonious realized volatility model. Inspired by the Heterogeneous Market Hypothesis (Müller et al. 1993) and by the asymmetric propagation of volatility between long and short time horizons, we propose an additive cascade of different volatility components generated by the actions of different types of market participants. This additive volatility cascade leads to a simple AR-type model in the realized volatility with the feature of considering volatilities realized over different time horizons. We term this model, Heterogeneous Autoregressive model of Realized Volatility (HAR-RV). In spite of the simplicity of its structure, simulation results, seem to confirm that the HAR-RV model successfully achieves the purpose of reproducing the main empirical features of volatility (long memory, fat tail, self-similarity) in a very simple and parsimoniously way. Results on the estimation and forecast of the HAR-RV model on USD/CHF data, show remarkably good out of sample forecasting performance which seems to steadily and substantially outperform those of standard models. Moreover, by extending the Heterogeneous Market Hypothesis idea to the leverage effect, we propose an asymmetric version of the HAR-RV model which considers asymmetric responses of the realized volatility not only to previous daily returns but also to past weekly and monthly returns.

In the final chapter we extend the approach of directly using all the available tick-by-tick data to the realized covariance and realized correlation estimation. As for the realized volatility, the presence of market microstructure (although of a different nature) can induce significant bias in standard realized covariance measures computed with artificially regularly spaced returns. Contrary to these standard approaches, we propose a simple and unbiased realized covariance estimator which does not resort on the construction of a regular grid, but directly and efficiently employs the raw tick-by-tick returns of the two series. Montecarlo simulations show that this simple tick-by-tick covariance estimator posses no bias and the smallest dispersion, resulting to be the best performing measure among the covariance estimators considered in the study. Combining the proposed covariance together with the DST volatility estimator we obtain a realized correlation measure where both the volatilities and the covariances are computed from tick-by-tick data. In the empirical analysis performed on S&P 500 and US bond data we apply the HAR model to the tick-by-tick correlation measure, obtaining highly significant coefficients for the heterogeneous components and remarkably good out of sample forecasting performance. We then suggest the use of a shrinkage approach with a newly proposed shrinkage target for the construction of a definite positive and more accurate correlation matrix. Finally, possible multivariate extensions of the HAR volatilty model are briefly outlined.

Chapter 1

Realized Volatility and Microstructure effects

1.1 Introduction

Asset returns volatility is a central feature of many prominent financial problems such as asset allocation, risk management and option pricing. For instance, in risk assessment, due to the increasing role played by the Value-at-Risk (VaR) approach, it is becoming more important to have a good measure and forecast of short-term volatility, mainly at the one to ten day horizon. Intuitively, the volatility measures how much a random process jitters. But despite its importance, volatility is still an ambiguous term for which there is no unique, universally accepted definition, and several volatility concept along this simple idea can be written down.

In a frictionless continuous-time no arbitrage price process framework, three different conditional volatility concepts can be defined[1]:

(i) the *Notional, Actual* or *Integrated* ex-post volatility over a non-vanishing interval,

(ii) the ex-ante *Expected* volatility over a non-vanishing interval and

(iii) the *Instantaneous* volatility.

The notional volatility refers to the ex-post cumulative sample-path return variability over a discrete time interval which under very general conditions corresponds to the increments in the quadratic variation of the return process.

[1]See Andersen, Bollerslev and Diebold (2005).

In practice, the approaches for empirically quantifying the concept of volatility fall into two distinct types of categories:

- estimation of *Parametric* models,
- direct *Nonparametric* methods.

So far, most of the studies have focused on the parametric approach considering volatility as an unobservable variable and using a fully specified functional model for the ex-ante expected volatility[2]. Modelling the unobserved conditional variance was one of the most prolific topics in the financial literature which led to all ARCH-GARCH models and stochastic volatility models. In the ARCH models only the most recent conditional volatility is not observable, while in the stochastic volatility ones, the whole history of the conditional variance is assumed unobservable. In general this kind of models suffer from a twofold weakness: first, they are not able to replicate main empirical features of financial data; second, the estimation procedures required are often rather complex (especially in the case of stochastic volatility models). This study focus instead on a nonparametric approach to develop ex-post observable proxies for the notional volatility (rather than the expected one) through new methodologies which fully exploits intraday information.

Within the class of nonparametric methods we can distinguish between the *Realized volatility Filter or Smoothers* and the *Realized Volatility Measures*. The first heavily rely on the continuous sample paths assumption on the price process in order to evaluate the instantaneous volatility. Filters exploit only the information contained in past returns while smoothers also use ex-post future returns (thus they can be seen as two-sided filters). These instantaneous volatility measures require that as the length of the time interval goes to zero the number of observations tends to infinity. However this strong condition (which implies a double limit theory and excludes jumps from both return and volatility process) are virtually never fulfilled in empirical data, making this approach unfeasible in practice.

1.2 Realized Volatility

Contrary to the ARCH filters, realized volatility affords the empirical measurement of the latent notional volatility on the discrete time interval $[t-h,t]$, with h a strictly positive non-vanishing quantity (typically one day). Similarly to the instantaneous volatility measures,

[2]Also the Implied Volatility approaches can be included in this category since they are based on a parametric model for the returns together with an option pricing model.

realized volatilities may be classified according to whether the estimation of the notional volatility only exploits returns observations falling in the interval $[t-h,t]$, which we call *Local*, or also incorporates returns outside $[t-h,t]$. Local measurements have the advantage to be asymptotically unbiased and fast adapting but the disadvantage to neglect potentially useful information contained in adjacent intervals. The most obvious local measure for daily volatility is the daily absolute return. However, as clearly shown by Andersen and Bollerslev (1998) this proxy can be extremely noisy. The inadequacy of volatility proxies obtained with daily observations clearly suggests the use of intraday data to obtain more accurate volatility estimates. In fact, in its standard form realized volatility is nothing more than the square root of the sum of squared high-frequency returns over a given time interval $[t-h,t]$, i.e. the square root of the second uncentered sample moment of the high-frequency returns.

This idea traces back to the seminal work of Merton (1980) who showed that the integrated variance of a Brownian motion can be approximated to an arbitrary precision using the sum of intraday squared returns. His intuition was that higher frequencies are not useful for the mean but essential for the variance. This is deeply rooted in that, for a random walk, a minimal exhaustive statistics for the mean is given by the start and end point of the walk, whereas a minimal exhaustive statistics for the volatility is essentially given by the full set of increments.

Yet only recently this idea has started to be exploited with intraday data. Earlier studies are: Taylor and Xu (1997) who rely on 5-minute returns in the measurement of daily exchange rate volatilities, Schwert (1998) which utilizes 15-minute returns to estimate daily stock market volatilities, while Dacorogna, Müller, Dav, Olsen and Pictet (1998) use 1-hour returns to compute a 1-day volatility benchmark for other volatility forecasting models.

More recently a series of breakthrough papers (Andersen, Bollerslev, Diebold and Labys 2001, 2003 and Barndorff-Nielsen and Shephard 2001, 2002a 2002b, 2005 and Comte and Renault 1998) has formalized and generalized this important intuition by applying the quadratic variation theory to the broad class of special (finite mean) semimartingales. This very general class encompasses processes used in standard arbitrage-free asset pricing applications, such as Ito diffusions, jump processes, and mixed jump diffusions. They showed that, for this vast class of processes and under very general conditions, the sum of intraday squared returns converges (as the maximal length of returns go to zero) to the notional volatility over the fixed time interval $[t-h,t]$. Thus, as the sampling frequency from a diffusion is increased, realized volatility provides us, in principle, with a consistent nonparametric measure of the notional volatility.

To illustrate this property more formally and establish the notation of these different concepts of volatility, lets consider the following stochastic volatility process[3]

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) \tag{1.1}$$

where $p(t)$ is the logarithm of instantaneous price, $\mu(t)$ is a continuos, finite variation process, $dW(t)$ is a standard Brownian motion, and $\sigma(t)$ is a stochastic process independent of $dW(t)$. For this diffusion process, the integrated variance associated with day t, is the integral of the instantaneous variance over the one day interval $[t-1d, t]$, where a full 24 hours day is represented by the time interval $1d$,

$$IV_t^{(d)} = \int_{t-1d}^{t} \sigma^2(\omega)d\omega. \tag{1.2}$$

Some authors refer to this quantity as integrated volatility, while we will devote this term to the square root of the integrated variance, i.e. in our notation, the integrated volatility is $\sigma_t^{(d)} = (IV_t^{(d)})^{1/2}$.

As shown by Andersen, Bollerslev, Diebold and Labys 2001a,b and Barndorff-Nielsen and Shephard 2001a,b 2002a,b, the integrated variance $IV_t^{(d)}$ can be approximated to an arbitrary precision using the sum of intraday squared returns. Hence, the standard definition (for an equally spaced returns series) of the realized volatility over a time interval of one day is

$$RV_t^{(d)} = \sqrt{\sum_{j=0}^{M-1} r_{t-j\cdot\Delta}^2} \tag{1.3}$$

where $\Delta = 1d/M$ and $r_{t-j\cdot\Delta} = p\,(t-j\cdot\Delta) - p\,(t-(j+1)\cdot\Delta)$ defines continuously compounded Δ-frequency returns, that is, intraday returns sampled at time interval Δ (here, the subscript t indexes the day while j indexes the time within the day t).

In the following, we will also consider latent integrated volatility and realized volatility viewed over different time horizons longer than one day. These multi-period volatilities will simply be normalized sums of the one-period realized volatilities (i.e. a simple average of the daily quantities). For example, in our notation, a weekly realized volatility at time t will be given by the average

$$RV_t^{(w)} = \sqrt{\frac{1}{5}\left((RV_t^{(d)})^2 + (RV_{t-1d}^{(d)})^2 + ... + (RV_{t-4d}^{(d)})^2\right)}. \tag{1.4}$$

In particular we will make use of weekly and monthly aggregation periods. Indicating the aggregation period as an upper script, the notation for weekly integrated and realized volatility

[3] We use (t) to denote instantaneous variables and subscripts t to denote discrete quantities.

will be respectively $\sigma_t^{(w)}$ and $RV_t^{(w)}$ while a monthly aggregation will be denoted as $\sigma_t^{(m)}$ and $RV_t^{(m)}$. In the following, in respect of their actual frequency, all return and volatility quantities are intended to be annualized.

Under very broad assumptions, the ex-post realized volatility is an unbiased volatility estimator. Moreover, as the sampling frequency from a diffusion (even with non zero mean process) is increased, realized volatility provides a consistent nonparametric measure of the integrated volatility over the fixed time interval[4]: $\text{plim}_{M\to\infty} RV_t^{(d)} = \sigma_t^{(d)}$. This convergence property of the volatility is very appealing, in theory: under those assumptions, the mesaurement error of the integrated volatility could be arbitrarily reduced by simply increasing the sampling frequency of returns.

An error-free estimation of volatility would allow us to treat realized volatility as an observable, rather than a latent variable as with a GARCH(1,1) model for example. This opens the possibility to directly analyze, model, forecast and optimize volatility itself. Hence, more sophisticated dynamic models can be directly estimated without having to rely on the complicated estimation procedures needed when the volatility is assumed to be unobserved (e.g. log-likelihood for the ARCH-type models or indirect methods for SV models). For forecasting purposes, a better estimate of the target function allows to better extract the real underlying signal and to improve forecasting performance (Andersen and Bollerslev 1998). There is, in fact, a fast growing literature on directly analysing, modelling and forecasting observed realized volatility time series. Some examples are: Andersen, Bollerslev, Diebold and Labys (2003), Oomen (2002), Maheu and McCurdy (2002), Thomakos and Wang (2002), Martens and Zein (2002), Corsi (2003) and others.

Notice that the definition of realized volatility (as any other definition of historical volatility) involves two time parameters, the intraday return interval Δ and the aggregation period $1d$. In order to have a statistically reliable measure of volatility, the parameters must be such that the aggregation period $1d$ is much greater than Δ.

For example, for a Gaussian random walk with constant variance σ^2, the order of magnitude of the statistical error of the variance estimator can be computed. For this model, the sum of square returns has a chi-square distribution with degree of freedom equal to the number of terms in the sum χ^2_M. Setting the mean to zero, the root mean square error (RMSE) for a χ^2_M distribution is given by $\sigma^2\sqrt{2/M}$. Hence, for a Gaussian random walk,

---

[4]For the unbiased property, formally the zero mean assumption should be made, but the results remains approximately true for stochastically evolving mean process, while for its consistency this assumption is not required. See Andersen et al. (2005).

typical values for the RMSE of the 24 hours daily variance measured with return at different time intervals Δ are:

$$\begin{aligned}
\text{RMSE } [\Delta = 1\text{d}] &= 141.4\%\ \sigma^2 \\
\text{RMSE } [\Delta = 1\text{h}] &= 28.8\%\ \sigma^2 \\
\text{RMSE } [\Delta = 1'] &= 3.7\%\ \sigma^2 \\
\text{RMSE } [\Delta = 1''] &= 0.5\%\ \sigma^2.
\end{aligned}$$

These numbers clearly show the advantage of taking returns at the highest frequency in order to measure the daily volatility.

Much more general results for the asymptotic distribution of the realized variance, realized volatility and log realized volatility has been derived by Barndorff-Nielsen and Shephard (2005) and Barndorff-Nielsen, Graversen, Jacod, Podolskij and Shephard (2005). They also show that the measurement errors of realized volatilities computed over consecutive periods are approximately uncorrelated. This property will turns out to be very convenient for the time series modelling of realized volatility.

1.3 Market Microstructure

1.3.1 Scaling analysis of the realized volatility

In practice, however, empirical data differs in many ways from the frictionless continuous-time price process assumed in those theoretical studies. Beside the obvious consideration that a continuous record of prices is not available, other reasons prevent the applications of the limit theory necessary to achieve consistency of the realized volatility estimator. In fact, because of market microstructure effects[5], the assumption that log asset prices evolve as a diffusion process becomes less and less realistic as the time scale reduces. At the tick time scale, the empirical data differs from the simple theoretical model, and the volatility computed with very short time intervals is no longer an unbiased and consistent estimator of the daily volatility computed with daily returns.

This effect can be empirically verified by studying the unconditional expectation of the squared returns[6] $\mathbb{E}\big[r^2(\Delta)\big]$ as a function of the returns frequency Δ (because $\mathbb{E}\big[\sigma^2(\Delta)\big] = \mathbb{E}\big[r^2(\Delta)\big]$ up to finite sample effects that are negligible). Since returns $r(\Delta)$ are intended to

[5] For a good empirically oriented overview of market microstructure effects, see Hasbrouck (1996).

[6] Here, the time subscript is dropped in order to simplify the notation.

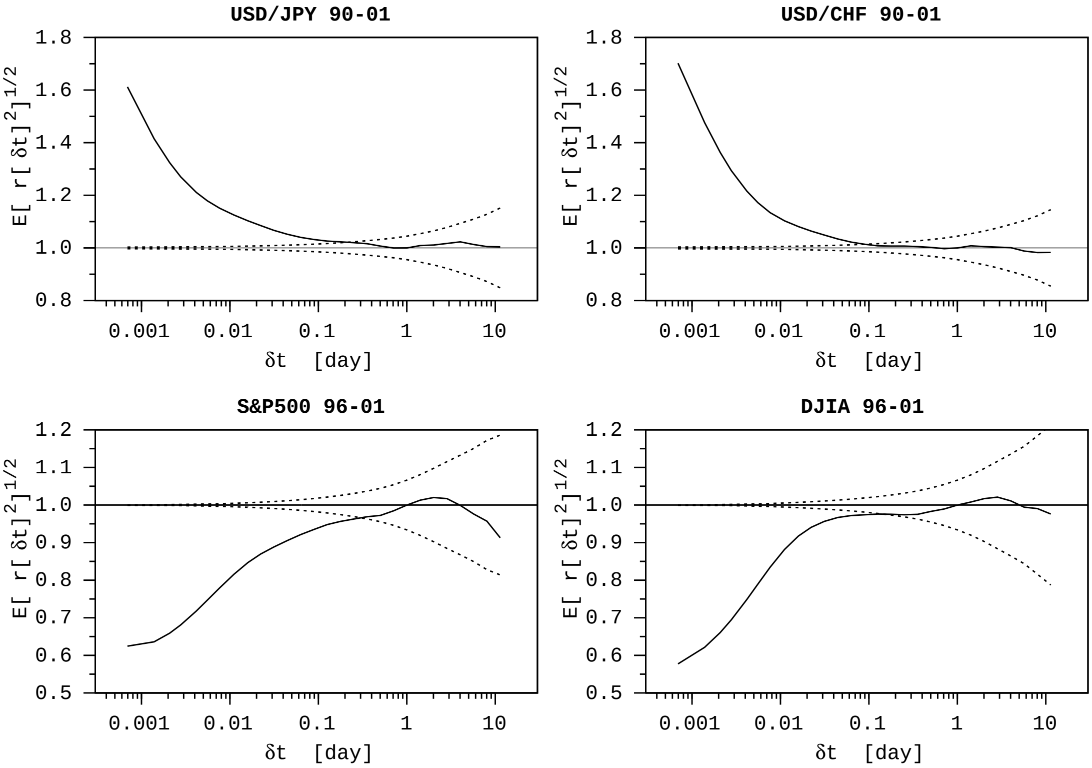


Figure 1.1: Scaling of the annualized volatilities $\mathbb{E}\big[r^2(\Delta)\big]$ for Δ ranging from 1 minute to 1 week. The dotted lines are error bounds assuming independant returns. Two FX rates, USD/JPY and USD/CHF, and two stock indices, Standard & Poors 500 and Dow Jones Industrial Average, are investigated.

be already annualized, for an uncorrelated diffusion process, $\mathbb{E}\big[r^2(\Delta)\big]$ should be independent of Δ.

As an example, figure 1.1 shows the scaling behavior of $\mathbb{E}\big[r^2(\Delta)\big]$ for two currencies and two stock indices. In order to ease the comparison between assets, the empirical average has been normalized such that $\mathbb{E}\big[r^2(1\text{d})\big] = 1$. The horizontal line at 1 corresponds to the expected volatility of an i.i.d. diffusion process. Clearly, the unconditional expectation of the variance computed with returns taken at small time intervals is not equal to the volatility obtained with daily returns:

$$\mathbb{E}\big[r^2(\Delta)\big] \neq \mathbb{E}\big[r^2(1\text{d})\big] \qquad \text{and} \qquad \mathbb{E}\big[\sigma^2(\Delta)\big] \neq \mathbb{E}\big[\sigma^2(1\text{d})\big]\,.$$

The variance estimator computed at a short time interval is strongly *biased* as compared to the mean squared daily return. We use the term bias because for a majority of the agents operating on financial markets, the relevant variable of interest is the daily return or the "one-day risk". Because of widely available daily data, the *de facto* time horizon of reference is one day. These agents are neither interested in the detailed price behavior happening at very short time intervals nor in the volatility observed at a 5-minutes time horizon. The use of high-frequency returns to compute the daily volatility is merely a measurement issue: short-term returns are used because we want to improve the estimation of volatility, rather than being interested in risks existing at the extremely short time frames. Therefore, tolerating this bias would lead to distorted daily risk measures when using high-frequency data instead of daily data.

The size of this bias is directly measured by the vertical distance between the theoretical i.i.d. horizontal line and the empirical one depicted in figure 1.1. From a study of many assets, the general behavior of the bias can be summarized as follow:

- FX: strong positive bias. At the 1-minute level, it ranges from 30% to about 80%.
- Futures and individual stock: positive bias.
- Stock indices: negative bias.

In general, the bias tends to be higher for assets with lower liquidity (for example USD/ITL has a bias of more than 80% at 3-minute level) than that of assets which exhibit higher liquidity.

However, a point of caution should be mentioned here: these results are fairly sensitive to the type of linear interpolation scheme chosen to convert the original time inhomogeneous (i.e. unequally spaced) tick-by-tick series of prices into a time homogeneous (equally spaced) price series. When two subsequent ticks are separated by a length of time longer than that of the time interval of the grid, linear interpolation implicitly assumes a minimum volatility in this interval, and introduces an artificial correlation between the subsequent returns of the generated regular time series. This leads to a systematic underestimation of volatility, that becomes larger as the tick frequency decreases and the number of empty intervals without ticks increases. This could explain the negative bias reported by some authors for less liquid FX rates, whereas, with a different interpolation scheme (previous-tick), we always found a positive bias for FX rates.

For these reasons, when generating a synthetic regular time series for volatility estimation purposes, it is more appropriate to use the "previous-tick interpolation" scheme, in which each tick remains valid until a new tick arrives. This interpolation scheme does not generate an underestimation of volatilities, nor a distortion of the autocorrelation. Using previous-tick interpolation on FX data, we found significant evidences of large biases at 30-20-minutes level, even for major exchange rates. In our analysis, the return interval at which the bias on FX data is no longer significant occurs only at the level of some hours. This means that even for the most liquid FX rates, the shortest return interval to obtain an unbiased volatility measure is of the order of 2-3 hours, leading to only 8-12 observations per day. Furthermore, this “unbiased return interval” considerably changes from asset to asset. Therefore, under these circumstances, without any specific treatment of the bias, the stochastic error of the volatility measure cannot be essentially reduced and cannot be computed with the same return interval for all instruments. Hence, if we want to have a general and homogeneous volatility estimation of optimal precision, an explicit treatment of the bias is required. The first step in this direction is to better understand the origin of the bias, which is the subject of the next two sections.

1.3.2 Autocorrelation analysis

The anomalous scaling of the second moment can only be explained by a non-zero autocorrelation $\rho(k)$ of the returns i.e. $\mathbb{E}\big[\sigma^2(\Delta)\big] \neq \mathbb{E}\big[\sigma^2(1\mathrm{d})\big]$ if and only if $\rho(k) \neq 0$ for some lags k. In short, returns must be not i.i.d. at short time scales. This can be derived by the following computations. From the return at high frequency Δ_1, the return at a time scale m times longer $\Delta_2 = m\,\Delta_1$, can simply be computed by aggregation (with an overall multiplication factor $1/\sqrt{m}$ to take care of the annualization). Then the variances computed with returns at scale Δ_1 and Δ_2 can be related. After some algebra, we find

$$\begin{aligned}\sigma^2(\Delta_2) &= \sigma^2(\Delta_1)\frac{1}{m}\sum_{k=1}^{m-1}\sum_{l=1}^{m-1}\rho(k-l) \\ &= \sigma^2(\Delta_1)\left[1+2\sum_{k=1}^{m-1}(1-\frac{k}{m})\rho(k)\right]\end{aligned} \tag{1.5}$$

where $\rho(k)$ is the autocorrelation function of the return $r(\Delta_1)$ at lag $k \cdot \Delta_1$. We can use this formula to compute the scaling of the mean volatility given the autocorrelation function of the returns. Qualitatively, we can have one of the three following cases:

- no autocorrelation $\rho = 0$ implies $\sigma^2(\Delta) = \sigma^2(1\mathrm{d})$, i.e. no anomalous scaling
- negative autocorrelation $\rho < 0$ implies $\sigma^2(\Delta) > \sigma^2(1\mathrm{d})$, i.e. positive anomalous scaling
- positive autocorrelation; $\rho > 0$ implies $\sigma^2(\Delta) < \sigma^2(1\mathrm{d})$, i.e. negative anomalous scaling

We then expect a positive autocorrelation for stock indices and a negative one for FX, stocks and futures. To confirm those expectations, we performed an intraday autocorrelation study of the returns for several stock indices and FX returns. For this computation, synthetic regular time series of price are generated, where "regular" is meant according to different time scales:

- The physical time scale. This is the usual physical time (including weekend).
- The tick time scale. This clock moves by one unit for every incoming tick. When computing the autocorrelation of tick returns on this scale, we ignore the varying time intervals between ticks.
- The dynamic ϑ-time scale. The ϑ-time scale is a sophisticated business time scale designed to remove intraday and intraweek seasonalities by compressing periods of inactivity while expanding periods of higher activity (Dacorogna, Müller, Nagler, Olsen and Pictet 1993, Breymann, Zumbach, Dacorogna and Müller 2000). It is essentially an intraday generalization of the usual daily business time scale that omits weekends and holidays.

The results are plotted in figure 1.2. The most evident result is the very strongly negative first-lag autocorrelation for FX computed in tick time, with a value around -40%. This contrasts with some mildly negative first-order autocorrelations reported in older studies on high-frequency FX data (Goodhart 1989, Goodhart 1991). Let us emphasize that the bid-ask bounce "a la Roll" cannot be invoked, since logarithmic middle prices are used in these computations. The bid-ask bouncing as described in Roll (1984) is due to the random hitting of the transactions at the bid or ask quotes, for a fixed bid and ask prices. Roll's explanation of the short-term bouncing thus exclusively applies to time series of transaction prices and not to those of middle prices as used in the present computations.

Compared to the autocorrelation evaluated in physical time and ϑ-time, the one computed in tick time presents a strongly negative value at the first lag, about twice as large. The decay at subsequent lags is faster. These differences between the autocorrelation in tick time and in the other time scales can be understood by considering the time deformation induced by the transformation of the tick time scale into a physical or ϑ-time scale. For example, figure 1.3 shows the probability density function (pdf) of the time intervals between ticks measured in physical time for 10 years of USD/CHF. Computing the autocorrelation function of, say, one-minute returns in physical time would imply that:

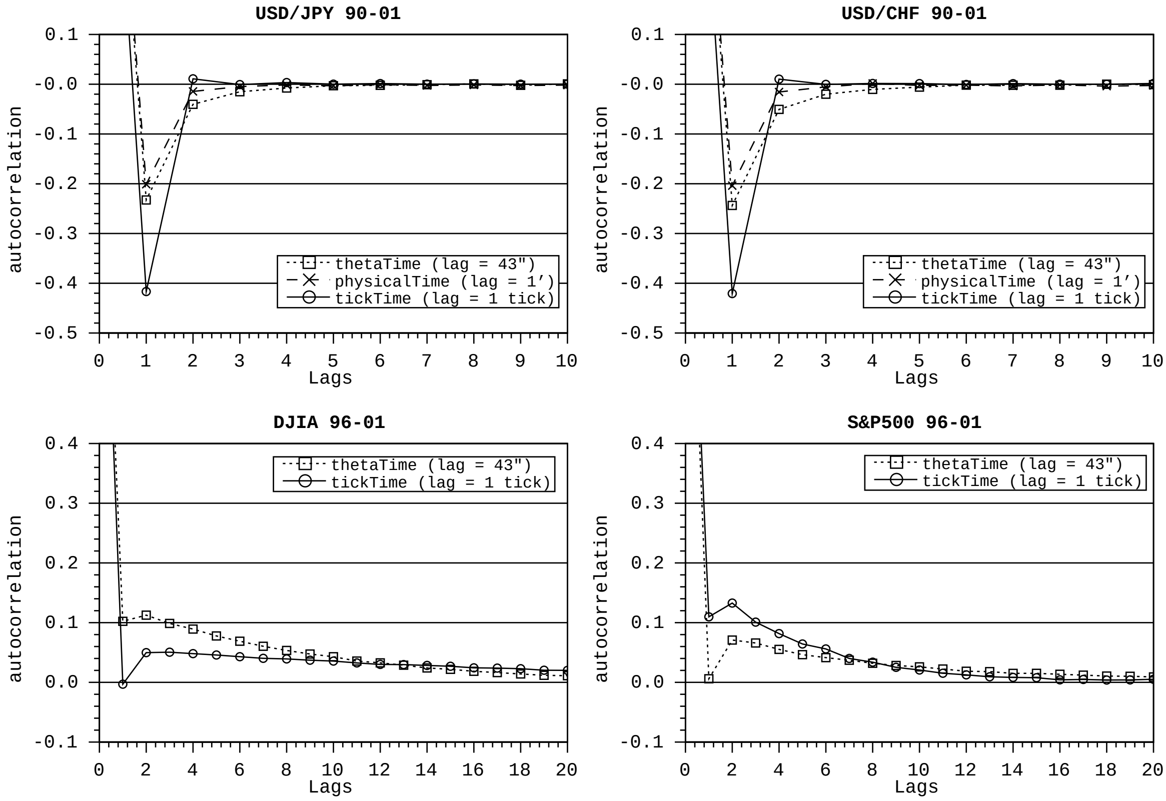


Figure 1.2: Returns autocorrelations for FX (top panels) and stock indices (bottom ones).

- If many ticks are in the same one-minute interval, most of them are ignored in the computation of one-minute returns, and the returns are determined more by the true process and less by microstructural effects.

- If the time interval between two ticks is (much) larger than one minute, the previous-tick interpolation leads to zero returns in one or more one-minute intervals. Zero returns dampen lagged covariances, but the overall expectation of squared returns is not affected by a change of time scale. Since the autocorrelation is the quotient of lagged covariance and variance of returns, the result is a reduced autocorrelation at lag one. Beside, two returns separated by a string of empty one minute intervals share the incoherent component (see the model in the next section). In this case, the strongly negative first-lag autocorrelation in tick time directly affects two distant one-minute intervals, leading to the slow decay at larger lags of the correlation.

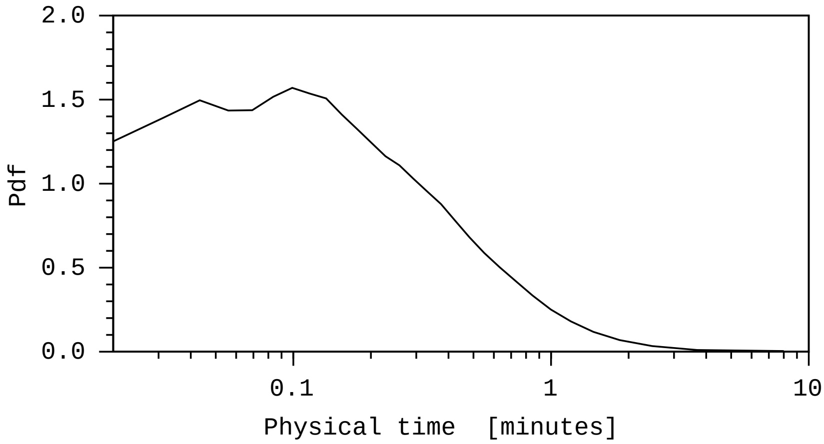


Figure 1.3: Probability density function (pdf) of the time intervals between ticks (semi-logarithmic scale) for USD/CHF. The data sample ranges from 1.1.1990 to 1.1.2000.

All these effects contribute to the empirically found behavior, notably the attenuated autocorrelation at the first lag.

The existence of a very pronounced mean-reverting behavior of the FX quotes is also revealed by a closer visual inspection of the price dynamics as displayed in figure 1.4. Thus, the considerable anomalous scaling of FX volatility is entirely due to the strong negative correlation occurring between subsequent returns.

1.3.3 Sources of microstructure effects

So far, no economic explanation has been presented to account for those statistical properties. In order to develop a microscopic model for the price, the real structure of the market must be taken into account as the return autocorrelation is directly related to microstructure effects arising from the price formation process.

The two main universal sources of microstructure effects are the bid-ask spread and price discreteness. Studies on the bid-ask spread are largely developed within the framework of quote-driven markets. However, the bid-ask spread is not unique to the dealer markets: Cohen et al. (1981) establish the existence of the bid-ask spread in a limit-order market when investors face transaction costs in assessing information, monitoring the market, and conveying orders to the market; Glosten (1994) shows that limit-order markets have a positive bid-ask spread arising from the possibility of trading on private information. On the effects of the bid-ask spread on the price process, already during the '80s, Roll (1984) and Blume and Stambaugh (1983), analysing the statistical properties of tick-by-tick returns, showed

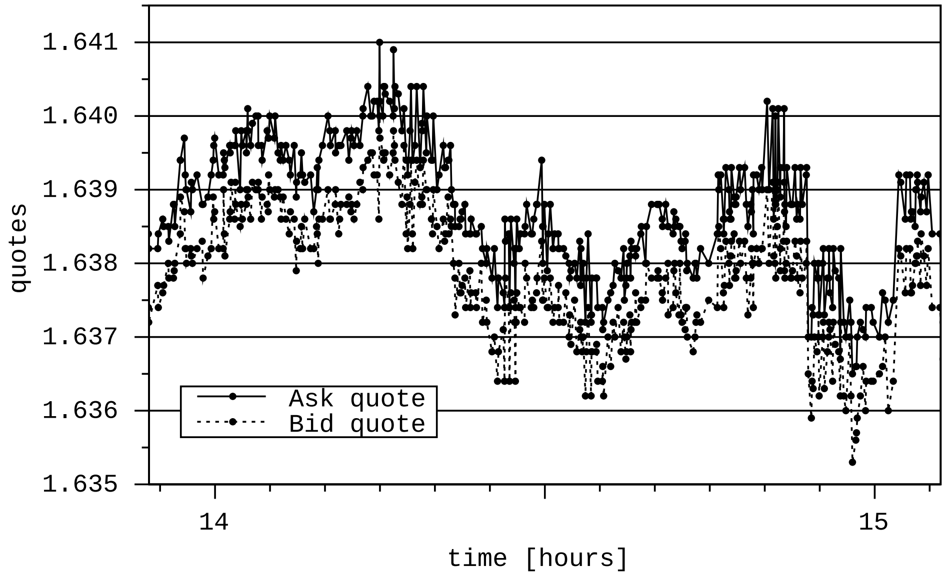


Figure 1.4: A sample price dynamics for the USD/CHF exchange rate. The data are from Reuters, the 06.04.2000, with GMT time.

that in general bid-ask spreads produce negative first-order autocovariances in observed price changes.

Similarly, if one makes the assumption that observed prices are obtained by rounding underlying true values, Glottlieb and Kalay (1985) and Harris (1990) showed that price discreteness induces negative serial covariance in the observed returns. A significant negative autocorrelation induces a bias of positive sign, which increases with the sampling frequency.

This two general microstructure effects are the one responsible for the negative autocorrelation (and hence positive bias) observed in the high frequency returns of individual stocks and future prices.

For the FX, however, since the data employed are not transaction prices but rather the mid point of a bid-ask quote, the bid-ask spread can not be invoke to explain the very large positive bias found in the scaling analysis. Nonetheless, the size of the bias remain too large to be completely attributable to price discreteness effects only. We attribute such strong mean-reverting behavior of the FX quotes to the presence of an "incoherence" effect in the price formation originating from the multiple contributor structure of the FX spot market. A large part of the FX market is in fact an over-the-counter market where all dealers publish their own price quotes. Some consequences of the multiple contributor structure are:

- Disagreement on what the "true price" should be, due to the fact that opinions on public information, strategies of individual contributors, and private information sets are not uniform.

- Market maker bias towards bid or ask prices. Depending on their inventory position, market makers have preferences for either selling or buying. Hence to attract traders to deal in the desired direction they publish new quotes so that either the bid or the ask is competitive. The other price of the bid-ask pair, being pushed far away, influences the level of the (logarithmic) middle price, inducing a very short-term random bouncing of it.

- Fighting-screen effects for advertising purposes. In order to maintain their name on the data suppliers' screens (such as those of Reuters, Bloomberg or Bridge), some contributors keep publishing fake quotes generated by computer programs that randomly modify the most recent quotes (or a moving average of them).

- Delayed quotes. Trader interviews, comparisons to transaction data from electronic trading systems and lead-lag correlation studies show that many contributors release quotes with a considerable time delay (in some cases larger than a minute).

While individual agents follow their different strategies in a coherent way, the whole market generates an incoherent component due to the price formation process that is responsible for the strong negative autocorrelation of FX returns. Note again that the random hitting of bid or ask prices by transactions as describe in Roll (1984) can not be considered here because it does not affect the logarithmic middle price.

For stock indices, on the contrary, we find a large significant positive autocorrelation that lasts up to few hours[7], as already reported by many other authors (G.Hawawini 1980, Conrad and Kaul 1988, Lo and MacKinlay 1988). This positive lagged correlation, pervasive across the three different time scales, sample periods and countries, can be explained by the so-called *lagged adjustment model* (Holden and Subrahamanyam 1992, Breymann, Jegadeesh and Swaminathan 1993). According to this model, among the stocks that compose the indices, there are "leading" stocks that react quickly to new information whereas others stocks partially adjust or adjust with a certain delay, due to either information transmission, non-trading or lower volume. Since the lagged covariance of a portfolio is a weighted average of the lagged cross-covariance between the stocks composing it, a positive autocorrelation results and hence a negative bias arises.

[7]While no positive autocorrelation is found in stocks returns themselves or in futures contracts on indices (Ahn, Boudoukh, Richardson and Whitelaw 1999).

1.4 Conclusions

Summarizing, market microstructure generates a transitory effect on the dynamics of the informationally efficient price. Perturbations of the underlying price induces a non-zero autocorrelation in the returns process which makes no longer true that the variance of the sum is the sum of the variances. This autocorrelation of returns is not negligible at short time scales (usually less than an hour), causing the volatility scaling to strongly deviate from that of a standard i.i.d. process. Thus, the volatility computed with short time intervals becomes a potentially highly biased estimator of the daily volatility. A significant negative autocorrelation induces a bias of positive sign, i.e. the expectation of daily realized volatility computed with high frequency returns is systematically larger than the volatility of the true unobservable process. On the contrary, a relatively long lasting persistence of stock index returns gives rise to a positive slope of the scaling function (i.e. a negative bias). Such biases increase with the sampling frequency.

Therefore, a trade-off arises: on one hand, efficiency considerations suggest to use a very high number of return observations to reduce the stochastic error of volatility estimation. On the other hand, market microstructure introduces a bias that grows as the sampling frequency increases. Therefore, without any explicit treatment of the bias, the stochastic error of the volatility measure cannot be substantially reduced. A solution to this trade-off which directly try to correct for the microstructure effects at the tick-by-tick level, would permits to fully exploit all the information contained in the high frequency data.

Chapter 2

Measuring Realized Volatility

2.1 Introduction

The previous chapter has shown how, in the standard estimation of realized volatility, a bias-variance trade-off arises: on one hand, variance reduction considerations suggest to use very high frequency returns while on the other hand, the bias generated by microstructure effects imposes a reduction of the sampling frequency.

Given such a trade-off between efficiency and bias, a simple approach to overcome this problem is to choose, for each financial instrument, the shortest return interval at which the resulting volatility is still not significantly affected by the bias; that is, to find that frequency wich strike an optimal (usually in mean square sense) compromise between bias and variance. This approach exploits the different aggregation properties between the integrated process of the efficient price and the non-scaling behaviour of the pricing error term. As the aggregation of returns increases the impact of the transitory component on the volatility decreases, reducing the size of the bias. This approach is simple, fully nonparamentric and robust to any source of microstructure effects. On the other hand, in practice this "unbiased return frequency" turns out to be fairly low[1], leaving us with only few return observations per day.

A better solution to this trade-off which permits to fully exploit the information contained in high frequency data, is to have an explicit treatment of the bias trying to correct for the microstructure effects at the tick-by-tick level. The purpose of this chapter is to present some

[1]The answer to this question also depends on the interpolation scheme employed when a regular time series is constructed.

alternative definitions of realized volatility that make use of very high-frequency data, without suffering from the aforementioned systematic deviation. In order to motivate and describe such alternative realized volatility definitions we first briefly review a standard discrete time model for the tick-by-tick price process.

2.2 Price process with microstructure effects

As described in Hasbrouk (1993, 1996), a general way to model the impact of various sources of microstructure effects is to decompose the observed price into the sum of two unobservable components: a martingale component representing the informationally efficient price process and a stationary pricing error component expressing the discrepancy between the efficient price and the observed one. The dynamics of the true latent price can be modelled as a general continuous time Stochastic Volatility (SV) process [2]

$$d\tilde{p}(t) = \mu(t)dt + \sigma(t)dW(t) \tag{2.1}$$

where $\tilde{p}(t)$ is the logarithm of the true instantaneous price, $\mu(t)$ is the finite variation process of the drift, $dW(t)$ is a standard Brownian motion, and $\sigma(t)$ is the instantaneous volatility. For this diffusion, the notional or actual variance is equivalent to the integrated variance for the day t (IV_t)[3] which is the integral of the instantaneous variance of the underling true process $\sigma^2(t)$ over the one day interval $[t-1;t]$, i.e. $IV_t = \int_{t-1}^{t} \sigma^2(\omega)d\omega$.

The observed (logarithmic) price, being recorded only at certain intraday sampling times and contaminated by market microstructure effects, is instead a discrete time process described in the “intrinsic transaction time” or ”tick time”[4] denoted with the integer index n:

$$p_{t,n} = \tilde{p}_{t,n} + \eta_t \omega_{t,n} \tag{2.2}$$

where $\tilde{p}_{t,n}$ is the unobserved true price at intraday sampling time n in day t and $\eta_t \omega_{t,n}$ represents the pricing error component with η_t the size of the perturbation. Depending on the structure imposed on the pricing error component, many structural models for microstructure

[2] Alternatively, a pure jump process as the compound Poisson process proposed by Oomen (2005) could be employed to model the dynamics of the true price process.

[3] We use the notation (t) to indicate instantaneous variable while subscript t denote daily quantities.

[4] That is, a time scale having the number of trades as its directing process (here we don’t make the distinction between tick time and transaction time).

effects could be recovered. Here we take a more statistical perspective assuming ω to simply be a zero mean nuisance component independent of the price process. In this section the assumption of an i.i.d. noise process for ω is made while it will be relaxed to allow for more general dependence structure in section 2.4.6 .

According to the Mixture of Distribution Hypothesis originally proposed by Clark (1973) and extended and refined in numerous subsequent works, the price process observed under the appropriate transaction time should appear as a diffusion with constant volatility. A Brownian motion in tick time is, in fact, a subordinate stochastic process which has been shown to properly accommodate for many empirical regularities. In other words, in tick time even a simple constant volatility process can reproduce stylized facts observed in physical time such as heteroskedasticity, volatility clustering, fat tails and others. Hence, the hypothesis of homoskedastic processes in tick time is far less restrictive compared to the same hypothesis made for processes defined in physical time (where this assumption would be clearly violated by the empirical data). Finally, the robustness of the different volatility measures against violation of the assumption of homoskedasticity in tick time is checked in the simulation study by explicitly employing an heteroskedastic DGP for the true tick-by-tick price process.

Computing daily volatility in tick or transaction time also presents several practical advantages (see Oomen 2005 for a detailed comparison of different sampling schemes). Intuitively, in tick time all observations are used so that no information is wasted. The interpolation error and noise arising from the construction of the artificial regular grid is avoided. Moreover, using a tick time grid the underlying price process tends to be sampled more frequently when the market is more active, that is, when it is needed more because the price moves more.

The observed tick-by-tick return $r_{t,n}$ of day t at time n can then be decomposed as

$$r_{t,n} = \sigma_t \tilde{\epsilon}_{t,n} + \eta_t \left(\omega_{t,n} - \omega_{t,n-1} \right) \tag{2.3}$$

where the unobserved innovation of the efficient price $\tilde{\epsilon}_{t,n}$ and the pricing error $\omega_{t,n}$, are independent $\mathrm{IID}\,(0,1)$ processes. Hence, for each day the tick-by-tick return process is a MA(1) with $\mathbb{E}\left(r_{t,n}\right) = 0$ and autocovariance function given by

$$\mathbb{E}\left[r_{t,n} r_{t,n-h}\right] = \begin{cases} \sigma_t^2 \quad + \quad 2\eta_t^2 & \text{for } h = 0 \\ - \;\; \eta_t^2 & \text{for } h = 1 \\ 0 & \text{for } h \geq 2 \end{cases} \tag{2.4}$$

where σ_t^2 represents the tick-by-tick variance of the unobserved true price for day t and η_t^2 the extra variance in the observed returns coming from the market microstructure noise observed during that day. Under these assumptions, the integrated variance of day t simply becomes $IV_t = N_t \sigma_t^2$, with N_t the number of ticks occurred in day t.

According to this model, the observed variance is equal to the "true integrated variance" (the variance of the process describing the dynamics of the true price) plus an additional term coming from the microstructure effects. This last term is responsible for the observed bias of the volatility. As long as the length of the return interval is sufficiently long (say a number of ticks equivalent to one day or one week in physical time) the contribution of the microstructure noise is negligible and so is the bias of the volatility estimation. But when high-frequency data are used the contribution of the additional component increases and the size of the bias is no longer negligible.

Equation (2.4) also implies $-0.5 \leq \rho(1) \leq 0$. Where the lower bound -0.5 is reached when σ_t^2 is completely negligible compared to η_t^2, and the return is the lag one difference of a noise. An empirical autocorrelation around -0.4, as observed for the USD/JPY and USD/CHF, implies $\eta_t^2 \simeq 2\sigma_t^2$. This indicates that at tick-by-tick level the volatility originating from the microstructure effects is largely predominant. Therefore, this effect should be carefully considered before using data at very high frequency.

The model can easily be extended for stock indices by introducing an autoregressive structure in the true return $\tilde{r}_{t,n}$:

$$\tilde{r}_{t,n} = \phi_t \, \tilde{r}_{t,n-1} + \epsilon_{t,n} \tag{2.5}$$

with $\epsilon \simeq$ i.i.d.$(0,\, \sigma_\epsilon^2)$. Then the autocovariance structure of the model becomes

$$\mathbb{E}\left[r_{t,n} r_{t,n-h}\right] = \begin{cases} \sigma_t^2 + 2\eta_t^2 & \text{for } h = 0 \\ \phi_t \sigma_t^2 - \eta_t^2 & \text{for } h = 1 \\ \phi_t^h \sigma_t^2 & \text{for } h \geq 2 \end{cases} \tag{2.6}$$

with $\sigma_t^2 = \sigma_{\epsilon,t}^2/(1-\phi_t^2)$. This lagged correlation replicates the empirical data for stock indices as shown on figure 1.2, where a small incoherent effect at lag one is present.

Thus, this simple model is capable of reproducing the main empirical evidences found for financial data. In particular it is able to replicate the strong negative first-lag autocorrelation of tick-by-tick returns and the observed anomalous scaling of realized volatility. Recently, this model (in some cases without the i.i.d assumption on the noise) as been employed to study in a more formal setting the impact of the microstructure noise on the realized volatility measure as in Hansen and Lunde (2004), Bandi and Russel (2004), Aït-Sahalia, Mykland and Zhang (2005), among others.

2.3 A simple EMA filter

On the basis of the above model, it is then possible to design high-frequency estimators for volatility that appropriately discount for the bias induced by the microstructure effects. Surprisingly, until recently, only very few studies has pursued along this line.

According to equation (2.4) the simplest approach is to applied a first order serial covariance correction as first proposed, outside the realized volatility literature but for a similar problem, by French and Roll 1986, Harris 1990 and Zhou 1996 and recently revived in the contest of realized volatility literature by Oomen (2005) and Hansen and Lunde (2005). Essentially, in this approach, the standard estimator which has expectation of $\sigma^2 + 2\eta^2$ is corrected by twice the sum of cross products of adjacent returns which on average should be equal to $-2\eta^2$ (here and in the following, since we will compute the Realized Volatility for each day separately, the daily subscript t is suppressed to simplify the notation).

However, this variance estimator, beside being very noisy due to the instability of the (local) covariance correction, suffers from the serious problem of having the possibility to become negative (because of the large cancellation between the two estimators for the variance and lag one covariance). The non positivity of this definition is a serious drawback, which become particularly manifest when the number of ticks in the interval is not large enough. To overcame those problems Corsi, Zumbach, Müller and Dacorogna (2001) has proposed a simple exponential moving average (EMA) filtering of the tick-by-tick price process.

The basic idea behind this simple estimator starts from the observation that r_n has a MA(1) representation

$$r_n = w_n - \theta w_{n-1} = (1 - \theta L) w_n \tag{2.7}$$

with $w_n = \text{i.i.d}(0,\ \Omega^2(\sigma, \eta))$ and $\theta = f(\sigma, \eta)$. This representation can then be inverted to give

$$w = (1 - \theta L)^{-1} r \tag{2.8}$$

the $(1 - \theta L)^{-1}$ operator is related to an exponential moving average (EMA) defined by the iterative equation

$$\text{EMA}[\theta; r]_n = \theta\, \text{EMA}[\theta; r]_{n-1} + (1 - \theta) r_n. \tag{2.9}$$

When iterating the EMA definition, we obtain

$$\begin{aligned} \text{EMA}[\theta; r]_n &= (1 - \theta) \left\{ r_n + \theta r_{n-1} + \theta^2 r_{n-2} + \cdots \right\} \\ &= \left\{ (1 - \theta)(1 - \theta L)^{-1} r \right\}_n . \end{aligned} \tag{2.10}$$

Therefore, the white noise term w can be computed by a simple EMA in tick time[5]

$$w = (1-\theta)^{-1}\mathrm{EMA}[\theta; r]. \tag{2.11}$$

Furthermore, as $\mathbb{E}\big[\mathrm{EMA}[\theta; r]^2\big] = (1-\theta)^2\mathrm{E}[w^2] = \sigma^2$, the EMA filtering does not change the volatility. The appropriate parameters θ can be estimated from the first-lag autocorrelation of the return in tick time. Using the representation 2.7, the first-lag correlation is

$$\rho(1) = \frac{-\theta}{1+\theta^2} \tag{2.12}$$

and solving for θ gives

$$\theta = -\frac{1}{2\rho(1)}\left(1 - \sqrt{1 - 4\rho(1)^2}\right). \tag{2.13}$$

Because the EMA operator is linear, filtering the return is equivalent to filtering the price. Therefore, a tick-by-tick price time series, filtered from the incoherent component, is defined by

$$\mathcal{F}(x) = \mathrm{EMA}[\theta; x]. \tag{2.14}$$

Then, a regular time series can be computed and the realized volatility estimated using the definitions 1.3. We will call this volatility estimator the *EMA-filtered realized volatility* . The parameter $\hat{\theta}$ must be estimated from the first-lag correlation $\rho(1)$ of the returns and equation (2.13). Since the bias component, though dynamically changing, is relatively stable over short time periods on FX data, we can estimate the first-lag correlation on a moving window longer than one day. The choice of the length and the kernel of the moving window only has a moderate influence on results. The procedure behaves well over a wide range of choices; our choice is a window length of the order of two weeks. With a moving window of T days ending at the time t to filter the price at t, we have a causal estimate where the information ahead of the current time of the price filtering procedure is not used.

Having the moving estimate for the lag one correlation $\hat{\rho}[T](t)$, the parameter $\hat{\theta}(t)$ can be obtained using equation (2.13), and the filtered price computed with $\mathcal{F}(x) = \mathrm{EMA}[\hat{\theta}; x]$ where $\hat{\theta}$ is now a time series.

To summarize, the EMA-filtered realized volatility is computed as follows:

- estimate the first-lag autocorrelations of tick returns on a moving sample,
- compute $\hat{\theta}(t)$, using equation (2.13),

[5] In this context, this is equivalent to the application of a Kalman filter since it can be shown that the EMA is the steady-state solution of this particular model. See (Harvey 1989) pag. 175.

- filter the log-price time series with an EMA in tick time with parameter $\hat{\theta}(t)$,

- compute the return and volatility at the desired frequency in a given time scale.

Let us emphasize that the filtering is a non local one i.e. it employs information outside the estimation interval of the realized volatility. However it is causal, namely the parameter of the filter is evaluated only on past data. Therefore, this filtering technique can be applied in real time, and not only to historical data. Moreover, it is "computationally cheap" since only the computations of EMAs are required.

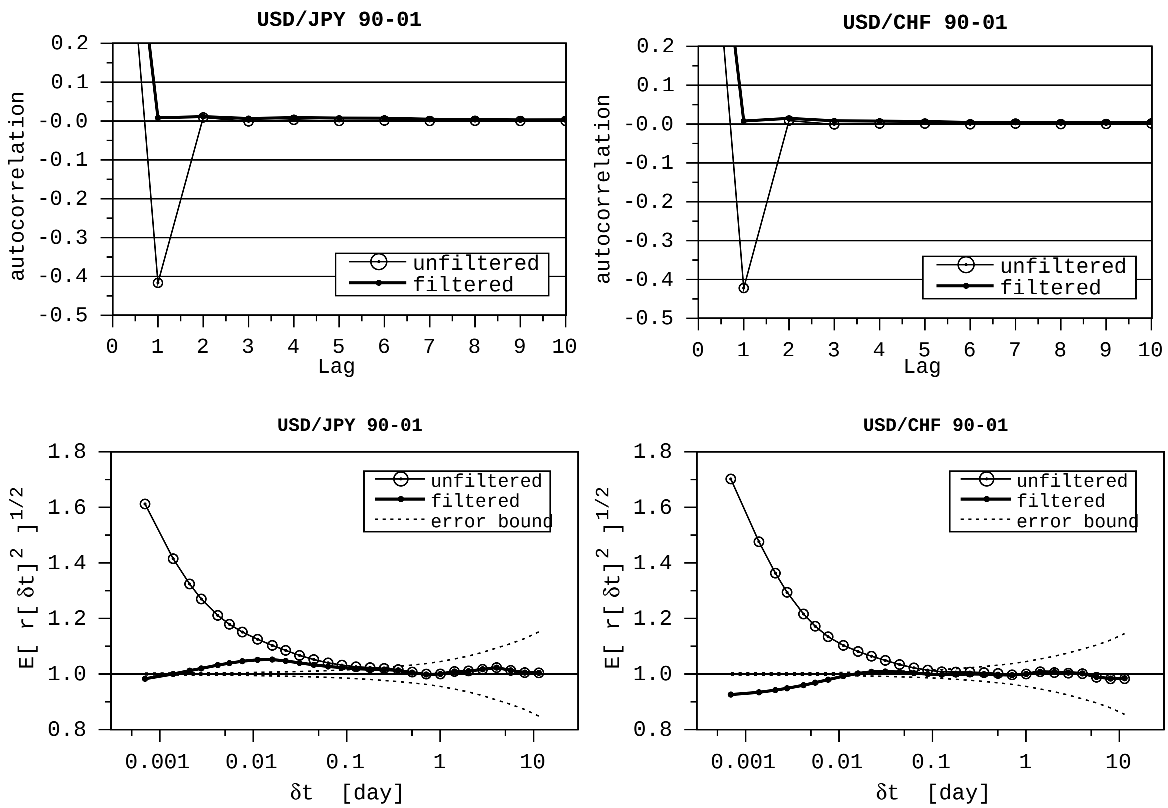


Figure 2.1: Results on tick time autocorrelations (top panels) and volatility scaling (bottom ones) of the application of the EMA-filter to tick-by-tick price series of USD/JPY and USD/CHF. The sample covers 11 years from January 1, 1990, to January 1, 2001. The first-lag correlation is evaluated on a moving sample of length 20 days.

The results of this EMA filtering to several tick-by-tick FX series are reported in figure 2.1 and table 1. The top panels of figure 2.1 display the autocorrelation structures in tick time of the original and filtered return series of USD/JPY and USD/CHF rates. The bottom panels

	$E[\Delta t]$ θ-time [min]	$\rho(1)$ original	$\rho(1)$ filtered
USD/JPY	0.74'	-41.5%	0.78%
USD/CHF	0.82'	-42.0%	0.95%
GBP/USD	0.88'	-40.0%	-0.17%
EUR/USD	0.25'	-41.4%	-1.28%
EUR/GBP	1.9'	-29.8%	-4.66%
USD/ITL	1.6'	-35.0%	-3.15%
USD/DKK	2.3'	-36.0%	-1.38%
GBP/JPY	6.86'	-46.9%	4.8%

Table 2.1: Reduction of the first order autocorrelation of tick-by-tick returns obtained with the EMA-filter.

show the scaling behavior of original and filtered volatility. Clearly, removing the strong negative first-order autocorrelation, considerably reduces the volatility bias. We report in table 1 values of the first-order autocorrelation with and without the application of the EMA-filter for different FX rates. Though not always perfect, the reduction of $\rho(1)$ is remarkable.

2.4 The Discrete Sine Transform Approach

Though effective in reducing the bias (especially on highly liquid data), the EMA filter is a non-local estimator which adapts only slowly to changes in the properties of the pricing error component. Moreover, all the estimators based on the first order covariance, correct only for the bias deriving from the first lag of the return autocorrelation function, while they are very sensitive to non zero higher lag coefficients.

The presence of significant autocorrelation at lags length greater than one and the possibility that each trading day may be characterized by different autocorrelation structures, makes the filtering problem rather complex. In theory, this problem could be tackled by a fully parametric approach where several ARMA models are first estimated every day. Then the best model is chosen on the basis of some loss criteria and finally an estimate for the daily volatility could be obtained from the residuals of the selected model. This parametric higher order covariance correction has been proposed, for instance, by Bollen and Inder (2002) which makes use of a series of AR models selected on the basis of the Schwarz BIC criteria and by Hansen and Lunde (2004) which employ MA(q) filters where q changes with the returns frequency so to keep the time spanned by the autocorrelation window constant. However, such fully parametric approach, beside being asset dependent and sensitive to the loss criteria chosen, relies on the estimation of a large number of parameters, conveying the estimation errors of those parameters to the volatility estimator substantially amplifying its variance. Moreover, Bustos and Yohai (1986) show how even very few outlying observations can largely increase the variance of the estimated residuals. Recently, Barndorff-Nielsen, Hansen, Lunde and Shephard (2004) proposed a modified kernel-based estimator which is asymptotically optimal. Concurrently, Zhang, Mykland and Aït-Sahalia (2005) proposed an estimator based on overlapping subsampling schemes and an appropriate combinantion of two realized volatilities computed at two different time scales. More recently, Zhang (2004) has generalized the Two Scales estimator to a multiple time scales estimator that combines realized volatilities computed at more than two return frequencies and reaches the same asymptotic efficiency of the kernel-based estimator. Our approach will follow the direction of this *Multi-Scales* methodology[6].

In this section new alternative Multi-Scales realized volatility measures based on linear regression approach and *Discrete Sine Transform* (DST) are presented (Curci and Corsi 2003). Multi-Scales estimators similar to that recently proposed by Zhang (2004) can, in fact, be constructed within a simple regression based approach by exploiting the linear relation existing between the market microstructure bias and realized volatilities computed at different frequencies. These regression based estimators can be further improved and robustified by using the DST approach to filtering out most of the market microstructure noise. The motivation for the employment of the DST approach rests on its ability to decorrelate signal

[6] Barndorff-Nielsen et al. (2004) show that a direct link between the Multi-Scales and the kernel-based estimators exists.

for data exhibiting MA type of behaviour which arises naturally in discrete time models of tick-by-tick returns. In fact, we show that the DST diagonalizes exactly MA(1) processes and approximately MA(q) ones. Hence, this nonparametric DST approach, turns out to be very convenient as it provides an orthonormal basis which permits to optimally[7] extract the volatility signal hidden in the noisy tick-by-tick return series. As a result, new nonparametric realized volatility estimators which fully exploit all the available information contained in high frequency data can be constructed. We also show that this approach produces robust and accurate results also in the presence of not i.i.d. microstructure noise which leads to more general MA(q) processes for the tick-by-tick returns. It is then robust against a wide class of noise contaminations and model misspecifications.

Moreover, thanks to this result we derive closed form expression for the score, the Fischer information matrix and the Cramer-Rao bounds of MA(1) processes and provide an efficient numerical procedure for the likelihood maximization.

Finally, although for the ease of exposition we will describe the model having the efficient price which follows a Brownian motion in tick time, the actual implementation of DST approach would only need to assume the volatility in tick time to be constant over a small window of very few ticks (20 or 30), so that the already weak and realistic assumptions of homoskedasticity in tick time is additionally weakened.

2.4.1 Discrete Sine Transform

Considering the vector of M tick-by-tick observed returns $R\left(M,n\right) = \left[\begin{array}{cccc} r_n & r_{n-1} & \cdots & r_{n-M+1} \end{array}\right]^\top$, we develop a Principal Component Analysis[8] of the associated variance-covariance matrix $\Omega^{(M)} = \mathbb{E}\left(R\left(M,n\right)R\left(M,n\right)^\top\right)$, which is a tridiagonal matrix of the form:

$$\Omega^{(M)} = \begin{bmatrix} \sigma^2+2\eta^2 & -\eta^2 & & \\ -\eta^2 & \sigma^2+2\eta^2 & \ddots & \\ & \ddots & \ddots & -\eta^2 \\ & & -\eta^2 & \sigma^2+2\eta^2 \end{bmatrix}$$

[7] In a linear sense.

[8] Also known as Karhunen-Loéve expansion or Hotelling transformation.

By solving the eigenvalue equation $\Omega^{(M)}\varphi_m^{(M)} = \lambda_m^{(M)}\varphi_m^{(M)}$ with $m = 1, 2, \ldots, M$, it can be shown that the eigenvalues of $\Omega^{(M)}$ are given by

$$\lambda_m^{(M)} = \sigma^2 + 4\eta^2 \sin^2 \frac{\pi m}{2(M+1)} \tag{2.15}$$

with $0 < \lambda_1^{(M)} < \lambda_2^{(M)} < \cdots < \lambda_M^{(M)}$. Therefore, the eigenvalues of the DST components are ordered, separated and all non degenerate. The corresponding eigenvectors are

$$\varphi_m^{(M)}(k) = \sqrt{\frac{2}{M+1}} \sin \frac{\pi m k}{M+1} \qquad k = 1, 2, \ldots, M \tag{2.16}$$

The remarkable fact is that, unlike common situations, the eigenvectors ($\varphi_m^{(M)}$) of a MA (1) process are universal and they coincide with the orthonormal basis used in the Discrete Sine Transform (DST). Given that such nonparametric orthogonalization represents the optimal solution to a linear filtering problem, it can be very useful for the analysis of high frequency return data as it provides an universal basis to optimally decorrelate the price signal from market microstructure noise.

2.4.2 The Minimal DST estimator

According to the Principal Component Analysis, the simple and computationally fast DST of the returns

$$c_m^{(M)}(n) = \sum_{k=1}^{M} \varphi_m^{(M)}(k)\, r_{n-k+1}$$

acts as a projector of the signal into its principal components. The variance of the DST components are directly the eigenvalues of the variance-covariance matrix:

$$\mathbb{E}\left(c_m^{(M)}(n)\, c_m^{(M)}(n)\right) = \left(\varphi_m^{(M)}\right)^{\top} \Omega^{(M)}\varphi_m^{(M)} = \lambda_m^{(M)} = \sigma^2 + 4\eta^2 \sin^2 \frac{\pi m}{2(M+1)}$$

Since we are interested in the permanent component of volatility the idea is to consider the projection of the returns on the minimal principal component which is the one less contaminated by the transient volatility coming from the microstructure noise. Therefore, an asymptotically unbiased estimator of the average variance per tick σ^2 is given by the mean value of the square of the DST component associated with the minimal eigenvalue of the correlation matrix ($c_{min}^{(M)}$) in the limit of a large window M

$$\sigma_M^2 \equiv Var\left[c_{min}^{(M)}\right] = \sigma^2 + 4\eta^2 \sin^2 \frac{\pi}{2(M+1)} \tag{2.17}$$

since for large M the effect of the price error vanishes as $\sigma_M^2 \simeq \sigma^2 + \eta^2 \dfrac{\pi^2}{M^2}$.

This clearly shows how the aggregation on the minimal component decreases the impact of the pricing error at a much higher speed compared with the standard aggregation of returns. In fact, in this second case, the bias is reduced at the rate M while on the minimal DST component the bias is cut down at rate M^2, allowing to substantially increase the "unbiased return frequency" and then improving the precision of the volatility estimation.

It is important to note that throughout the paper, in order to reduce estimation errors and assure consistency of the estimators, the computation of any variance estimator at any level of aggregation, is always performed by adopting a full overlapping scheme i.e. (using the terminology introduced by Zhang et al. 2004) by subsampling and averaging. Having an estimate of the average volatility of the tick-by-tick returns for a given day, the corresponding daily volatility is readily obtained by rescaling σ^2 with the number of ticks occurred in that day. We term this volatility measure the *Minimal DST estimator*.

2.4.3 The Multi-Scales Least Square estimator

Recently Zhang, Mykland and Aït-Sahalia (2004) have introduced the "Two-Scales" estimator while Zhang (2004) has generalized this approach to a "Multi-Scales" estimator that combine realized volatilities computed at more than two return frequencies. Here, we present a different approach to the construction of realized volatility estimators computed with multiple time scales.

Under the assumption of i.i.d. noise the conditional expectation of the daily realized variance $RV^{(k_j)}$ computed with observed returns of different tick-lengths k_j is[9]

$$\mathbb{E}\left[RV^{(k_j)}\right] = IV + 2N^{(k_j)}\,\eta^2 \tag{2.18}$$

where $N^{(k_j)}$ is the number of k_j-returns in the day. Hence, a consistent and unbiased estimator can be obtained by computing the realized variance at different frequencies k_j and then estimating IV and η^2 by means of a simple OLS, WLS or GLS linear regression of $RV^{(k_j)}$ on $N^{(k_j)}$. Moreover, in order to relax the i.i.d. assumption for the noise, those frequencies k_j could be selected by choosing only those on the linear part of the plot (analogous to the volatility signature plot) of $\mathbb{E}\left[RV^{(k_j)}\right]$ against the number of observations $N^{(k_j)}$). We will denote this class of estimators as *Multi-Scales Least Square* estimators.

[9] As aforementioned, for returns with a length in ticks $k_j > 1$, subsampling and averaging (i.e. a full overlapping scheme) is adopted.

It is interesting to note that applying this Multi-Scales Least Square approach to only two different frequencies k_1 and k_2, one gets a very simple linear system of two equations in two unknowns (IV and η^2) which can be directly solved, giving as estimator of IV

$$TS = \frac{\alpha RV^{(k_2)} - RV^{(k_1)}}{\alpha - 1} \tag{2.19}$$

where $\alpha = N^{(k_1)}/N^{(k_2)}$ is the ratio between the number of returns sampled at the "base" frequency k_1 and that obtained by sampling at the lower "auxiliary" frequency k_2.

Equation (2.19) is exactly the expression of the Two-Scales estimator for serially dependent noise (with the small sample bias correction) recently proposed by Aït-Sahalia, Mykland and Zhang (2005) as an extention of the Zhang et al. (2005) estimators. Therefore, this alternative "Jack Knife style" derivation of the Aït-Sahalia et al. (2005) estimator shows that the Multi-Scales Least Square approach can be seen as another natural generalization of the Two-Scales estimator to more than just two sampling frequencies.

2.4.4 The Multi-Scales DST estimator

The idea of exploiting linear relations among realized volatility measures computed at different aggregation frequencies by means of simple linear regressions, could also be extended to the DST estimators. Hence, regression based Multi-Scales estimators can be further improved and robustified by using the DST approach to prefilter market microstructure noise. The idea is to exploit the linear relation existing between the realized variance of the minimal principal component $c_{min}^{(M)}$ (i.e. the Minimal DST estimator) and the window length of the PCA M.

In fact, denoting $RV_{min}^{(M)} = \widehat{V[c_{min}^{(M)}]}$ we have:

$$\mathbb{E}\left[RV_{min}^{(M_j)}\right] = \sigma^2 + \eta^2 \mathcal{N}^{(M_j)} \tag{2.20}$$

where $\mathcal{N}^{(M_j)} = 4\sin^2 \frac{\pi}{2(M_j+1)}$.

Therefore, a more effective way of employing the DST decomposition is to evaluate the Minimal DST estimator $RV_{min}^{(M_j)}$ for different values of M_j and then perform a simple linear regression. Then the intercept is an unbiased (not only asymptotically but also in finite sample) and consistent estimator of the tick-by-tick volatility σ^2, while the slope is an estimate of η^2. Once appropriately rescaled by the number of ticks per day, the resulting IV estimator will be called the *Multi-Scales DST estimator*.

2.4.5 Exact MA(1) Likelihood, Cramer-Rao bounds and absolute efficiency

Thanks to the previous results on the universality of the eigenvectors, we can obtain a diagonalization of the variance-covariance matrix of MA(1) processes which does not depend on the parameters to be estimated.

Collecting the M eigenvectors of Ω in the MxM characteristic matrix $\Psi = [\varphi_1 \varphi_2 ..., \varphi_M]$, we can project the return vector onto the orthogonal space of the principal component $C = \Psi^\top R$, which is a Mx1 vector distributed as $C \sim N(0, \Lambda)$, where Λ is the MxM diagonal matrix containing the M eigenvalues of the tridiagonal matrix Ω[10]. Therefore, the likelihood function of R can be rewritten in terms of the principal components vector C as

$$f(C) = \frac{1}{\sqrt{(2\pi)^M \det \Lambda}} \exp\left[-\frac{1}{2} C^\intercal \Lambda^{-1} C\right] = \frac{1}{\sqrt{(2\pi)^M \prod_{n=1}^M \lambda_n}} \exp\left[-\frac{1}{2}\sum_{n=1}^M \frac{c_n^2}{\lambda_n}\right]$$

and

$$\ln f_\theta(C) = -\frac{M}{2}\ln 2\pi - \frac{1}{2}\sum_{n=1}^M \ln \lambda_n - \frac{1}{2}\sum_{n=1}^M \frac{c_n^2}{\lambda_n}$$

Then, from the linear equation (2.15) we readily obtain

$$\frac{\partial \lambda_n}{\partial \theta} = \begin{pmatrix} 1 \\ 4\sin^2 \frac{\pi n}{2(M+1)} \end{pmatrix} \quad \text{and} \quad \frac{\partial^2 \lambda_n}{\partial \theta_i \partial \theta_k} = 0 \quad \text{for} \quad i, k = 1, 2$$

and hence, we are now able to analytically derive the equations for the Score and the Hessian

$$\frac{\partial \ln f_\theta(C)}{\partial \theta_i} = \frac{1}{2}\sum_{n=1}^M \left(\frac{c_n^2}{\lambda_n^2} - \frac{1}{\lambda_n}\right)\frac{\partial \lambda_n}{\partial \theta_i}; \qquad \frac{\partial^2 \ln f_\theta(C)}{\partial \theta_i \partial \theta_k} = -\sum_{n=1}^M \left(\frac{c_n^2}{\lambda_n^3} - \frac{1}{2\lambda_n^2}\right)\frac{\partial \lambda_n}{\partial \theta_i}\frac{\partial \lambda_n}{\partial \theta_k}$$

Therefore, thanks to equation (2.15) and (2.16) we are able to explicitly compute the Fisher Information matrix of an MA(1) process, which reads

$$\mathcal{I}_{ik} = -\mathbb{E}\left(\frac{\partial^2 \ln f_\theta(C)}{\partial \theta_i \partial \theta_k}\right) = \frac{1}{2}\sum_{n=1}^M \frac{1}{\lambda_n^2}\frac{\partial \lambda_n}{\partial \theta_i}\frac{\partial \lambda_n}{\partial \theta_k}$$

With each element of the matrix given by

$$\mathcal{I}_{11} = \frac{1}{2}\sum_{n=1}^M \frac{1}{\lambda_n^2}, \quad \mathcal{I}_{22} = 8\sum_{n=1}^M \frac{1}{\lambda_n^2}\sin^4\left(\frac{\pi n}{2(M+1)}\right), \quad \mathcal{I}_{12} = \mathcal{I}_{21} = 2\sum_{n=1}^M \frac{1}{\lambda_n^2}\sin^2\left(\frac{\pi n}{2(M+1)}\right)$$

[10] In fact, $\mathbb{E}[CC^\top] = \mathbb{E}[\Psi^\top RR^\top \Psi] = \Psi^\top \mathbb{E}[RR^\top]\Psi = \Psi^\top \Omega \Psi = \Lambda$.

Then the Cramer-Rao bounds of $\hat{\sigma}^2$ and $\hat{\eta}^2$ can now be given in closed form as

$$\operatorname{var}\left(\hat{\sigma}^2\right) \geqslant \frac{\mathcal{I}_{22}}{\mathcal{I}_{11}\mathcal{I}_{22} - \mathcal{I}_{12}^2} \qquad \text{and} \qquad \operatorname{var}\left(\hat{\eta}^2\right) \geqslant \frac{\mathcal{I}_{11}}{\mathcal{I}_{11}\mathcal{I}_{22} - \mathcal{I}_{12}^2}$$

These results, which are original to our knowledge, have two important implications. First they obviously permit to evaluate the absolute efficiency of volatility estimators. Second numerical optimization of the exact likelihood is greatly simplified. In fact, given that the principal components do not depend on the parameter, the orthogonalization of the returns process needs to be done only once rather then at each iteration as it occurs using Cholesky factorization (see Hamilton 1994).

From simulations (Table 2.2) it turns out that the Multi-Scales DST estimator for σ^2 possesses a variance very close to the Cramer-Rao bounds. Moreover, if desired, this small loss of efficiency could be easily eliminated by using the Multi-Scales DST estimator as initial value in ML numerical optimization performed with the Newton-Raphson method. It is well known, in fact, that in order for the Newton-Raphson method to be stable and quickly converge, good starting points are required. The Multi-Scales DST estimator seems to be the most appropriate starting point as it guarantees the convergence of the Newton-Raphson algorithm in less than 10 iterations (usually even only 3-4 iterations are enough). Then the combination of the Multi-Scales DST estimator with the Newton-Raphson algorithm (MS-DST + NR) would quickly lead to the fully efficient ML estimator.

2.4.6 Stability and robustness

To judge the stability and robustness of the DST filter with respect to more general specifications of the nuisance component, this section relaxes the i.i.d. assumption for the noise structure which leads to more general MA(q) processes for the observed returns.

It should be noted however, that the presence of a dependent noise process could, in some cases, be an artificial result of the construction of the equidistant series in physical time. In fact, the time deformation induced by the transformation from a tick time scale to a physical one, can transform an MA(1) process into an MA(q) or ARMA(p, q). In other words, the time deformation induced by the equidistant grid construction could have the effect of spreading the mass of the first autocorrelation lag onto higher order lags[11]. This possible artificial increase of the autocorrelation order induced by the regular grid construction is, in fact, an additional important reason to favor, in the computation of the realized volatility, the use of a tick time scale instead of the commonly used regular grid in physical time.

[11]See Corsi et al. (2001) for an empirical example and Oomen (2005) for a detailed theoretical analysis.

	σ^2	η^2	std(σ^2)	std(η^2)
True values & Cramer-Rao bounds	1	4	0.0951	0.1698
MS-DST	0.9964	4.0045	0.0957	0.2036
MS-DST + NR	0.9982	4.0001	0.0939	0.1685

Table 2.2: Evaluation of the absolute efficiency of the Multi-Scales DST estimator and the Multi-Scales DST estimator + Newton-Raphson algorithm (MS-DST + NR) with true volatility of 1, noise to signal ratio $\frac{\eta}{\sigma} = 2$, 2048 observations per day and 5,000 simulations.

In the presence of dependence in the noise process, the observed returns in tick time become an MA(q) process which can be written as

$$r_n = \sigma \tilde{\epsilon}_n + \sum_{i=1}^{q} \eta_i \left(\omega_n^{(i)} - \omega_{n-i}^{(i)} \right)$$

with $\tilde{\epsilon}_n \sim \mathrm{IID}\,(0,1)$ and $\omega_n^{(i)} \sim \mathrm{IID}\,(0,1)$. It can be shown that in this more general case the variance of the Minimal DST component is given by

$$\sigma_M^2 = \mathbb{E}\left(c_{min}^{(M)}(n)\, c_{min}^{(M)}(n) \right) = \sigma^2 + \sum_{i=1}^{q} \eta_i^2 F\,(M,i)$$

where $F\,(M,i) = \frac{2}{M+1}\left[M + 1 - (M+1-i)\cos\frac{\pi i}{M+1} - \cot\frac{\pi}{M+1}\sin\frac{\pi i}{M+1} \right]$.

Because $F\,(M,i)$ can be approximated as $F\,(M,i) = \pi^2 \frac{i^2}{M^2} - 2\pi^2 \frac{i^3}{M^3}\left(1 + \frac{3}{i} - \frac{1}{i^2}\right) + O\left(\frac{i^4}{M^4}\right)$, when $M/q \to \infty$, we obtain that $\sigma_M^2 \simeq \sigma^2 + \frac{\pi^2}{M^2}\sum_{i=1}^{q}(i\,\eta_i)^2$ which indicates that also the bias coming from higher order autocorrelations is cut down at the same rate M^2, guaranteeing the robustness of the DST estimators respect to a wide class of noise contaminations and model misspecifications.

2.5 Monte Carlo Simulations

The DGP used in the simulations is a combination of the Heston (1993) SV model for the dynamics of the true price process and the model proposed by Hasbrouck (1999) for the microstructure effects.

In the Heston model the true log price assumes the following continuous time dynamics

$$
\begin{aligned}
dp(t) &= (\mu - v(t)/2)dt + \sigma(t)dB(t) && (2.21)\\
dv(t) &= k(\alpha - v(t))dt + \gamma v(t)^{1/2}dW(t) && (2.22)
\end{aligned}
$$

where $v = \sigma^2$ and the initial value $v(0)$ is drawn from the unconditional Gamma distribution of v. The value of the parameters are the same as in Zhang et al. (2005), which in annualized terms are: $\mu = 5\%$, $k = 5$, $\alpha = 0.04$ corresponding to an expected annualized volatility of 20%, $\gamma = 0.5$ and the correlation coefficient between the two Brownian motions $\rho = -0.5$. Those parameters, who are reasonable for stocks, will be held constant throughout the simulations. The continuous time model of the true price is simulated at the usual Euler clock of one second.

To this SV model for the dynamics of the true price, we add the Hasbrouck bid-ask model for the observed price. The Hasbrouck model views the discrete bid and ask quotes as arising from the efficient price plus the quote-exposure costs (information and processing costs). Then the bid price is the efficient price less the bid cost rounded down to the next tick and the ask quote is the efficient price plus the ask cost rounded up to the next tick. As in Alizadeh et al. (2002) the model is simplified by assuming that the bid cost and the ask cost are both equal to the minimum tick size.

Then, according to the Hasbrouck model the bid and ask prices are respectively

$$B_n = \Delta \left\lfloor \tilde{P}_n/\Delta - 1 \right\rfloor \quad \text{and} \quad A_n = \Delta \left\lceil \tilde{P}_n/\Delta + 1 \right\rceil \tag{2.23}$$

where Δ represents the tick size, $\lfloor x \rfloor$ is the floor function, $\lceil x \rceil$ the ceiling one and the unobserved efficient price is $\tilde{P}_n = e^{\tilde{p}_n}$.

Hence, the observed price is given by the following bid-ask model

$$P_n = B_n q_n + A_n (1 - q_n) \tag{2.24}$$

with $q_n \sim \text{Bernoulli}\,(1/2)$. Therefore, the observed logarithmic return can be written as

$$r_n = \ln \frac{P_n}{P_{n-1}} = \ln \frac{\lceil P_n/\Delta + 1 \rceil}{\lceil P_{n-1}/\Delta + 1 \rceil} + q_n \ln \frac{\lfloor P_n/\Delta - 1 \rfloor}{\lceil P_n/\Delta + 1 \rceil} - q_{n-1} \ln \frac{\lfloor P_{n-1}/\Delta - 1 \rfloor}{\lceil P_{n-1}/\Delta + 1 \rceil} \tag{2.25}$$

which would be an MA(1) process in case the true price followed a Brownian motion. Notice that, by adopting the Heston model for the dynamics of the true price, the observed prices do not follow an MA(1) process anymore, making the DST approach formally misspecified. We choose this misspecified simulation setting expressly to show the robustness of the DST approach against more general heteroskedastic process in tick time as discussed in section **??**.

We first follow Hasbrouck and Alizadeh et al. and choose parameter values which imply a high level of the noise to signal ratio: $\Delta = 1/16$ and $P_0 = 45$. These values, together with the average annualized volatility of 20% given by the Heston model for the true price, induce an average noise to signal ratio of about 3.5[12]. Such high level of noise manifests itself as a strong price fluctuation between bid and ask quotes, which generates a highly negative first order-autocorrelation $\rho(1) \approx -48\%$ for the tick-by-tick returns r_n.

This noise to signal ratio reflects a microstructure impact on the return process which is remarkably large and rarely observed on real data. However, such an extreme setting provides a useful stress test for realized volatility measures and hardens the competition versus daily range-based estimators which are favored under these circumstances.

We simulate one-day sample paths of 6.5 hours (the typical opening time for stock markets) for 25,000 days. The simulation is repeated for two different values of the total number of price observations per day: $M = 390$ which corresponds to an average intertrade duration of one minute, and $M = 4,680$ which corresponds to an average tick arrival time of 5 seconds.

The competing estimators are:

- the two DST estimators: the Minimal DST (Min-DST) is computed with a window length of 30 ticks, while, for the Multi-Scales DST (MS-DST), we construct a series of minimal DST estimator using a sequence of M_j ranging from 2 to 20 ticks and then fit equation (2.20) through simple OLS.

- three "simple" Multi-Scales estimators: two Two-Scales estimators with frequency ratio α of 5 and 10, denoted respectively TS(5) and TS(10), and the Multi-Scales Least Square (MS-LS) estimator. The values chosen for the frequency ratio α, are just two representative values among those who seem to give the best results across the different settings considered in this analysis;

[12]Following (Oomen 2006) we define the noise to signal ratio as the standard deviation of the noise divided by the average standard deviation per tick of the true price process. This standardization has been chosen first because it seems reasonable to normalize both the noise and signal standard deviation respect to the same time interval and second because doing that at the tick-by-tick level facilitates comparison across different assets and over time, being such ratio not affected by the different market activity.

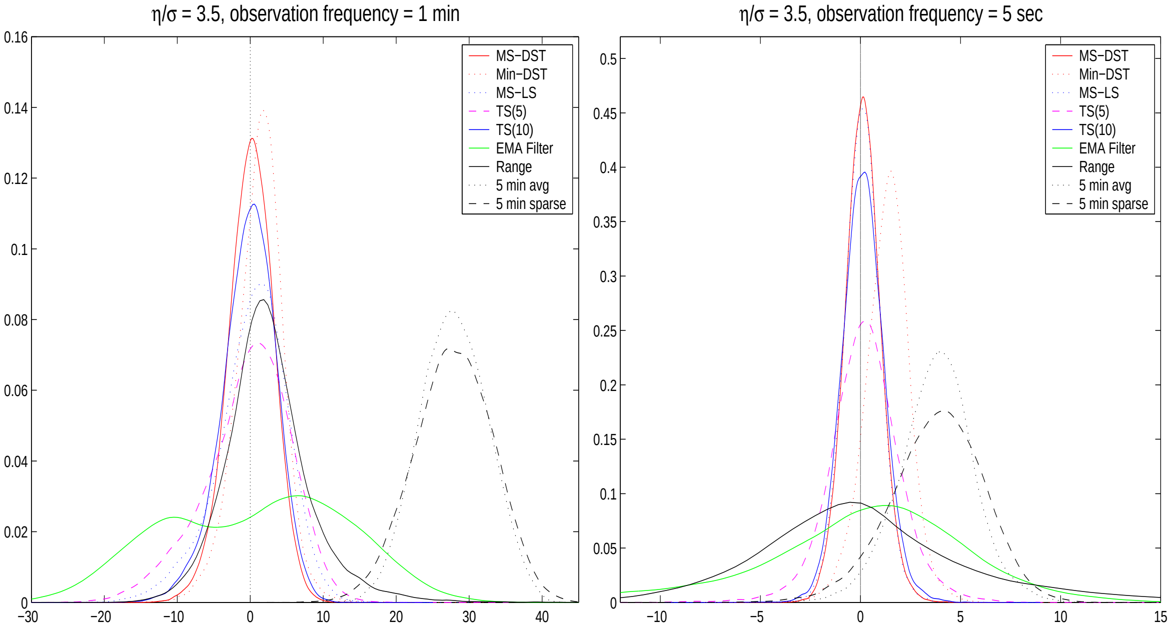


Figure 2.2: Comparison of the pdf of the estimation errors on the annualized percentage volatility (on average 20%) obtained with an average observation frequency of 1 minute (left panel) and 5 seconds (right panel) and a noise to signal ratio $\frac{\eta}{\sigma}$ =3.5.

- the local EMA filter (i.e. calibrated on a single day), which then simply corresponds to a daily MA(1) filter;
- two standard realized volatility measures both computed with 5 minute returns but one sampled with an overlapping scheme and then averaged;
- the daily range, as proposed by Parkinson (1980) and recently advocated by Alizadeh et al. (2002) in the contest of SV models estimation[13] .

We first consider the case of having 390 observations per day (corresponding to an average one minute frequency) and a noise to signal ratio of 3.5. Table 2.3 reports the mean, standard deviation and Root Mean Square Error (RMSE) of the estimation errors on the annualized volatility (express as a percentage). Figure 2.2 shows the probability density functions of those volatility estimation errors.

[13]Comparison with the recently proposed high frequency range, the so called "Realized Range-based Variance" will be deferred to future research.

VOLATILITIES ESTIMATES WITH $\frac{\eta}{\sigma}$ =3.5

	1 min			5 sec		
	mean	std	RMSE	mean	std	RMSE
MS-DST	-0.0524	3.1033	3.1037	0.0904	0.8910	0.8955
Min-DST	1.1831	3.2068	3.4181	1.2957	1.1135	1.7084
MS-LS	0.0806	4.8183	4.8190	0.0669	0.9095	0.9119
TS(5)	-0.5610	5.9293	5.9557	0.1483	1.7775	1.7837
TS(10)	-0.3715	3.7116	3.7302	0.0904	1.0410	1.0449
EMA Filter	0.3930	12.4018	12.4080	-0.0181	5.2224	5.2225
daily Range	2.2526	5.9256	6.3393	0.1492	5.9072	5.9091
5 min avg	27.7033	5.3758	28.2201	3.7356	2.3190	4.3969
5 min sparse	27.7525	4.6204	28.1345	3.7242	1.8551	4.1607

Table 2.3: The table report the mean, standard deviation and RMSE of the estimation errors on the annualized percentage volatility (on average 20%) obtained with an average observation frequency of 1 minute (left panel) and 5 seconds (right panel), and a noise to signal ratio $\frac{\eta}{\sigma} = 3.5$.

Given the high level of noise and the relatively small number of observations per day, the estimation of the first order autocorrelation required to calibrate the EMA filter, is very noisy and does not always satisfy the theoretical bound for MA(1) process $| \rho(1) | < 1/2$ (in the 30% of the cases), leading to a complex MA(1) coefficient θ. In such cases, the EMA filter would fail and we are then forced to impose an artificial floor to $\rho(1)$. But, besides its arbitrariness, this procedure induces unreasonably low volatility estimates (responsible for the left bump presents in the EMA estimator pdf on the left panel of Figure 2.2). Moreover, under these conditions, the variance of the estimator is extremely large. For the 5 minute realized volatility, the fact that the aggregation from 1 to 5 minute returns is not able to eliminate all the negative autocorrelation, makes this estimator strongly upward biased. In the case of the Minimal DST estimator instead, the aggregation works much better but, due to a relatively low window length of 30 ticks, a small upward bias is still present. Even the daily range suffers from a significant bias but it also has a much larger variance (both, the bias and the variance, are about two times those of the Minimal DST one). Among the three simple Multi-Scales estimators the TS(5) and TS(10) have small negative biases while the MS-LS is virtually unbiased. However, the variance and the RMSE of the TS(10) are lower than the other two simple multi-frequency estimators. Under this extreme setting, the only measure which is still able to remain unbiased and sufficiently precise is the MS-DST estimator, which has, in fact, the lowest RMSE. Moreover, comparing the realized volatility estimators with the one based on the daily range shows that, even in the most unfavorable setting for the realized volatilities, they remain much more accurate than the daily range: the best realized volatility estimator, the MS-DST, possesses, in fact, a RMSE 48% smaller than that of the daily range.

Keeping the same level of noise, we repeat the simulation at 5 second frequency (which means 4,680 observations per day). With twelve times more data the realized volatility measures are much more precise: the local EMA filter has less failings (5%) and lower variance, while the 5 minute realized volatilities (thanks to the longer aggregation period) have smaller, but still significant, biases. Although smaller than the biases of the 5 minute realized volatilities, the Minimal DST still shows a bias with this high level of noise. The Zhang et al. estimators become both unbiased with the TS(10) having a smaller variance than the TS(5). The MS-DST and the MS-LS estimators are both unbiased and equally very accurate, remaining the best choices among the estimators considered.

In practice, however, financial time series present a noise to signal ratio at tick-by-tick level that usually lies between 0.5 and 2. But, even with such a moderate level of noise, a naive high frequency realized volatility measure would be from one to three times the actual one. We then repeat the simulation with a more realistic noise to signal ratio of 1.5 for both observation frequencies. Table 2.4 and Figure 2.3 summarize the results.

At the 1 minute frequency the daily range and the local EMA filter are unbiased but quite inaccurate while the realized volatilities based on 5 minutes have again a large bias. The MS-DST and the other Multi-Scales estimators are the most accurate with the MS-LS having a slightly smaller bias and variance than the others.

At the 5 second frequency, with a moderate level of noise and a large number of data, the EMA filter starts to have a much lower variance and the 5 minute measures much lower biases. Nevertheless, they can still not compete with the MS-DST and the three other Multi-Scales estimators which become extremely precise and accurate under this setting.

Empirical studies on the autocorrelation of tick-by-tick data often show significant values not only for the first order but also for higher order lags (though, usually, of much smaller amplitude). A possible explanation, and way to model it, is by relaxing the assumption of i.i.d. microstructure noise by introducing a correlation in the sequence at which bid and ask

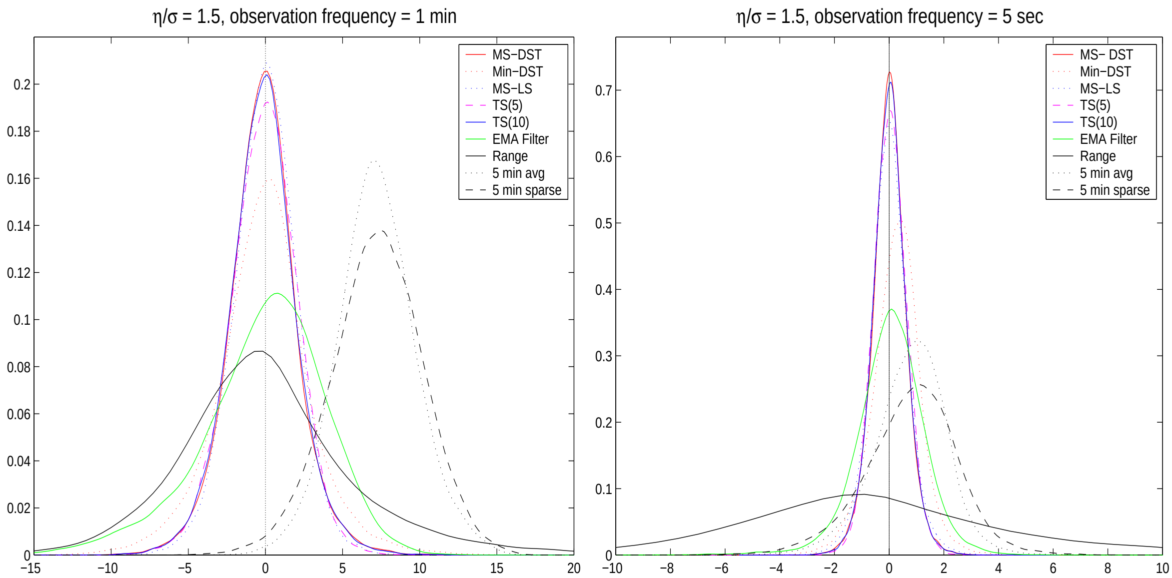


Figure 2.3: Comparison of the pdf of the estimation errors on the annualized percentage volatility (on average 20%) obtained with an average observation frequency of 1 minute (left panel) and 5 seconds (right panel) and a noise to signal ratio $\frac{\eta}{\sigma} = 1.5$.

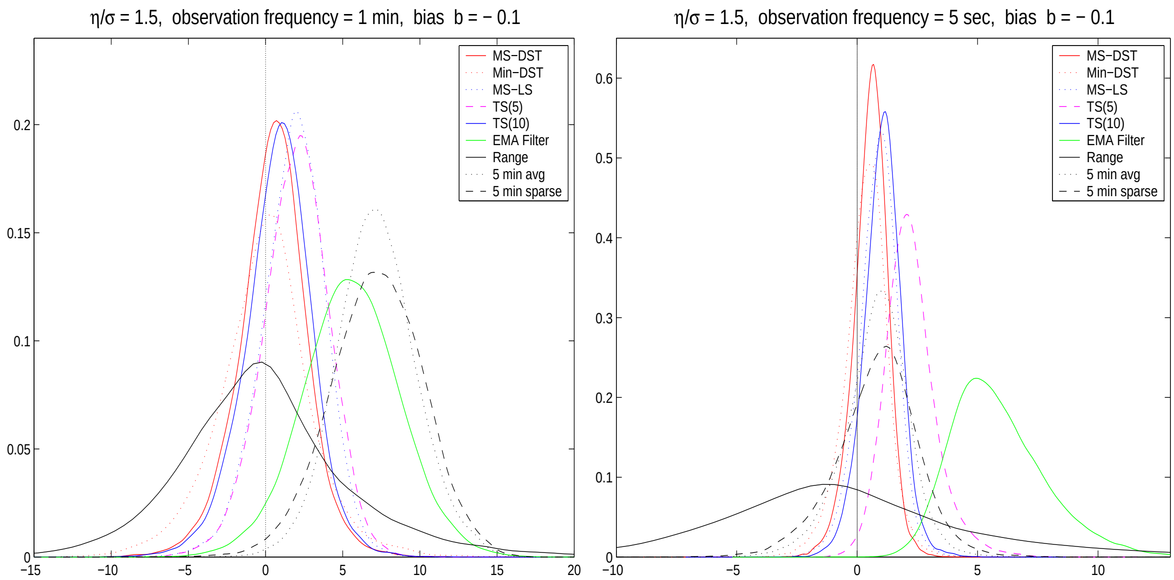


Figure 2.4: Comparison of the pdf of the estimation errors on the annualized percentage volatility (on average 20%) obtained with an average observation frequency of 1 minute (left panel) and 5 seconds (right panel), a noise to signal ratio $\frac{\eta}{\sigma} = 1.5$ and a biased Bernoulli process with bias $b = -0.1$.

VOLATILITIES ESTIMATES WITH $\frac{\eta}{\sigma} = 1.5$

Unbiased Bernoulli

	1 min			5 sec		
	mean	std	RMSE	mean	std	RMSE
MS-DST	-0.2200	2.2131	2.2240	-0.0124	0.6270	0.6271
Min-DST	-0.2263	3.0683	3.0767	0.2101	0.8992	0.9234
MS-LS	-0.0969	2.0306	2.0329	-0.0287	0.7119	0.7125
TS(5)	-0.2077	2.1551	2.1651	-0.0043	0.6128	0.6128
TS(10)	-0.2380	2.2127	2.2254	-0.0130	0.6255	0.6257
EMA Filter	-0.2969	4.0460	4.0569	-0.0308	1.2683	1.2687
daily Range	-0.0096	5.9989	5.9989	-0.4718	5.8244	5.8435
5 min avg	7.1261	2.9711	7.7206	0.7326	1.7042	1.8550
5 min sparse	7.1217	2.5285	7.5572	0.7006	1.3455	1.5170

Biased Bernoulli

	1 min			5 sec		
MS-DST	0.3894	2.2340	2.2677	0.6751	0.7170	0.9848
Min-DST	-0.1534	3.0779	3.0817	0.3293	0.9049	0.9630
MS-LS	1.7224	2.0709	2.6935	0.8317	0.8174	1.1662
TS(5)	1.8883	2.1509	2.8622	2.2000	1.0817	2.4516
TS(10)	0.7261	2.2142	2.3302	1.0239	0.7912	1.2940
EMA Filter	5.4676	3.1365	6.3034	5.8407	1.9792	6.1670
daily Range	-0.1283	5.8236	5.8250	-0.4573	5.9265	5.9441
5 min avg	7.1038	2.9780	7.7028	0.7540	1.7418	1.8980
5 min sparse	7.1053	2.5490	7.5487	0.7195	1.3707	1.5480

Table 2.4: The table reports the mean, standard deviation and RMSE of the estimation errors on the annualized percentage volatility (on average 20%) obtained with an average observation frequency of 1 minute (left panel) and 5 seconds (right panel), a noise to signal ratio $\frac{\eta}{\sigma} = 1.5$ and, for the bottom panel, a biased Bernoulli process with bias $b = -0.1$.

prices arrive[14]. Hence, instead of having an "unbiased" Bernoulli(1/2) for the q_n process, we construct a Bernoulli process which produces autocorrelation in q_n. This "biased" Bernoulli is obtained by taking $q_n = \text{Bernoulli}\,(1/2 + b)$ if $q_{n-1} = 1$ and $q_n = \text{Bernoulli}\,(1/2 - b)$ if $q_{n-1} = 0$. We choose $b = -0.10$ which induces a second order autocorrelation of about -6%.

Now, in the presence of not i.i.d. microstructure noise, the local EMA filter, which was unbiased, becomes highly biased at both frequencies (see Figure 2.4). Also the TS and MS-LS estimators are now showing a positive bias. Among all the realized volatility estimators the DST measures are the ones with the smallest bias and smallest RMSE, showing a high degree of robustness against more general microstructure noise contaminations (as analytically described in the previous section).

Summarizing the results of the simulation study, we can draw the following conclusions. The daily range estimator is always inferior to the realized volatility ones. The realized volatilities with 5 minute returns are often significantly biased and inaccurate. The local EMA filter gives satisfactory results only in the presence of a high number of observations and a low level of i.i.d. noise. Although more precise in general, similar considerations can be made for the simple MS estimators (TS and MS-LS). When the microstructure noise is moderate and i.i.d. the simple MS estimators are almost as accurate as the MS-DST and hence close to the optimal Cramer-Rao efficiency bound (since, as shown in section 2.4.5, the MS-DST is very close to the full efficiency of the Cramer-Rao bounds). In particular, the MS-LS seems to be particularly efficient in exploiting the information contained in the data when a relatively small number of observations is available (perhaps due to its ability to extract information from many frequencies), while the TS(5) and TS(10) are at their best when the number of observations increases. However, when the microstructure noise increases and deviations from the i.i.d. structure arise, the discrepancy between the simple MS estimators and the MS-DST starts to increase due to a higher level of robustness of the DST approach. Therefore, the overall winner that seems to arise from this volatility estimation "horse race" is the MS-DST which shows the highest level of precision and robustness across a wide range of microstructure noise contaminations.

[14] Hasbrouck and Ho (1987) suggest that positive autocorrelation at lag lengths greater than one may be the result of traders working an order: "a trader may distribute purchases or sales over time". However also significant negative autocorrelation at lag two are often observed.

2.6 Empirical application

To verify the behavior of volatility estimators when the microstructure noise is not an i.i.d. process, we analyze six years of tick-by-tick data (from January 1998 to October 2003) for the following three future contracts: the S&P 500 stock index future, the 30 years U.S. Treasury Bond future and the Italian stock index future FIB 30. In a base asset mapping approach (as the one of RiskMetrics), those three major future contracts can be seen as the reference liquid base assets for, respectively, the US stock and bond market and the Italian stock market.

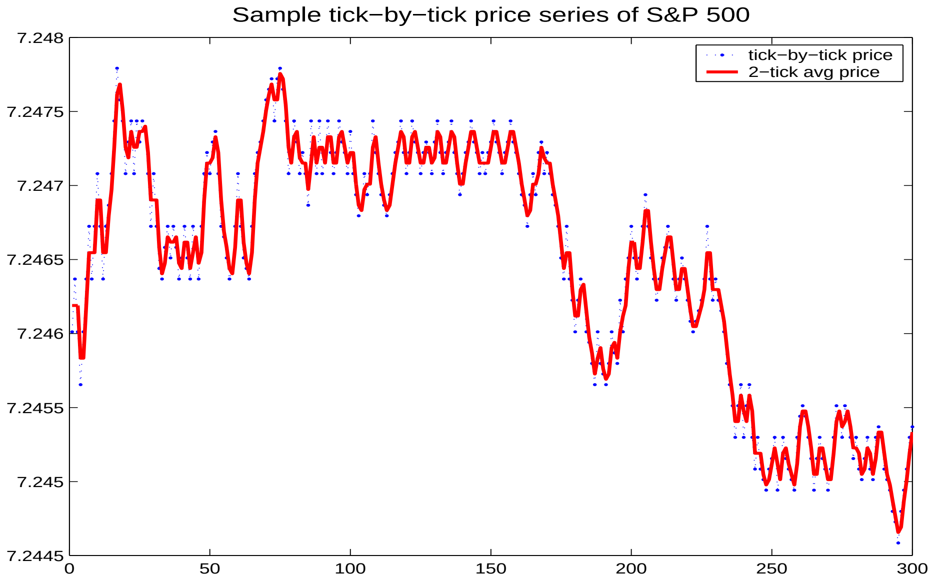
In order to analyse the dependence structure of the microstructure noise in those series, we investigate the behaviour of the autocorrelation of tick-by-tick returns. This tick-time autocorrelation analysis shows significant departure from the standard i.i.d. assumption for the microstructure noise. In fact, more complex structures than that of a simple MA(1) expected under the standard i.i.d. assumption, were found in all the three series. These patterns are independent of the inclusion or censoring of all zero trade-by-trade returns which, in all the three assets, usually represent only a small percentage of the total number of trade-by-trade returns.

Such autocorrelation patterns of the tick-by-tick returns are instead consistant with more complex ARMA structure for the microstructure noise. In fact, simulating the Hasbrouck model with those ARMA structures for the noise, leads to exactly the same autocorrelation functions observed in the data. In particular, those patterns are consistent with a microstructure noise having an MA(1) structure for the FIB, an MA(2) (at least) for the S&P and a strong oscillatory AR(1) for the US bond[15] (see Figure 2.5).

To overcome the problem of a complex ARMA structure in the autocorrelation of tick-by-tick returns of the S&P and US bond, we first notice that, in both cases, a simple aggregation of two ticks returns almost restore the MA(1) autocorrelation pattern typical of the i.i.d. assumption for the microstructure noise (see Figure 2.6). Therefore, applying the MS-DST estimator to the two-ticks returns of the S&P and US bond series, we can still obtain a highly precise evaluation of the realized volatilities of the two assets and closely follow their time series dynamics (see Figure 2.7).

Obviously, in empirical analysis the true volatility is not observable, hence no direct evaluation criteria of the quality of the volatility estimators exist. However, general indirect criteria can be employed.

[15]The analysis of the market microstructure determinants and specific institutional constraints that would lead to such empirical evidences is beyond the scope of the present study.



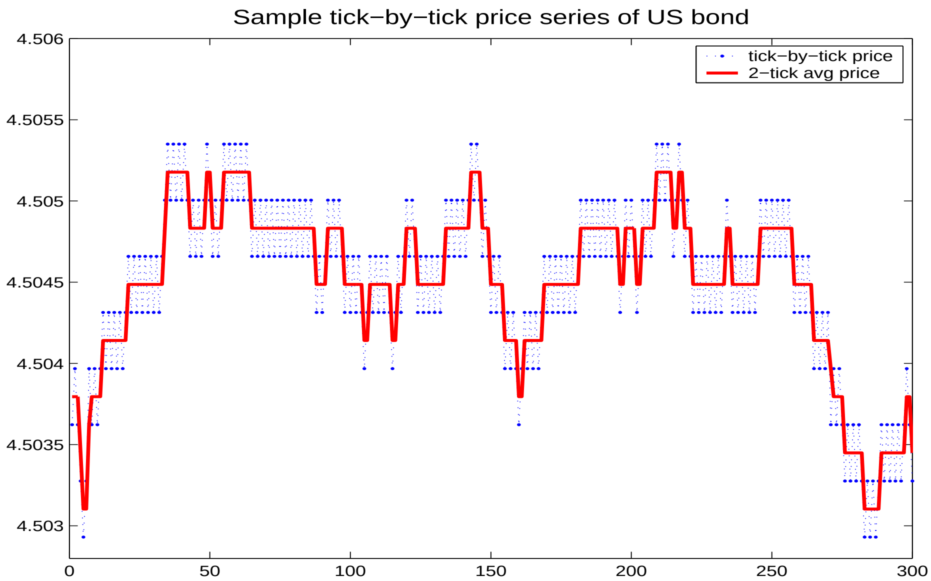


Figure 2.5: Sample path of the tick-by-tick price process (dotted line) with its two ticks moving average (solid line) for the S&P 500 (top panel) and US Bond (bottom panel) future.

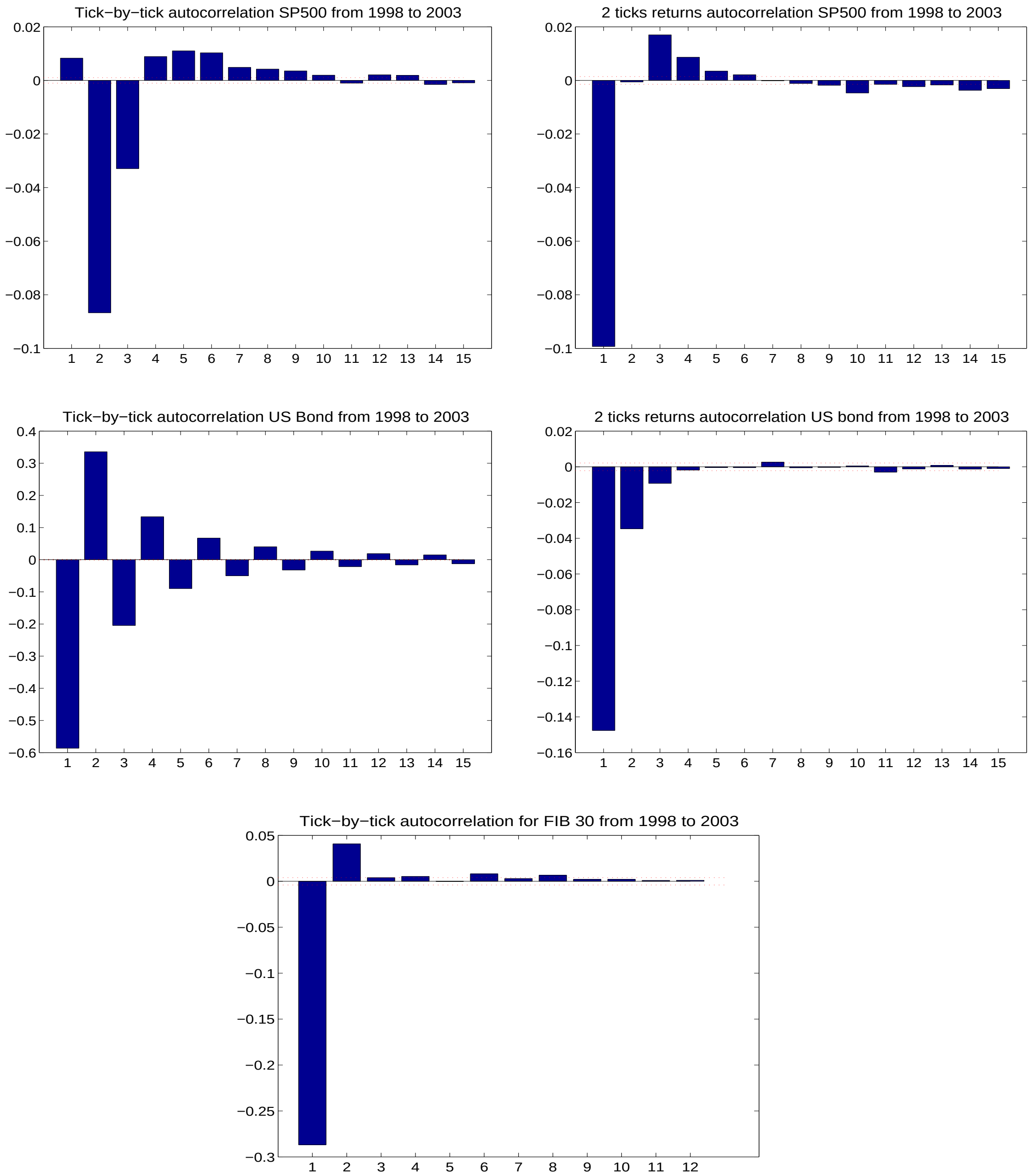


Figure 2.6: Sample autocorrelation of S&P (top) and US Bond (middle) for tick-by-tick returns (left) and 2-ticks returns (right) together with the tick-by-tick autocorrelation of FIB30 (bottom). All autocorrelation functions are computed over the six year sample, from 1998 to 2003.

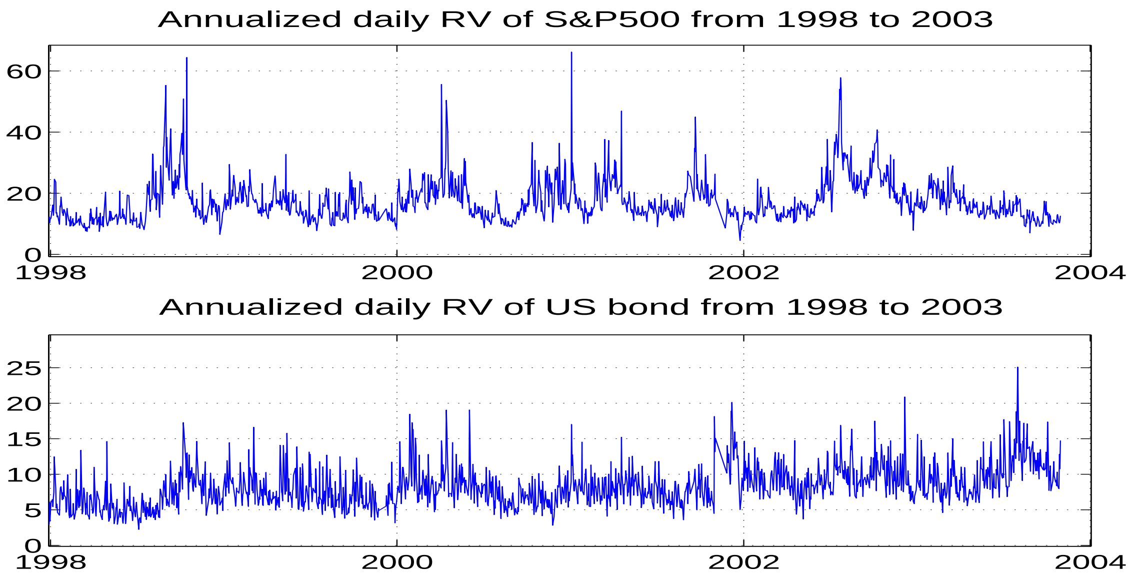


Figure 2.7: Time series of the annualize realized volatility computed with the MS-DST estimator on the two ticks returns for the S&P 500 (top panel) and US Bond (bottom panel) futures from 1998 to 2003.

First of all, the unconditional mean of daily volatilities obtained with high frequency estimators should not be significantly different from the unconditional mean volatility obtained with lower frequency returns. We asses this property for the MS-DST estimator by computing its volatility signature plot. Figure 2.8 shows the volatility signature plot of the standard and MS-DST realized volatility measures for the three assets, averaged over the whole six years period. Ideally, for estimators which are robust against microstructure effects the scaling should appear as a flat line in the volatility signature plot. The top panel of Figure 2.8 refers to the scaling of S&P 500 future showing a moderate but clear impact of the market microstructure on the standard realized volatility measure and the presence of a mild lower frequency autocorrelation. The MS-DST estimator correctly discounts market microstructure effects on volatility while it retains the residual lower frequency autocorrelation which is responsible for its scaling behavior to be not completely flat. The middle and bottom panels are respectively the FIB 30 and U.S. Bond future. In both cases market microstructure noise has a strong impact on the standard measure of realized volatility inducing larger biases as the frequency increases while the MS-DST estimator, remaining reasonably flat at any frequency, confirms its ability to properly filter out market microstructure effects.

Under the hypothesis of an underlying continuous time diffusion process for the logarithm

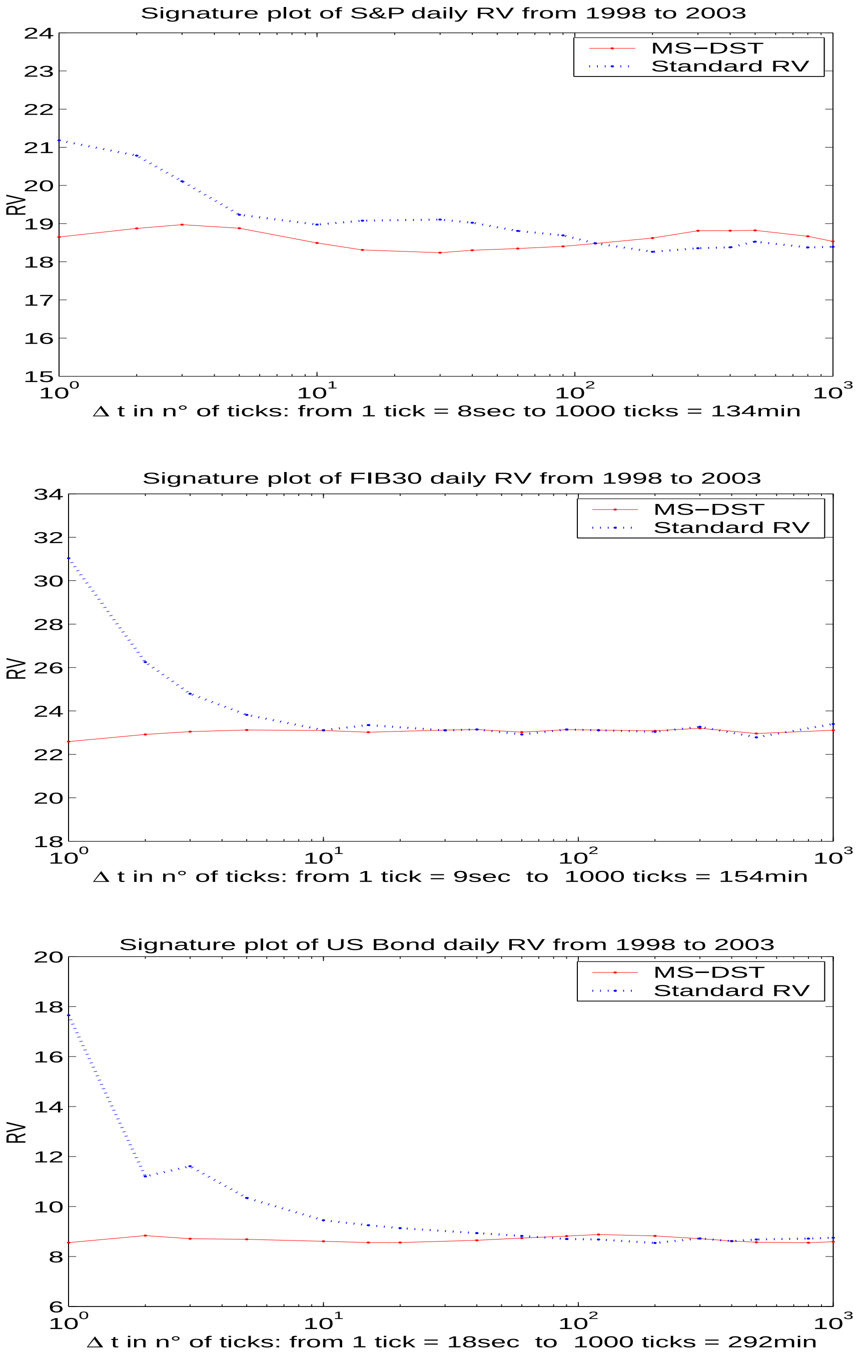


Figure 2.8: Signature Plot in semilog scale of the standard (dotted) and DST (solid) realized volatility for S&P 500, FIB 30 and US Bond from January 1998 to October 2003.

price, another indirect criterion can be considered to asses the quality of realized volatility measures in empirical applications. In fact, if the log-price follows a SV diffusion the model for daily returns could be written as $r_t = \sigma_t z_t$ where $z_t \sim$ i.i.d. $N(0,1)$. Hence, the 1-day return would be conditionally Gaussian with variance equal to the integrated variance. The normality of z_t is justified by appealing to the Central Limit Theorem for mixing process aggregated over a reasonable length of time (such as daily for highly traded assets). Therefore, if a volatility measure adequately estimates the integrated volatility, the corresponding standardized returns should be normally distributed. We test this condition using the Jarque-Bera normality test on returns standardized by the 30 minute realized volatility[16], MS-DST realized volatility and the daily range[17].

Table 2.5 reports the results. In all three cases, daily raw returns are highly leptocurktic as expected, while returns standardized by daily ranges become highly thin tailed and remain far from normal. Returns standardized by 30 minute realized volatilities become excessively thin tailed for the S&P and US Bond while remaining too fat tailed for the FIB 30 and clearly failing the Jarque-Bera tests in all the three cases. Whereas for the MS-DST standardized returns, Jarque-Bera test cannot be rejected for both the stock index future S&P and FIB. However, for the US Bond future, even though among the three competing estimators the MS-DST standardized returns remain by far the closest to the standard normal, the Jarque-Bera test is rejected. The rejection is due to a value of the kurtosis excessively smaller than three, meaning that the MS-DST measure tends to overestimate the "true" integrated volatility of the Bond future process. However, since the realized volatility consistently estimates the quadratic variation (which includes the contributions of jumps) and not the integrated volatility (which only considers the contribution of the continuous part), such overestimation could be due to the presence of a large jump components in the Bond future series. Indeed, the fact that the relative contribution of jumps is higher in bond series compared to stock indices, has been recently found by Andersen Bollerslev and Diebold (2003) and is consistent with the

---

[16]This choice of a somewhat lower frequency of 30 minutes instead of an higher one, is motivated by the need of having an unbiased estimator of the daily volatility.

[17]The EMA filter estimator has not been included here because, as shown in the simulations, it is sensitive to the presence of significant higher order autocorrelation in the tick-by-tick returns which results to be significantly different from zero in all the three series considered here. While, on this kind of data, the simple MS estimators give results that under these weak empirical tests are almost indistinguishable from the ones of the MS-DST, thus confirming the results of the Monte Carlo simulations where the MS-DST and the simple MS estimators were all very close when the level of noise were moderate and the number of observations relatively high.

	Std. Dev	Kurtosis	Skewness	Jarque-Bera	Probability
S&P 500					
Raw returns	19.3959	6.5403	-0.0106	734.8280	0.0000
MS-DST-std. returns	1.0249	2.7642	0.0190	3.3448	0.1878
30 min-std. returns	1.0663	2.4325	-0.0266	19.0443	0.0001
Range-std. returns	0.9223	1.7542	-0.0372	91.3168	0.0000
FIB 30					
Raw returns	24.4786	6.8536	-0.2368	829.1147	0.0000
MS-DST-std. returns	1.1069	2.8901	0.1464	5.3702	0.0682
30 min-std. returns	1.5778	4.1715	-0.0365	75.7205	0.0000
Range-std. returns	0.9120	1.7520	0.0493	86.1375	0.0000
US Bond					
Raw returns	8.6578	4.1011	-0.4231	113.0565	0.0000
MS-DST-std. returns	0.9664	2.5179	-0.1101	16.4640	0.0003
30 min-std. returns	1.0001	2.2831	-0.0941	32.2116	0.0000
Range-std. returns	0.8877	1.7664	-0.0948	91.3266	0.0000

Table 2.5: Comparison of sample distribution properties of daily raw and standardized returns of FIB 30, S&P 500 and thirty years Bond futures from 1998 to 2003. Standardized returns are computed using MS-DST, 30 minute realized volatility and daily range.

empirical evidence of the fixed income market being the most responsive to macroeconomic news announcements (Andersen, Bollerslev, Diebold and Vega 2003).

In summary, the analysis conducted on the empirical data confirms the ability of the MS-DST estimators to accurately and reliably estimate daily realized volatility, thus confirming the results obtained in the Monte Carlo simulation analysis.

2.7 Conclusions

The autocorrelation induced by microstructure effects represents a challenging problem for realized volatility measures. It makes the naive realized volatility computed at short time intervals highly biased, while filters based on first order covariance correction become prone to misspecification (due to the frequent significance of higher order lags) and to the unpleasant possibility of becoming negative.

In this study new realized volatility measures based on Multi-Scale regression and *Discrete Sine Transform* (DST) approaches are presented. We show that Multi-Scales estimators similar to that recently proposed by Zhang (2004) can be constructed within a simple regression based approach by exploiting the linear relation existing between the market microstructure bias and the realized volatilities computed at different frequencies. These regression based estimators can be further improved and robustified by using the DST approach to filter market microstructure noise. This approach is justified by the theoretical result regarding the ability of the DST to diagonalize exactly an MA(1) process and approximately an MA(q) one. Hence, we utilize the DST orthonormal basis decomposition to optimally disentangle the underlying efficient price signal from the time-varying nuisance component contained in tick-by-tick return series. The robustness of the DST approach with respect to more general dependent structures of the microstructure noise is also analytically shown.

The combination of such a Multi-Scale regression approach with the DST gives us a Multi-Scales DST realized volatility estimator which is then robust against a wide class of noise contaminations and model misspecifications. Thanks to the DST orthogonalization, which also allows us to analytically derive closed form expressions for the Cramer-Rao bounds of MA(1) processes, an evaluation of the absolute efficiency of volatility estimators under the i.i.d. noise assumption becomes available, indicating that the Multi-Scales DST estimator possesses a finite sample variance very close to the optimal Cramer-Rao bounds.

Monte Carlo simulations based on a realistic model for microstructure effects and volatility dynamics, show the superiority of MS-DST estimators compared to alternative local volatility proxies such as the TS and MS-LS estimators, the daily range, the EMA filter and 5 minute realized volatilities. The MS-DST estimator results to be the most accurate and robust across a wide range of noise to signal ratios and types of microstructure noise contaminations. The empirical analysis based on six years of tick-by-tick data for S&P 500 index-future, FIB 30, and 30 years U.S. Tresaury Bond future, seems to confirm Monte Carlo results.

Both empirical and simulation results show that the DST approach is a powerful non-parametric method able to cope with very general and time-varying microstructure noise, providing robust and accurate volatility estimates under a wide set of realistic conditions. Moreover, its computational efficiency, makes it well suitable for real-time analysis of tick-by-tick data.

Chapter 3

Modelling Realized Volatility

3.1 Introduction

The main purpose of this chapter is to obtain a conditional volatility model based on realized volatility which is able to account for all the main empirical features observed in the data and, at the same time, which remains parsimonious and easy to estimate.

Inspired by the Heterogeneous Market Hypothesis (Müller et al. 1993, Dacorogna et al. 2001) which led to the HARCH model of Müller et al. (1997) and Dacorogna et al. (1998) and by the asymmetric propagation of volatility between long and short time horizons, we propose an additive cascade model of different volatility components each of which generated by the actions of different types of market participants. This additive volatility cascade leads to a simple AR-type model in the realized volatility with the feature of considering volatilities realized over different time horizons. We thus term this model, Heterogeneous Autoregressive model of Realized Volatility (HAR-RV). Surprisingly, in spite of its simplicity and the fact that it does not formally belong to the class of long memory models, the HAR-RV model is able to reproduce the same memory persistence observed in volatility as well as many of the other main stylized facts of financial data.

3.2 Stylized Facts of Financial Data

Summarizing the vast literature on the empirical analysis of financial markets, the main characteristics of financial data are:

1. Fat tails: the kurtosis of the returns is much higher than that of a normal distribution at intraday frequency and tends to decrease as the return length increases. Thus return

pdf are leptokurtic with shapes depending on the time scale and presenting a very slow convergence of the Central Limit Theorem to the normal distribution. For instance, for USD/CHF the kurtosis of hourly returns is 15.58, for daily return is 4.72 and 3.78 at the weekly horizons.

2. Long memory in the volatility: although the autocorrelation of the returns is insignificant at all scales, the autocorrelation of the square and absolute returns shows very strong persistence which lasts for long time interval. This persistence reflects on the hyperbolic autocorrelation of realized volatilities from which the long memory of the process becomes even more evident. For example, the autocorrelation of USD/CHF realized volatility remain very significant for at least 6 months. This result holds true for realized volatilities aggregated at all frequencies (hourly, daily, weekly and monthly).

3. Distributional properties of realized volatility: the unconditional distributions of realized variances possess high level of skewness and kurtosis which decrease with temporal aggregation but remain far from normal even at monthly scale. Realized volatility and logarithmic realized volatility are instead much closer to normal distributions.

4. Scaling and multiscaling: computing the power spectrum of logarithmic prices often result in an approximated straight line in the relative log-log plot which is consistent with the power law behavior of scaling processes. Moreover, structure function analysis also shows strong evidences of multifractality in financial data (as we will later discuss).

3.3 Some desired properties of a volatility model

Standard GARCH and SV models are not able to reproduce the features described above. Observed data contains noticeable fluctuations in the size of price changes at all time scales while standard GARCH and SV short memory models appear like white noise once aggregated over longer time periods. For instance, in the GARCH(1,1) models there is a stringent trade-off between the possibility of having sharp changes in the short term volatility (high value of the parameter α) and the ability to capture the long memory behavior of volatility (through high values of β). Moreover, even with high value of $\beta < 1$ GARCH models are subject to exponential decline in the autocorrelation, which is at odds with the observed hyperbolic decline observed in the data. Hence, the recent interest in long memory processes.

Long memory volatility is usually obtained by employing fractional difference operators like in the FIGARCH models of returns or ARFIMA models of realized volatility. Fractional integration achieves long memory in a parsimonious way by imposing a set of infinite-dimensional restrictions on the infinite variable lags. Those restrictions are transmitted by the fractional difference operators. However fractionally integrated models also pose some problems. Fractional integration is a convenient mathematical trick but completely lacks a clear economic interpretation. The use of the fractional difference operator $(1-L)^d$ may destroy some useful information on the process and may happen to be not flexible enough to capture the real structure of the data (especially if this structure is dynamically changing in time). Fractionally integrated models are often non trivial to estimate and not easily extendible to multivariate processes. These shortcomings are evident in the FIGARCH case. But also for ARFIMA models it has been shown that the heuristic method of estimating d separately (via a Geweke, Porter-Hudak method, for instance), gives notably biased and inefficient estimates especially in the presence of large AR or MA roots (which seems to be our case). Joint ML estimation of all the parameters in ARFIMA(p,d,q) models, would then be necessary, making the estimation procedure more complex and even more difficult to extend to the multivariate case. Moreover, the application of the fractional difference operator requires a very long build up period which results in a loss of many observations. Finally, these kind of models are able to reproduce only the unifractal (or monofractal) type of scaling but not the empirical multifractal behavior found in many recent works.

Formally, a random process $X(t)$ is said to be *fractal* or self-similar if it satisfies the following scaling rule:

$$\mathbb{E}[\mid X(t) \mid^q] = c\, t^{\,\zeta(q)} \tag{3.1}$$

where c is a constant, $q > 0$ is the order of the moment and $\zeta(q)$ the scaling function or structure function exponent which is linked to the Hölder exponent[1] simply by $H(q) = \frac{\zeta(q)}{q}$. For *unifractal* processes $\zeta(q)$ is linear and then fully determined by its unique parameter $H(q) = H$, hence the terminology unifractal or monofractal. *Multifractal* processes, on the contrary are characterized by continuously changing $H(q)$ and this leads to a nonlinear (concave) $\zeta(q)$ function.

In practice, the scaling function $\zeta(q)$ of financial data is estimated by studying the scaling behavior of the moments of returns computed at different scale, the so called empirical structure (or partition) function $S(\Delta t, q)$. If the price process $p(t)$ is scaling, then:

$$S(\Delta t, q) = \sum_{t=1}^{int[T/\Delta t]} \mid p(t+\Delta t) - p(t) \mid^{\,q} \;\sim\; \Delta t^{\zeta(q)}. \tag{3.2}$$

[1] A generalization of the Hurst exponent.

Estimation of $\zeta(q)$ is then obtained by regressing $S(\Delta t, q)$ on Δt in log-log plots for different values of q.

Structure function analysis can then be seen as a study of "generalized" average volatilities (since only moments of order 1 and 2 are usually employed to define volatility) computed at different scales Δt. For this reason structure function has been already implicitly studied in the financial literature without explicitly referring to the structure function and scaling formalism. In fact, variability of the scaling exponent H for various powers of the returns has been already found to be a pervasive feature of financial data: Ding, Granger and Engle (1993), Lux (1996), Mills (1996), Andersen and Bollerslev (1997), Lobato and Savin (1998). Although the above authors did not refer to the concepts of multifractality in their papers, their findings identify the existence of multifractal processes in financial data. Their basic message is, therefore, the same as that of recent contributions from physicists (Schmitt, Schertzer and Lovejoy 1999, Vandewalle and Ausloos 1998a, Vandewalle and Ausloos 1998b, Pasquini and Serva 2000).

Formally, any additive process can be shown to have only linear $\zeta(q)$ or constant $H(q)$[2]. Hence, theoretically, only multiplicative processes can lead to multifractal behavior. It is in fact often stated, mainly by physicists, that only random multiplicative cascade models, as those encountered in turbulent flows analysis and fragmentation processes, are able to reproduce the long memory and multifractal properties found in the empirical financial data. Does this mean that we should refrain to continue to employ additive models and resign ourself to use multiplicative cascade processes which will be extremely difficult to identify and estimate?

The crucial point is that the long memory and multiscaling features observed in the data could also be only an *apparent behavior* generated from a process which is not really long memory or multiscaling. In fact, if the aggregation level is not large enough compared to the lowest frequency component of the model, truly asymptotic short memory and monoscaling models can be mistaken for long memory and multiscaling ones. In other words, the usual tests employed on the empirical data can indicate the presence of long memory and multiscaling even when none exists, just because the largest aggregation level that we are able to consider is actually not large enough. This means that the set of stochastic processes able to generate

[2]For Brownian motion for instance we have $\zeta(q) = \frac{q}{2}$ which implies $H(q) = \frac{1}{2}$. More generally for fractional Brownian motion with an order of fractional integration of d, $\zeta(q) = q(d - \frac{1}{2}) = qH$. It has been shown numerically that ARCH-GARCH process quickly converge to giving $\zeta_(q) = \frac{q}{2}$. Even in the case of more exotic Lévy flight (additive process with Lévy noise) and truncated Lévy flight the behavior of $\zeta(q)$ is still linear.

the stylized facts found in the data is much larger than commonly thought. In particular, LeBaron (2001) shows that a very simple additive model defined as the sum of only three different linear processes (AR(1) processes) each operating on a different time frames can display hyperbolic decaying memory and multiscaling, provided that the longest component has a half life that is long relative to the tested aggregation ranges. The appearance of long memory as a combination of short memory processes is not surprising given the result of Granger (1980) which shows that the sum of an infinite number of short memory processes can give rise to long memory. However, what is surprising is that those results can be obtained with only three different time scales.

As a result, it would be empirically impossible to statistically discern between true multiplicative processes and simple additive models with more than one (but far from infinite) time scales. Since it would be desirable to have a volatility model which, in addition to replicate the main stylized facts, is also simple to estimate and possibly possesses a clear economic justification and interpretation, it seems reasonable to go in the direction of simple additive volatility models with a small number of components rather than in that of complicated multiplicative systems.

3.4 Background ideas

In the light of the above considerations, we will propose a multi-component volatility model with an additive hierarchical structure which will lead to a very simple additive time series model of the realized volatility.

The basic idea stems from the so called *"Heterogeneous Market Hypothesis"*presented by Müller et al. 1993 , which recognizes the presence of heterogeneity in the traders. This view of financial markets can be easily related with the "Fractal Market Hypothesis" of Peters (1994) and the "Interacting Agent View" of Lux and Marchesi (1999). The idea of a presence of multiple components in the volatility process has been also suggested by Andersen and Bollerslev (1997) in their "mixture of distribution" hypothesis. Yet in this latter view the multi-component structure stems from the heterogeneous nature of the information arrivals rather than from the heterogeneity of the agents.

The Heterogeneous Market Hypothesis tries to explain the empirical observation of a strong positive correlation between volatility and market presence. In fact, in a homogeneous market framework where all the participants are identical, the more agents are presents, the faster the price should converge to its real market value on which all agents agreed. Thus, the volatility should be negatively correlated with market presence and activity. On the contrary, in an heterogeneous market, different actors are likely to settle for different prices and decide to execute their transactions in different market situations, hence they create volatility.

The heterogeneity of the agents may arise from various reasons: differences in the endowments, degree of information, prior beliefs, institutional constraints, temporal horizons, geographical locations, risk profiles and so on. Here we concentrate on the heterogeneity which originates from the difference in the time horizons. Typically a financial market is composed by participants having a large spectrum of dealing frequency. On one side of the dealing spectrum we have dealers, market makers and intraday speculator, with very high intraday frequency, on the other side there are central banks, commercial organization and, for example, pension fund investors with their currency hedging. Each such participant has different reaction times to news, related to his time horizon and characteristic dealing frequency. The basic idea is that agents with different time horizons perceive, react and cause different types of volatility components. Simplifying a bit, we can identify three primary volatility components: the short-term with daily or higher dealing frequency, the medium-term typically made of portfolio manager who rebalance their positions weekly, and the long-term with a characteristic time of one or more months.

Although this categorization finds its justification in the simple observation of financial markets and has a clear and appealing economic interpretation, it has been mainly overlooked in financial modelling. A noteworthy exception is the HARCH model of Müller et al. (1997) and Dacorogna et al. (1998). The HARCH process belongs to the wide ARCH family but differs from all other ARCH-type processes in the unique property of considering squared returns aggregated over different intervals. The equation of the latent variance is then a linear combination of the squared returns aggregated over different time horizons. The heterogeneous set of return interval sizes leads to the name HARCH for "Heterogeneous interval ARCH" (but the first "H" may also stand for "Heterogeneous market"). Because of the long memory of volatility, the HARCH process in its initial formulation requires a large number of returns measured at different frequency, making the log-likelihood optimization very difficult and computationally demanding. To overcome these problems Dacorogna et al. (1998) propose a new formulation of the HARCH process in terms of exponential moving averages (EMA): the EMA-HARCH process. The idea is to keep in the variance equation only a handful of representative interval sizes instead of having all of them, and replace the influence of the neighboring interval sizes by an exponential moving average of the few representative returns. This introduces a sort of GARCH-type elements in the HARCH process. In fact, broadly speaking, the variance equation of the EMA-HARCH process can be seen as a combination of several IGARCH processes defined over square returns aggregated at different frequencies.

Each IGARCH component can be regarded as the contribution of the corresponding market component to the the total market volatility and is hence termed *partial volatility* .

Studying the interrelations of volatility measured over different time horizons, permits to reveal the dynamics of the different market components. It has been recently observed that volatility over longer time intervals has stronger influence on those over shorter time intervals than conversely. This asymmetric behavior of the volatility has been found with different statistical tools. Müller et al. (1997) employ a lead lag correlation analysis of "fine" and "coarse" volatility to investigate causal relation in the sense of Granger, while Arneodo, Muzy and Sornette (1998) and Gencay and Selcuk (2004) perform a wavelets analysis. Recently, Zumbach and Lynch (2001) clearly visualize the asymmetric propagation of volatility by plotting the level of the correlation between the volatility first difference and the realized volatility for a grid of many different frequencies. These correlations measure the response function (in terms of induced volatility) of a given market component to changes of volatilities at various time scales.

The overall pattern that emerges is a volatility cascade from low frequencies to high frequencies. This can be economically explained by noticing that for short-term traders the level of long term volatility matters because it determines the expected future size of trends and risk. Then, on the one hand, short term traders react to changes in long term volatility by revising their trading behavior and so causing short term volatility. On the other hand, the level of short-term volatility does not affect the trading strategies of long-term traders. This hierarchical structure has induced some authors to propose formal analogy between FX dynamics and the motion of turbulent fluid where a energy cascade from large to small spatial scales is present. Then, borrowing from the Kolmogorov model of hydrodynamic turbulence, multiplicative cascade processes for volatility have been proposed (Ghashaghaie et al. 1999, Muzy et al. 2000 and Breymann et al. 2000). Although these types of models are able in theory to reproduce the main features of the financial data, their empirical estimation still remains an open question. Moreover, Kolmogorov model refers to the so called homogeneous cascade where the energy is homogeneously dissipated over an infinite number of scales; while in financial markets, only a limited number of scales (corresponding to the predominant components of the market) are the carriers of the financial turbulence (Lynch 2000).

Motivated by previous consideration on the ability of simple additive stochastic models to replicate equally well in practice the empirical behavior of the data, and from the observation that heterogeneous market structure generate an heterogeneous cascade with only few relevant time scales, we propose a stochastic additive cascade model of the volatility with three components.

3.5 The HAR-RV model

Defining the *latent partial volatility* $\tilde{\sigma}_t^{(\cdot)}$ as the volatility generated by a certain market component, the proposed model can be described as an additive cascade of partial volatilities, each of them having an "almost AR(1) structure"[3]. We assume a hierarchical process where at each level of the cascade the future partial volatility depends on the past volatility experienced at that time scale (the "AR(1)" component) and on the partial volatility at the next higher level of the cascade i.e. the next longer horizons volatility (the hierarchical component). To simplify, we consider a hierarchical model with only 3 volatility components corresponding to time horizons of one day ($1d$), one week ($1w$) and one month ($1m$) denoted respectively $\tilde{\sigma}_t^{(d)}$, $\tilde{\sigma}_t^{(w)}$ and $\tilde{\sigma}_t^{(m)}$.

We assume that the market dynamics is completely determined by the behavior of the dealers. Hence the high frequency return process is determined by the highest frequency volatility component in the cascade (the daily one in this simplified case) with $\tilde{\sigma}_t^{(d)} = \sigma_t^{(d)}$ the daily integrated volatility. Then the return process is

$$r_t = \sigma_t^{(d)} \epsilon_t \tag{3.3}$$

with $\epsilon_t \sim NID(0,1)$.

The model for the unobserved partial volatility processes $\tilde{\sigma}_t^{(\cdot)}$ at each level of the cascade (or time scale), is assumed to be a function of the past realized volatility experienced at the same time scale and, due to the asymmetric propagation of volatility, of the expectation of the next period values of the longer term partial volatilities. For the longest time scale (monthly) only the "AR(1)" structure remains. Then the model reads:

$$\tilde{\sigma}_{t+1m}^{(m)} = c^{(m)} + \phi^{(m)} RV_t^{(m)} + \tilde{\omega}_{t+1m}^{(m)} \tag{3.4}$$

$$\tilde{\sigma}_{t+1w}^{(w)} = c^{(w)} + \phi^{(w)} RV_t^{(w)} + \gamma^{(w)} \mathbb{E}_t[\tilde{\sigma}_{t+1m}^{(m)}] + \tilde{\omega}_{t+1w}^{(w)} \tag{3.5}$$

$$\tilde{\sigma}_{t+1d}^{(d)} = c^{(d)} + \phi^{(d)} RV_t^{(d)} + \gamma^{(d)} \mathbb{E}_t[\tilde{\sigma}_{t+1w}^{(w)}] + \tilde{\omega}_{t+1d}^{(d)} \tag{3.6}$$

Where $RV_t^{(d)}$, $RV_t^{(w)}$, and $RV_t^{(m)}$ are respectively the daily, weekly and monthly (ex post) observed Realized Volatilities as previously described, while the volatility innovations $\tilde{\omega}_{t+1m}^{(m)}, \tilde{\omega}_{t+1w}^{(w)}$

[3] Since on the right hand side there won't be the lagged latent volatility itself but the corresponding realized volatility, strictly speaking the process is not a true AR(1), but the fact that the realized volatility is a close proxy for the latent one, makes this process similar to an AR(1). More formally, this model could be classified in the broad class of Hidden Markov Models.

and $\tilde{\omega}^{(d)}_{t+1d}$ are contemporaneously and serially independent zero mean nuisance variates with appropriately truncated left tail to guarantee the positivity of partial volatilities[4].

The economic interpretation is that to each volatility component in the cascade corresponds a market component which forms expectation for the next period volatility based on the observation of the current realized volatility and on the expectation for the longer horizon volatility (which is known to affect the future level of their relevant volatility).

By straightforward recursive substitutions of the partial volatilities, such cascade model can be simply written as

$$\sigma^{(d)}_{t+1d} = c + \beta^{(d)} RV^{(d)}_t + \beta^{(w)} RV^{(w)}_t + \beta^{(m)} RV^{(m)}_t + \tilde{\omega}^{(d)}_{t+1d}. \tag{3.7}$$

Equation (3.7) can be seen as a three factor stochastic volatility model, where the factors are directly the past realized volatilities viewed at different frequency. From this process for the latent volatility it is easy to derive the functional form for a time series model in terms of realized volatilities by simply noticing that, ex-post, $\sigma^{(d)}_{t+1d}$ can be written as

$$\sigma^{(d)}_{t+1d} = RV^{(d)}_{t+1d} + \omega^{(d)}_{t+1d} \tag{3.8}$$

where $\omega^{(d)}_t$ represent latent daily volatility measurement as well as estimation errors. Equation (3.8) makes clear that we are not treating realized volatility as an error-free measure of latent volatility. Here the importance of a proper treatment of microstructure effect in the computation of the realized volatility measures (as discussed in section 2.2) becomes apparent. The consistency of the realized volatility (which is directly valid for a broad class of processes) is not enough to state that $\omega^{(d)}_t$ is a mean zero error term. Unbiased estimators of latent volatilities are needed. Equation (3.8) links our ex post volatility estimate $RV^{(d)}_{t+1d}$ to the contemporaneous measure of daily latent volatility $\sigma^{(d)}_{t+1d}$. Substituting equation (3.8) in equation (3.7) and recalling that measurement errors on the dependent variable can be absorbed into the disturbance term of the regression, we obtain a very simple time series representation of the proposed cascade model:

$$RV^{(d)}_{t+1d} = c + \beta^{(d)} RV^{(d)}_t + \beta^{(w)} RV^{(w)}_t + \beta^{(m)} RV^{(m)}_t + \omega_{t+1d} \tag{3.9}$$

with $\omega_{t+1d} = \tilde{\omega}^{(d)}_{t+1d} - \omega^{(d)}_{t+1d}$.

Equation (3.9) has a simple autoregressive structure in the realized volatility. In general, denoting l and h respectively the lowest and highest frequency in the cascade, equation (3.9) is

---

[4] An alternative way to ensure positiveness of the partial volatilities would be to write the model in terms of the log of RV.

an AR($\frac{l}{h}$) model reparametrized in a parsimonious way by imposing economically meaningful restrictions. In other words, equation (3.9) is an AR-type process but with the feature of considering volatilities realized over different interval sizes; it could than be labeled as an Heterogeneous Autoregressive model for the Realized Volatility (HAR-RV).

3.6 Simulation results

In spite of its simplicity the proposed model is able to produce rich dynamics for the returns and the volatility which closely resemble the empirical ones. These dynamics are generated by the heterogeneous reaction of the different market components to a given price change which in turns affect the future size of price changes. This causes a complex process by which the market reacts to its own price history with different reaction times. Thus market volatilities feed on themselves[5].

To asses the ability of the model to replicate the main stylized facts of the empirical data, we compare the time series returns and volatilities produced by the simulation with those of twelve years of USD/CHF. In order to give the model the time to unfold its dynamics at daily level, the HAR-RV(3) process is simulated at the frequency of 2 hours ($2h$). The simulated model then reads:

$$r_t^{(2h)} = \sigma_t^{(d)} \epsilon_t \tag{3.10}$$

$$\sigma_{t+2h}^{(d)} = c + \beta^{(d)} RV_t^{(d)} + \beta^{(w)} RV_t^{(w)} + \beta^{(m)} RV_t^{(m)} + \omega_{t+2h}^{(d)}. \tag{3.11}$$

The parameters of the model ($\beta^{(\cdot)}$) are just hand made calibrated to obtain realistic results.

The analysis begins with a simple visual inspection of the two time series for the returns (figure 3.1) and the realized volatilities (figure 3.2). In both figures 3.1 and 3.2, the upper panels show the empirical data for USD/CHF from December '89 to July '01, while the lower panels display a sample realization of the simulated process for a similar period. From the visual inspection alone is difficult to discern much difference.

Figure 3.3 summarizes the character of the simulated and actual return distribution for 1,5 and 20 day interval. In these and the subsequent comparison figures, the number of observations for the real and simulated data is very different. The twelve years of USD/CHF gives 3001 daily observations, while the HAR-RV(3) process is simulated (at 2 hours frequency) for a period corresponding to approximately 600 years i.e. 150,000 daily observations.

[5]This mechanism is sometimes called "price-driven volatility" in contrast to the "event-driven volatility" consistent with the EMH and the "error-driven volatility" due to over and under reaction of the market to incoming informations.

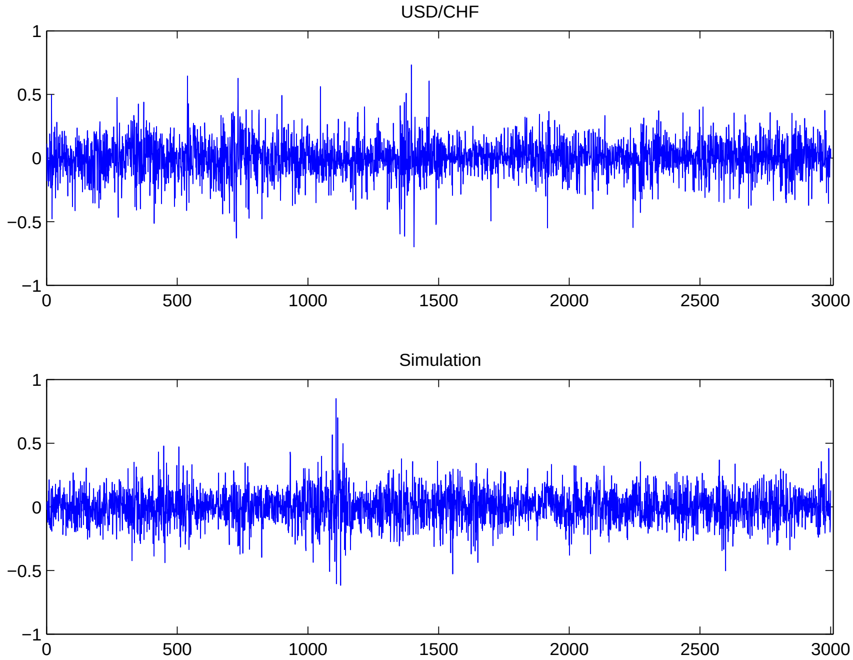


Figure 3.1: Comparison of actual (top) and simulated (bottom) daily returns series.

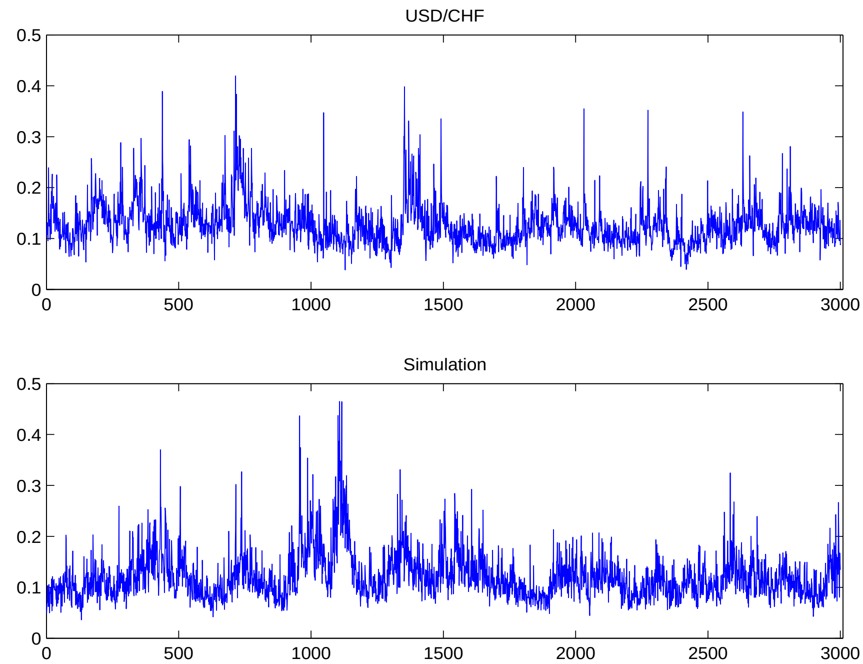


Figure 3.2: Comparison of actual (top) and simulated (bottom) daily RV series.

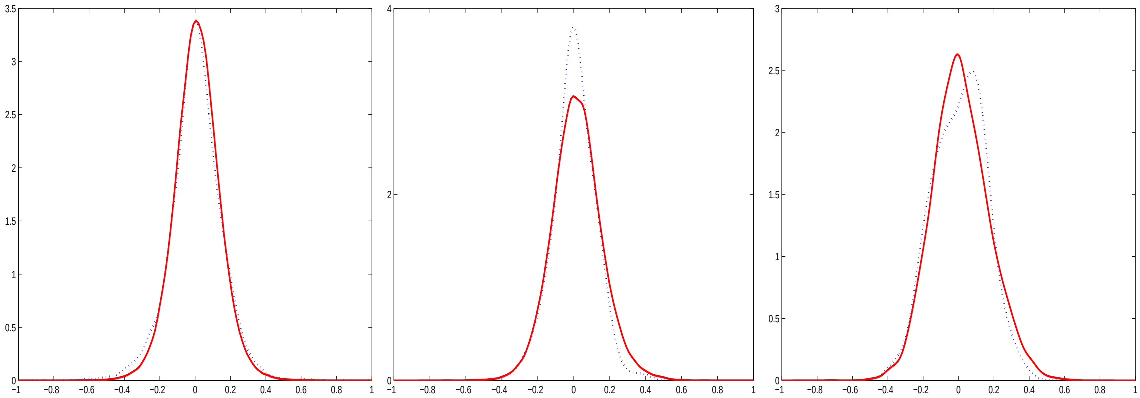


Figure 3.3: Comparison of actual (dotted) and simulated (solid) PDF of returns for different time horizons. Respectively from left to right: daily, weekly and monthly.

Kurtosis	daily returns	weekly returns	monthly returns
USD/CHF	4.72	3.78	3.04
HAR-RV(3)	4.89	3.90	3.50

Table 3.1: Comparison of actual and simulated kurtosis of returns for different time horizons.

Table 3.1 reports the values of the kurtosis of those distributions for the three aggregation interval. This Table clearly shows how the simple HAR model for the realized volatility is able to reproduce not only the excess of kurtosis of the daily returns, but also the empirical cross-over from fat tail to thin tail distributions as the aggregation interval increases.

But what we are mainly interested in, is the ability of the model to reproduce the volatility persistence of empirical data. Figure 3.4 shows the actual autocorrelation function of USD/CHF daily realized volatility together with the autocorrelation of HAR daily realized volatility simulated over a period corresponding to 600 years. This figure shows that the purpose of reproducing the long memory of empirical volatility seems to be very well fulfilled. It is important to remark that theoretically the HAR model for volatility is a short memory process which asymptotically should not exhibit hyperbolic decay of the autocorrelation. However, for the aggregation interval considered, the simulated model shows a volatility memory which is at least as long as that of actual data (actually it could be even much longer for different choices of the parameters). Also the partial autocorrelation functions show quite good agreement.

In figure 3.6 we also compare the distribution of the daily realized volatility, finding reasonable agreement between the real data and the simulated one.

Finally, we investigate the scaling behavior of the real and simulated data. In figure 3.7 the periodogram of the daily returns for the two series is plotted in a log-log plane. Again, real data cover a period of twelve years while the simulation is performed for a virtual period of more than 600 years. Both series display high degree of linearity as the one expected for true self-similar process.

3.7 Estimation and Forecast

3.7.1 The data

Our data set consists in almost 12 years (from December '89 to July 2001) of tick-by-tick logarithmic middle prices of several FX rates. Log mid prices are computed as averages of the logarithmic bid and ask quotes obtained from the Reuters FXFX screen. The whole data set amounts to millions of quotes kindly provided by Olsen&Associates. In the following univariate analysis we will concentrate on the USD/CHF exchange rate (as a proxy for the USD/EUR).

In order to avoid to explicitly model the seasonal behavior of trading activity induced by the weekend we exclude all the realized volatility taking place from Friday 21:00 GMT to Sunday 22:00 GMT. Moreover, a confounding influence comes from low trading days associated to fixed and moving holidays. Since the FX market is a world market, it is not easy to identify on the calendar the relevant holidays which affect such a global market. We then decided to use a more flexible approach by deleting those days presenting a number of ticks smaller than a certain threshold. Highly liquid rates such as USD/CHF have an average daily quotes number on the sample period of approximately 2,800. For this rate we choose a conservative threshold of 200 ticks per day. With this criteria 41 days (partially corresponding to the major US holidays) have been removed leaving us with a final sample of 3,001 full working days. The realized volatility estimates are aggregated at different scales in order to have realized volatility measures of the integrated volatility over different periods: daily, weekly and monthly.

Given that the true volatility is not observable there is no direct evaluation criteria of the quality of the volatility estimators. However, as we have already discussed in section 2.6

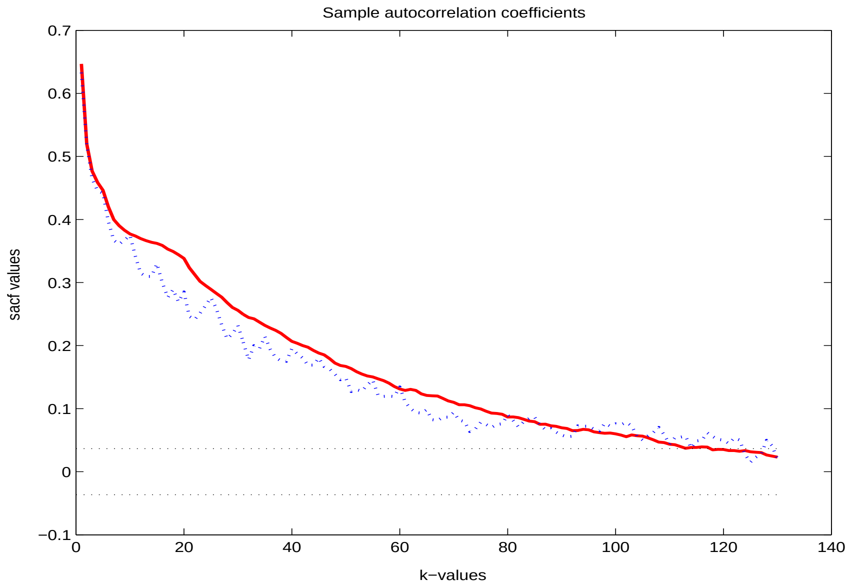


Figure 3.4: Comparison of actual (dotted) and simulated (solid) autocorrelation of daily realized volatility.

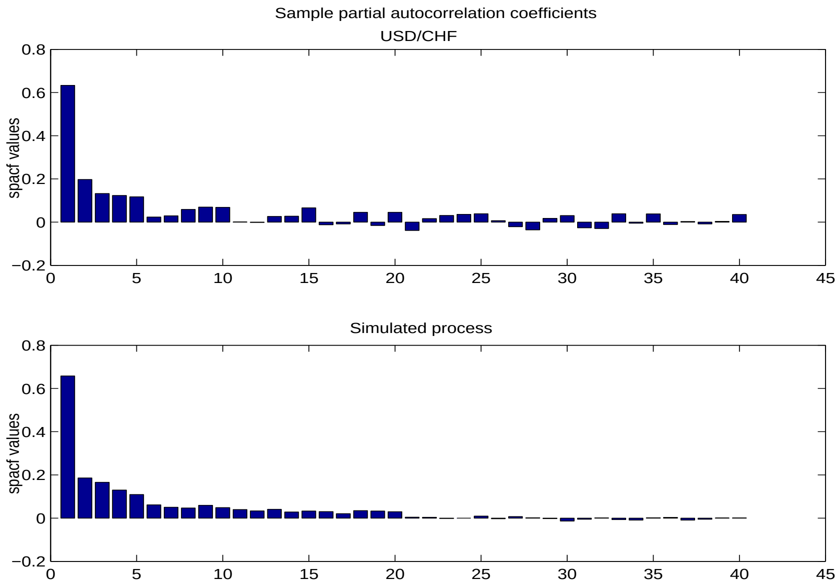


Figure 3.5: Comparison of actual (top) and simulated (bottom) partial autocorrelation of daily realized volatility.

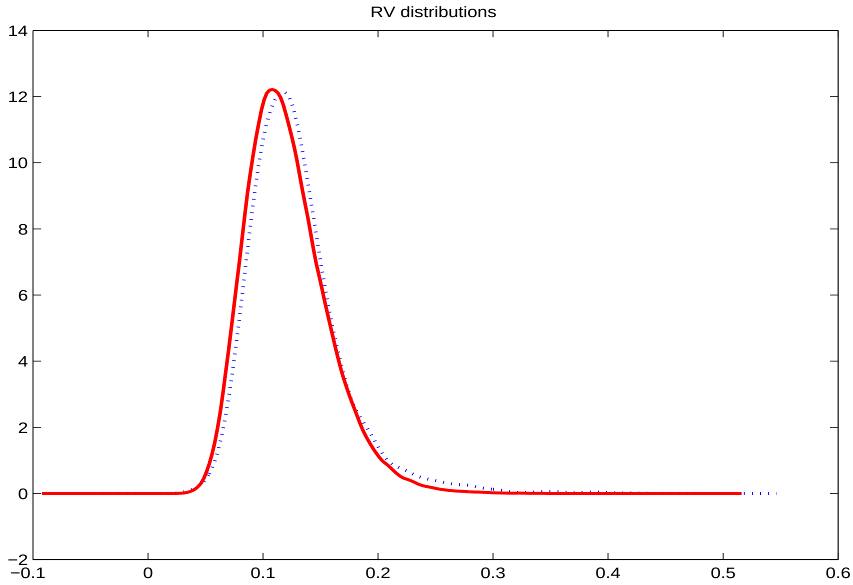


Figure 3.6: Comparison of actual (dotted) and simulated (solid) distribution of daily realized volatility.

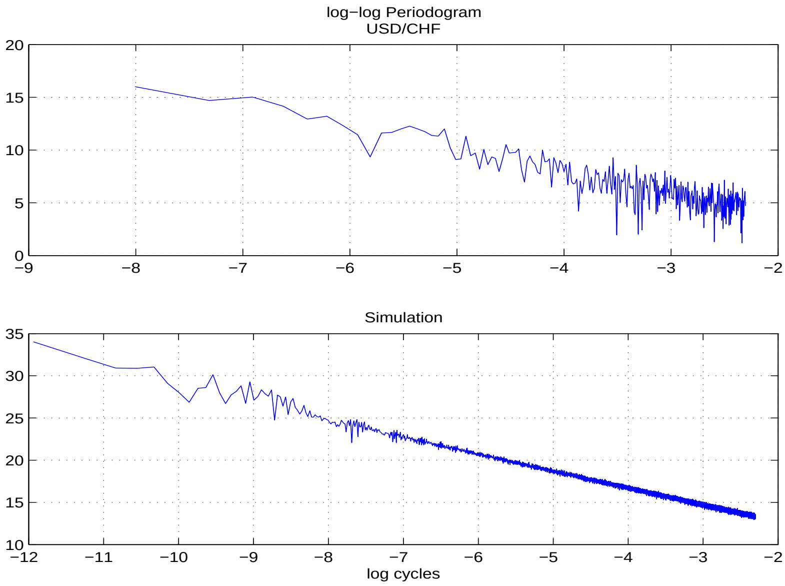


Figure 3.7: Comparison of actual (top) and simulated (bottom) periodogram of daily returns on log-log plane.

	Mean	Std. Dev	Kurtosis	Jarque-Bera	Probability
Raw returns	0.0005	0.1425	4.7262	382.89	0.0000
RV-std. returns	0.018	1.0191	2.9951	0.2464	0.8840

Table 3.2: Comparison of daily raw and RV-standardized return distributions.

general benchmark criteria can be easily constructed under the hypothesis of an underlying continuous time diffusion process for the logarithm price. In fact if the log-price follows a stochastic volatility diffusion as in (1.1) with a negligible conditional mean dynamics, the model for daily returns could be written as $r_t^{(d)} = \sigma_t^{(d)}\epsilon_t$ where $\epsilon_t \sim iid\ N(0,1)$. Hence the 1-day return is conditionally Gaussian with variance equal to the integrated variance. The normality of ϵ_t is justified by appealing to the Central Limit Theorem for mixing process to argue that the returns over a reasonable aggregation time (such as daily for highly traded assets) should tend towards normality. Then if $RV^{(d)}$ adequately estimates the integrated volatility $\sigma^{(d)}$, the RV-standardized returns should be normally distributed with a variance of unity. Table 3.2 shows that this is exactly the case for our realized volatility estimator.

3.7.2 Estimation

Following the recent literature on the realized volatility, we can consider all the terms in (3.9) as observed and then easily estimate its parameters $\beta^{(\cdot)}$ by applying simple linear regression. Standard OLS regression are consistent and normally distributed, but when multi-step ahead forecast are considered, the presence of regressors which overlap, makes the usual inference no longer appropriate. We then employ the Newey-West covariance correction with a conservative number of lags equal to 20.

Since the uses of intraday measures of realized volatility poses problems either of measurement accuracy and strong intraday seasonalities, we choose to estimate the variance equation (3.9) at daily frequency[6]. Table 3.3 reports the results of the estimation of the HAR-RV model

[6]A point of caution should be considered here. As for the GARCH, the $\beta^{(\cdot)}$ parameters of the variance equation (3.9) have a time frequency dimension; i.e. they are defined for a certain time frequency which is the frequency at which the model is estimated. The parameters will then have different values consistent with the time frequency employed and, in general, for the HAR-RV model time aggregation will tend to reduce the

for twelve years of USD/CHF daily realized volatilities. From the values of t-statistic and their associated probabilities it is very clear how all the three realized volatilities aggregated over the three different horizons are all highly significant.

It is worth noticing that if we are ready to believe that realized volatilities aggregated over different horizons are reasonable proxies for volatilities generated by the corresponding market components, an interesting byproduct of this simple OLS regression is a direct estimate of the market components weights, that is, a readily evaluation of the contribution of each market component to the overall market activity. For the USD/CHF considered here, it seems that the importance of the market components decreases with the horizon of the aggregation. Moreover, if a moving window regression is performed, a time series evolution of such weights is easily attained as well.

As we have already seen, the HAR-RV process is an autoregressive model reparametrized in a parsimonious way by imposing economically meaningful restrictions. We can then evaluate if those restrictions are valid by comparing the restricted HAR model with the unrestricted AR one. Since the HAR model considered here employs monthly realized volatility (which corresponds to 20 working days) the corresponding unrestricted autoregressive model is an AR(20). A multiple hypothesis test based on the difference between restricted and unrestricted residual sums of squares is then computed. The result of this F-test is 2.48 which is significant. Looking at the information criteria, instead, gives less clear results: on the basis of the AIC, the unrestricted AR(20) model would be slightly preferred, while on the basis of the SIC (which imposes larger penalty for additional coefficients) the HAR-RV is preferred.

3.7.3 Forecast

The in-sample 1 day ahead forecasts of the model are shown if figure 3.8 and in Table 3.4 and 3.5. These forecasts are obtained by first estimating the parameters of the models on the full sample and then performing a series of static one-step ahead forecasts. The visual impression of a quite accurate forecast shown in the top panel of figure 3.8 is confirmed by the remarkably high R^2 of the regression of 45%. From the bottom panel of figure 3.8, which displays the time series of the forecasting errors, the presence of a significant heteroskedasticity in the residuals is apparent. This observation has led Corsi, Kretschmer and Pigorsch (2005) to consider more sophisticated estimation procedures that, being able to take in to account this GARCH effect in the volatility residuals, may increases the estimation efficiency of the HAR-RV model.

---

impact of shorter realized volatilities and increases that of longer horizons.

HAR-RV(3) ESTIMATION

Included observations: 3000 after adjusting endpoints

Newey-West Standard Errors and Covariance (lag=20)

Variable	Coefficient	Std. Error	t - Statistic	Probability
C	0.017845	0.002824	6.319572	0.00000
$RV_{t-1}^{(d)}$	0.369542	0.028449	12.98979	0.00000
$RV_{t-1}^{(w)}$	0.265822	0.041865	6.349472	0.00000
$RV_{t-1}^{(m)}$	0.215011	0.037637	5.712704	0.00000

R-squared	0.45592	Mean dependent var	0.12764
Adjusted R-squared	0.45538	S.D. dependent var	0.04081
S.E. of regression	0.03012	Akaike info criter	-4.16567
Sum squared resid	2.71869	Schwarz criterion	-4.15766

Table 3.3: In sample estimation results of the least squares regression of HAR-RV(3) model for the USD/CHF exchange data from December '89 to July 2001.

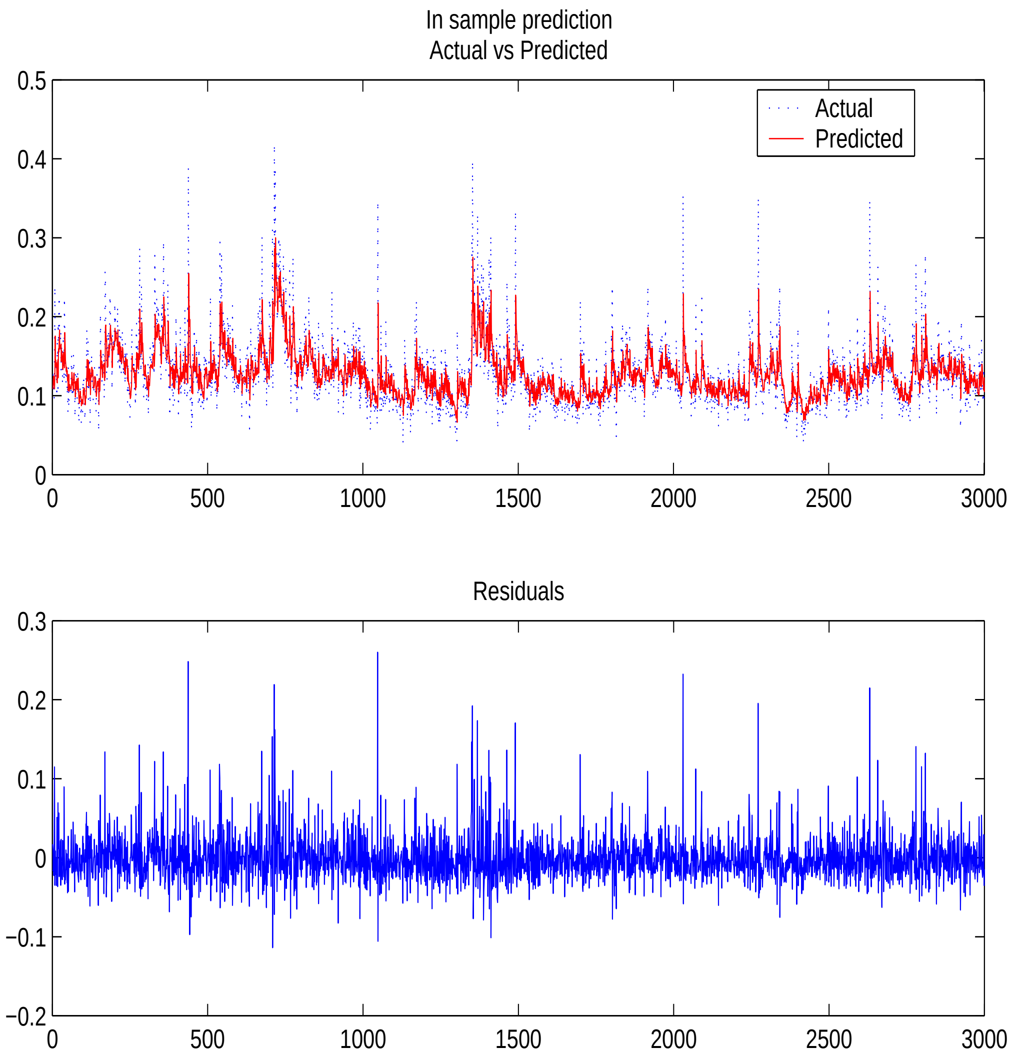


Figure 3.8: Top: comparison of actual (dotted) and in sample prediction (solid) of the HAR model for daily realized volatilities of USD/CHF exchange data from December '89 to July 2001. Bottom: residuals of the prediction.

For comparison purposes other models are added: the standard GARCH(1,1) and J.P. Morgan's RiskMetrics, together with an AR(1) and AR(3) model of the realized volatility. Moreover a fractionally integrated model for the realized volatility as employed by Andersen et al. 2002 is considered. They propose a fractional differentiation of the realized volatility series with a fractional coefficient estimated on the full sample with the GPH algorithm (which gives $d = 0.401$) followed by an AR(5) fit. Hence the model is an ARFIMA(5, 0.401, 0) estimated with a two steps procedure.

In Table 3.4 the forecasting performance are evaluated on the basis of Root Mean Square Errors (RMSE), Mean Absolut Error (MAE), Mean Absolute Percentege Error and Theil's Inequality coefficient. Following the analysis of Andersen and Bollerslev (1998) Table 3.5 reports the results of the Mincer-Zarnowitz regressions of the ex-post realized volatility on a constant and the various model forecasts based on time $t-1$ information. That is

$$RV_t^{(d)} = b_0 + b_1 \mathbb{E}_{t-1}\left[(RV_t^{(d)}) \right] + \ error. \tag{3.12}$$

In both Tables the difference in forecasting performance between the standard models and the ones based on realized volatility is evident.

But what we are mainly interested in, is to compare the models on the basis of truly out of sample forecasts. Figure 3.9 and Table 3.6 and 3.7 report the results for out of sample forecast of the realized volatility in which the models are daily reestimated on a moving window of 1000 observations[7]. An exception is made for the ARFIMA model for which the fractional difference operator requires a longer build up period equal to the cut off of its Taylor expansion. We choose the standard cut off limits of 1000 which for a value of d of 0.401 induces a cut off error of 4.2%. After fractional differentiation, the optimal length of the moving window used in the estimation of the AR parameters turns out to be of about 250 days. The forecasting performances are compared over three different time horizons: 1 day, 1 week and 2 weeks. The multi-step ahead forecasts are evaluated considering the aggregated volatility realized and predicted over the multi-period horizon. For a h steps ahead forecast the target function is then $\sum_{j=0}^{h} RV_{t+j}^{(d)}$ and the Mincer-Zarnowitz regression becomes:

$$\sum_{j=0}^{h} RV_{t+j}^{(d)} = b_0 + b_1 \mathbb{E}_{t-h}\left[\sum_{j=0}^{h} RV_{t+j}^{(d)} \right] + \ error \tag{3.13}$$

It turns out that, out of sample, the parsimonious HAR(3) model steadily outperforms the others at all the three time horizons considered (1 day, 1 week and 2 weeks). Moreover, the

[7]Hence, these results refer only to the last 2000 observations of the sample and are, therefore, not directly comparable with those in Table 3.4 and 3.5.

IN SAMPLE PERFORMANCE

	GARCH	RM	AR(1)	AR(3)	ARFIMA	HAR(3)
RMSE x 100	3.6579	3.6789	3.1600	3.0706	3.0787	3.0041
MAE x 100	2.8299	2.7742	2.1978	2.1214	2.0593	2.0577
MAPE %	25.24%	23.54%	17.87%	17.11%	15.80%	16.57%
Theil Inequality coefficient.x100	13.238	13.293	11.958	11.610	11.882	11.355

Table 3.4: Comparison of the in-sample performances of the 1 day ahead forecast of GARCH, RiskMetrics, AR(1), AR(3), ARFIMA(5,0.401,0) and HAR(3) RV models for 12 years of USD/CHF.

IN SAMPLE MINCER-ZARNOWITZ REGRESSION

	b_0	b_1	R^2
GARCH	-0.027631	1.101517	0.3055
	(-0.0361, -0.0192)	(1.0420, 1.1610)	
RM	0.032200	0.688880	0.3254
	(0.0271, 0.0373)	(0.6534, 0.7244)	
AR(1)	-0.000339	1.002469	0.4007
	(-0.0061, 0.0054)	(0.9586, 1.0464)	
AR(3)	-0.000553	1.004037	0.4341
	(-0.0059, 0.0048)	(0.9630, 1.0451)	
ARFIMA	0.005210	1.002926	0.4496
	(0.0002, 0.0102)	(0.9632, 1.0427)	
HAR(3)	-0.000861	1.006168	0.4589
	(-0.0060, 0.0043)	(0.9669, 1.0454)	

Table 3.5: In-sample Mincer-Zarnowitz regression for the GARCH, RiskMetrics, AR(1), AR(3), ARFIMA(5,0.401,0) and HAR(3) model for the 1 day ahead realized volatility of USD/CHF (95% confidence interval in parentesis).

HAR(3) model is the only one always presenting the values of 0 and 1 falling in the confidence interval of respectively b_0 and b_1 (the sufficient condition for unbiased forecasts).

It is noteworthy noticing that though the superior performance of the ARFIMA and HAR(3) were already apparent at daily horizon, it becomes striking at weekly and biweekly horizons. The reason is that the other models have a memory which is too short compared to the forecasting horizon (AR(1) and AR(3)) or they adjust too late to the movements of the realized volatility (RiskMetrics). This explanation is confirmed by figures 3.10 and 3.11 which compare the dynamic behavior of the forecasts of the different models for one week and two weeks periods ahead. For these time horizons the importance of long memory becomes manifest. What is surprising is the ability of the HAR-RV model to attain these results with only few parameters.

3.8 The Asymmetric HAR-RV model

So far no asymmetric volatility effect has been considered. It is well known, however, that equities often exhibit the so called "leverage effect", meaning that the volatility tends to increase more after a negative shock than after a positive shock of the same magnitude. In a regression style approach as the one in the HAR model, the introduction of this type of asymmetric effect simply amounts to add returns as new explanatory variables in to equation (3.9). Moreover, by extending the Heterogeneous Market Hypothesis idea to the leverage effect, we consider asymmetric responses of the realized volatility not only to previous daily returns but also to past weekly and monthly returns as a result of the activities and reactions of trading subjects operating at those frequencies. Positivity concerns on the resulting volatility would suggest to express the HAR model in terms of logarithms of the realized volatilities.

So the Asymmetric-HAR(3) model for realized volatility reads

$$\begin{aligned} \mathcal{RV}_{t+1}^{(d)} = c & + \beta^{(d)}\mathcal{RV}_t^{(d)} + \beta^{(w)}\mathcal{RV}_t^{(w)} + \beta^{(m)}\mathcal{RV}_t^{(m)} + \\ & + \gamma_-^{(d)} r_{-,t}^{(d)} + \gamma_-^{(w)} r_{-,t}^{(w)} + \gamma_-^{(m)} r_{-,t}^{(m)} + \\ & + \gamma_+^{(d)} r_{+,t}^{(d)} + \gamma_+^{(w)} r_{+,t}^{(w)} + \gamma_+^{(m)} r_{+,t}^{(m)} + \varepsilon_{t+1} \end{aligned} \tag{3.14}$$

where $\mathcal{RV}_t = \log(RV_t)$ and $r_{-,t}^{(d)}$, $r_{-,t}^{(d)}$ and $r_{-,t}^{(d)}$ are daily, weekly and monthly negative returns i.e. $r_{-,t}^{(\cdot)} = I(r_t^{(\cdot)} < 0)$, while the $r_{+,t}^{(d)}$, $r_{+,t}^{(d)}$ and $r_{+,t}^{(d)}$ are the positive ones.

We apply the Asymmetric-HAR-RV(3) model to twelve years of realized volatility series for the S&P 500 and US Bond future. These daily realized volatilities are computed with the Fitted DST estimator applied on the two-ticks return series as described in section 2.6.

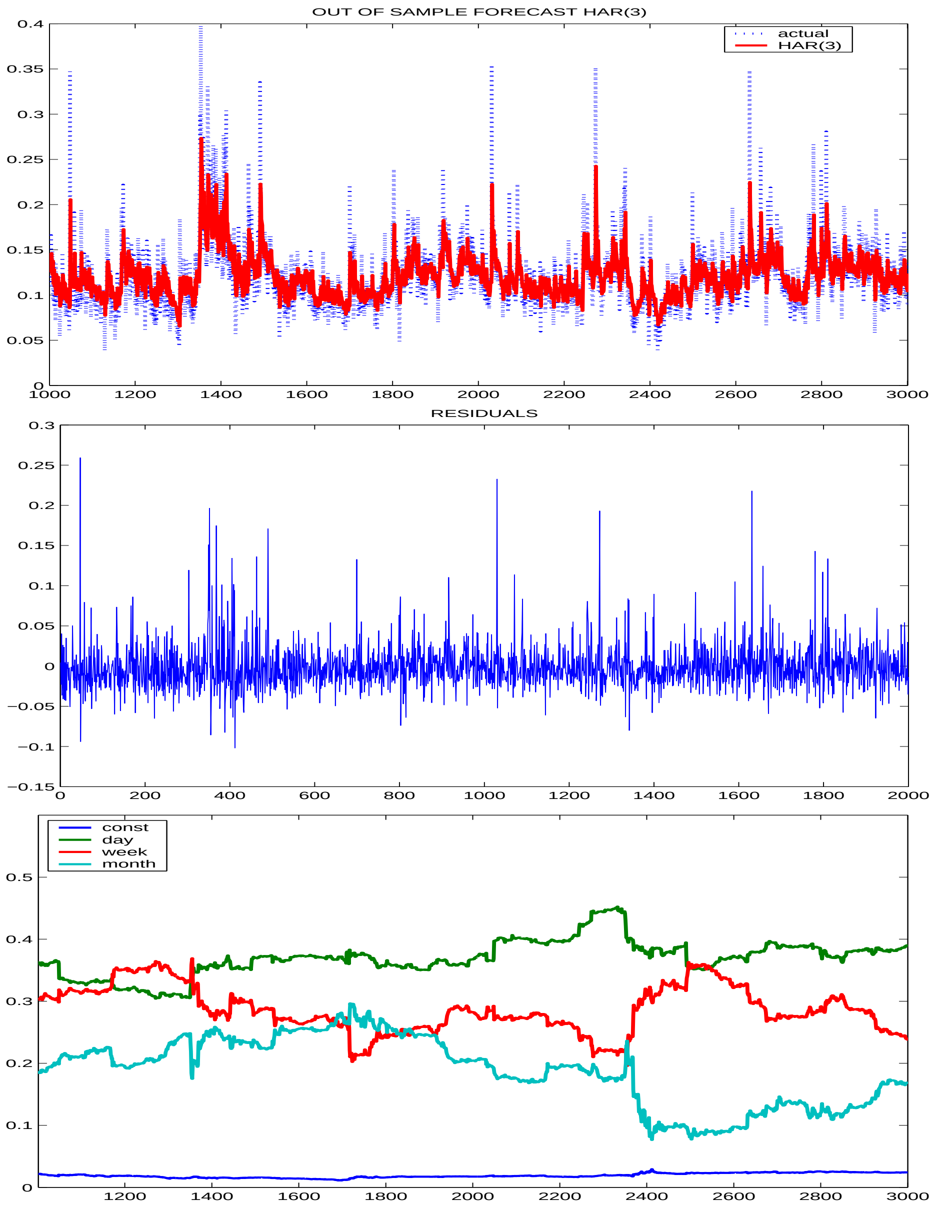


Figure 3.9: Top: comparison of actual (dotted) and out of sample prediction (solid) of the HAR(3) model for daily realized volatilities. Middle: residuals. Bottom: time evolution of the regression coeffiecients which according to the model represent market component weights.

OUT OF SAMPLE PERFORMANCE

		RM	AR(1)	AR(3)	ARFIMA	HAR(3)
1 DAY	RMSE x 100	3.5945	2.9404	2.9088	2.8916	2.8472
	MAE x 100	2.6786	2.0520	2.0061	1.9842	1.9477
	MAPE %	24.01%	17.55%	16.91%	16.90%	16.27%
	Theil Inequality coefficient.x100	13.901	11.741	11.619	11.682	11.384
1 WEEK	RMSE x 100	3.0065	2.7788	2.4372	2.7864	2.2939
	MAE x 100	2.3426	2.1324	1.8089	2.0589	1.6403
	MAPE %	22.08%	19.19%	16.074%	18.05%	14.15%
	Theil Inequality coefficient.x100	11.601	11.124	9.774	11.258	9.205
2 WEEKS	RMSE x 100	2.9734	2.8111	2.4660	2.3743	2.1713
	MAE x 100	2.4254	2.3004	1.9254	1.7728	1.6339
	MAPE %	23.28%	21.46%	17.91%	15.76%	14.47%
	Theil Inequality coefficient.x100	11.421	11.212	9.880	9.593	8.722

Table 3.6: Comparison of the out of sample performances of the RiskMetrics, AR(1), AR(3), ARFIMA(5, 0.401, 0) and HAR(3) RV model of 12 years of USD/CHF for 1 day, 1 week and 2 weeks ahead aggregated realized volatility of USD/CHF.

OUT OF SAMPLE MINCER-ZARNOWITZ REGRESSION

		b_0	b_1	R^2
	RM	0.044168	0.580141	0.2552
1		(0.0384, 0.0500)	(0.5367, 0.6236)	
	AR(1)	0.002169	0.977008	0.3764
D		(-0.0051, 0.0095)	(0.9179, 1.0361)	
A	AR(3)	0.004260	0.961717	0.3896
Y		(-0.0027, 0.0113)	(0.9052, 1.0182)	
	ARFIMA	0.010049	0.916610	0.3982
		(0.0035, 0.0166)	(0.8637, 0.9695)	
	HAR(3)	0.002030	0.982624	0.4150
		(-0.0047, 0.0088)	(0.9278, 1.0374)	
	RM	0.048970	0.536633	0.1333
1		(0.0410, 0.0570)	(0.4705, 0.5929)	
	AR(1)	-0.021963	1.164388	0.0801
W		(-0.0448, 0.0009)	(0.9786, 1.3502)	
E	AR(3)	-0.055300	1.444399	0.3196
E		(-0.0675, -0.0431)	(1.3452, 1.5436)	
K	ARFIMA	0.007047	0.938308	0.3569
		(-0.0002, 0.0143)	(0.8790, 0.9976)	
	HAR(3)	0.000191	0.997471	0.3692
		(-0.0073, 0.0077)	(0.9361, 1.0588)	
2	RM	0.072274	0.370246	0.0371
		(0.0614, 0.0831)	(0.2901, 0.4504)	
W	AR(1)	0.148271	-0.229492	0.0063
E		(0.1323, 0.1642)	(-0.3562, -0.1028)	
E	AR(3)	-0.027788	1.214038	0.1142
K		(-0.0474, -0.0082)	(1.0544, 1.3737)	
S	ARFIMA	0.006737	0.939015	0.2969
		(-0.0016, 0.0151)	(0.8708, 1.0072)	
	HAR(3)	0.002461	0.978118	0.3079
		(-0.0060 0.0109)	(0.9089 1.0474)	

Table 3.7: Out of sample Mincer-Zarnowitz regression for the RiskMetrics, AR(1), AR(3), ARFIMA(5,0.401,0) and HAR(3) model for the 1 day, 1 week and 2 weeks ahead aggregated realized volatility of USD/CHF (95% confidence interval in parentesis).

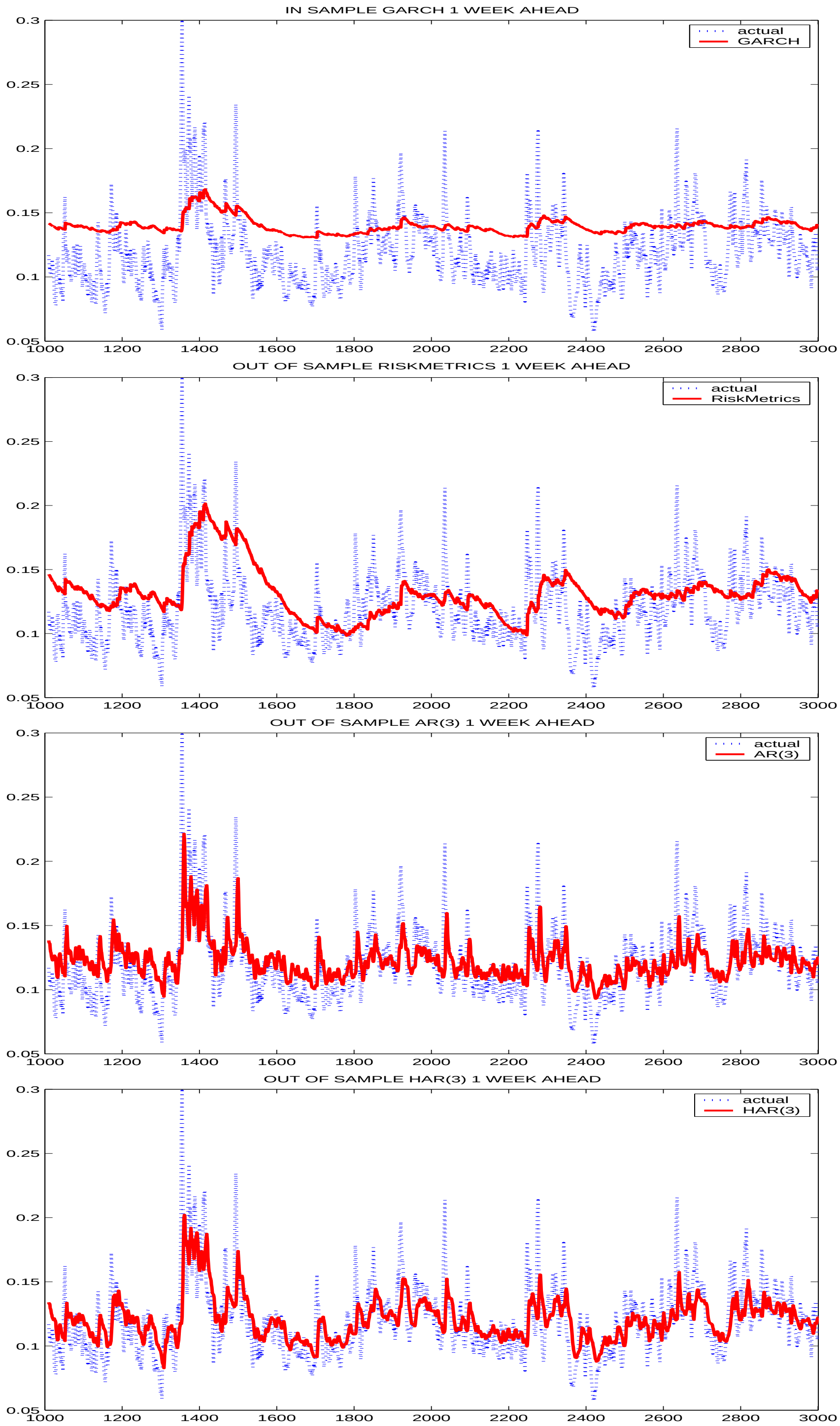


Figure 3.10: Comparison of out of sample 1 week aggregated volatility predictions for respectivly from top to bottom, GARCH, RiskMetrics, AR(3) and HAR(3) model. The continuos line is the prediction while the dotted line is the ex-post realized volatility over a 1 week period.

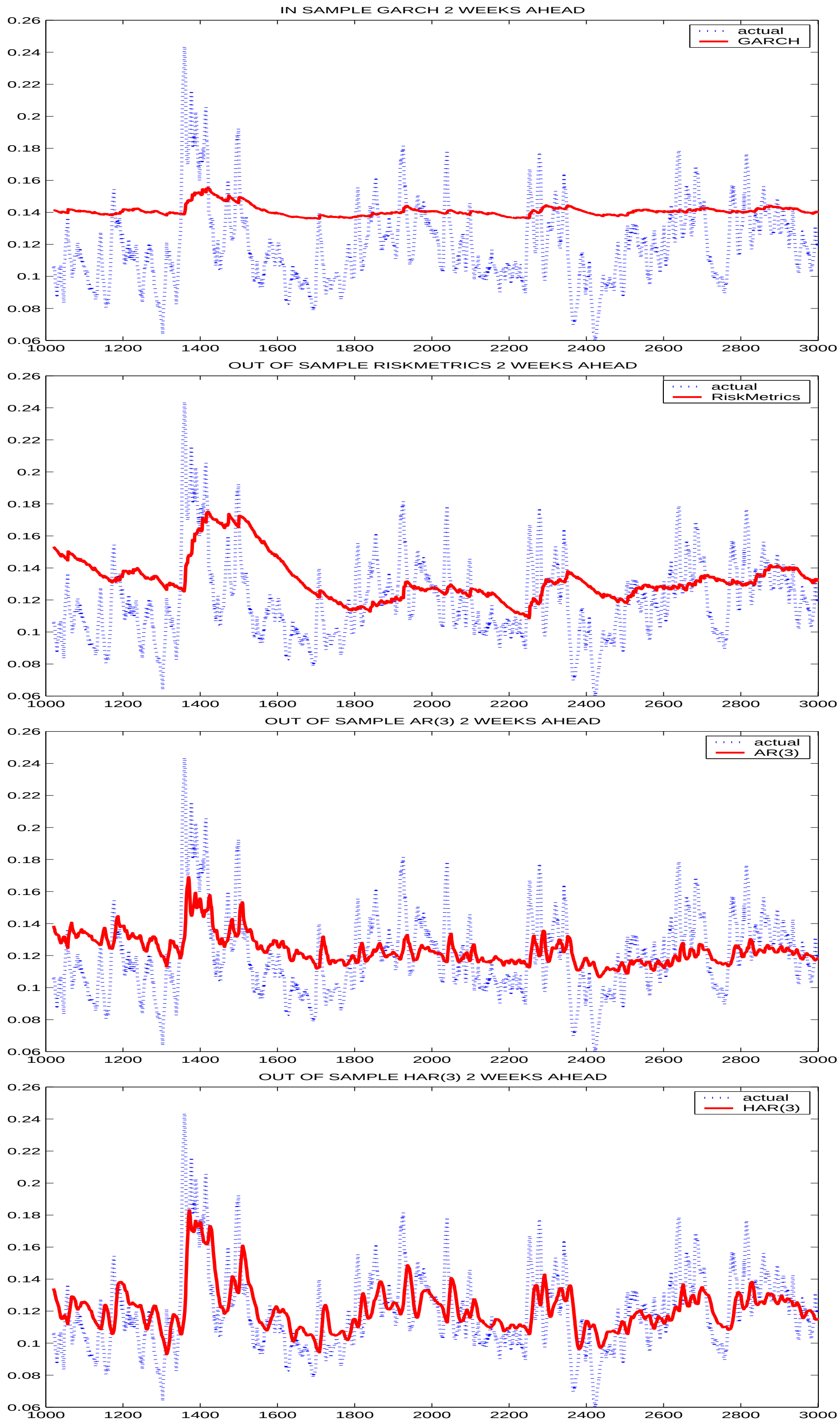


Figure 3.11: Comparison of out of sample 2 weeks aggregated volatility predictions for, respectivly from top to bottom, GARCH, RiskMetrics, AR(3) and HAR(3) model. The continuos line is the prediction while the dotted line is the ex-post realized volatility over a 2 weeks period.

Asymmetric-HAR-RV(3) ESTIMATED ON THE S&P 500 FUTURE
Newey-West Standard Errors and Covariance (n^o lags = 5)
Included observations: 3391
R-squared = 0.7365

Variable	Coefficient	t - Statistic	Probability
constant	0.243174	5.328021	0.000000
$\mathcal{RV}^{(d)}$	0.237635	9.095454	0.000000
$\mathcal{RV}^{(w)}$	0.341790	8.849822	0.000000
$\mathcal{RV}^{(m)}$	0.293096	9.328244	0.000000
$r_+^{(d)}$	-0.000100	-0.218239	0.827256
$r_-^{(d)}$	-0.005573	-9.815596	0.000000
$r_+^{(w)}$	-0.000824	-0.833938	0.404375
$r_-^{(w)}$	-0.005691	-4.474479	0.000008
$r_+^{(m)}$	0.008099	3.550714	0.000389
$r_-^{(m)}$	-0.008046	-2.833695	0.004629

IN SAMPLE MINCER-ZARNOWITZ REGRESSION

	b_0	b_1	R^2
Asymmetric HAR(3)	-0.024257 (-0.0750, 0.0265)	1.009447 (0.9894, 1.0295)	0.7427

Table 3.8: In sample estimation results of the Newey-West adjusted heteroscedastic-serial consistent least-squares regression of Asymmetric-HAR-RV(3) model for the S&P 500 from 1990 to 2003.

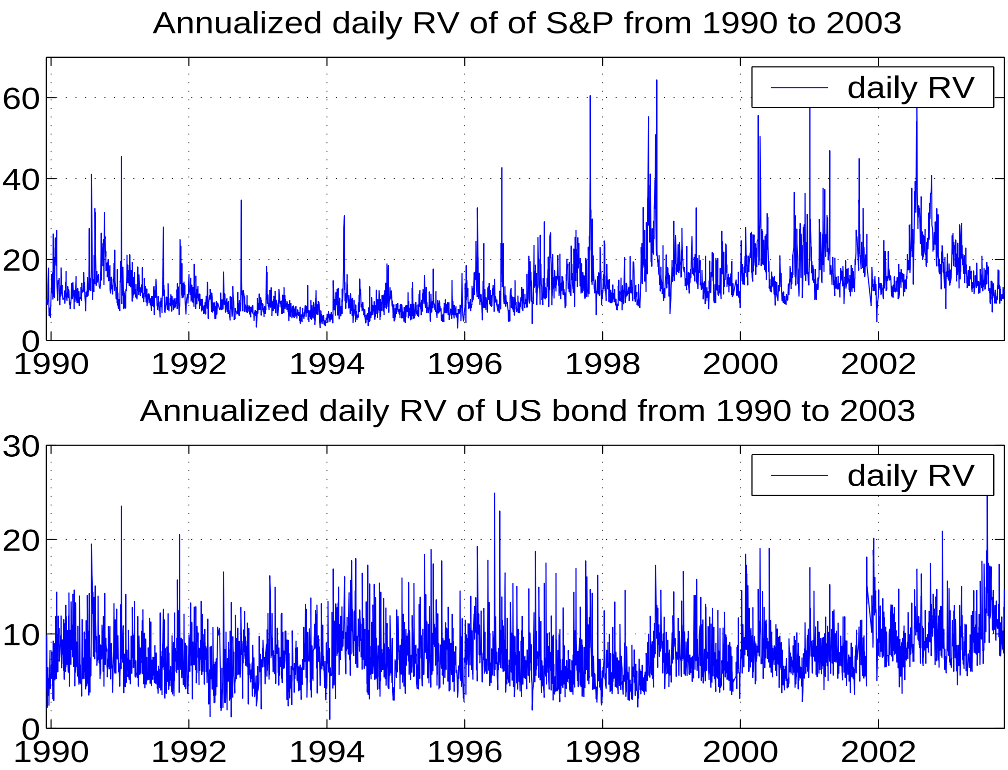


Figure 3.12: Annualized daily realized volatility of S&P 500 (top panel) and US Bond (bottom panel) future from 1990 to 2003.

The results of the estimation of the Asymmetric-HAR-RV(3) on the S&P data are reported in Table 3.8. All the three realized volatility component are again highly significant as for the FX data, but with a slightly less weight on the daily component and more on the weekly and monthly ones.

The most interesting result, however, is the strong significance (with the expected sign) of the negative returns at all the daily, weekly and monthly frequencies which unveils an heterogeneous component structure also in the leverage effect; i.e. not only the negative sign of the past day return affects the next day volatility (as well documented) but also the sign of the returns of the past week and past month have an impact on the volatility of tomorrow, which is over and above that of the previous day. This means that, also for the leverage effect, the market posses not only a daily memory but also a weekly and monthly one, observing and reacting to price trends happened in the past week and month. To our knowledge this is a novel empirical finding that further confirm the view of the Heterogeneous Market Hypothesis.

It is also interesting to note that for the positive returns the significance increases with the length of the aggregation period, being insignificant at daily and weekly level while becoming significant at monthly level.

The performance of the one day ahead in sample prediction of the Asymmetric-HAR(3) model on the S&P series is quite remarkable reaching an R^2 of 74% for the log realized volatility and one of 70% for the realized volatility itself. Figure 3.13 visualizes this notable fit and shows the time series of the residuals that clearly exhibits significant degree of heteroskedasticity which would justify, also for the S&P, more sophisticated estimation procedures robust to GARCH effects on volatility resiudals, as proposed in Corsi, Kretschmer and Pigorsch (2005).

Before analyising the estimation results of the Asymmetric-HAR(3) on the US bond data, it should be noted how the time series of US bond realized volatility seems to show a higher level of noise than that of the S&P. This is confirmed by the comparations of the autocorrelation function of the two series shown in figure 3.14. The much lower level and faster decay of the autocorrelation function of US bond indicate the presence of a weaker signal and hence a lower degree of predictability and higher level of noise in the series.

This higher level of noise in the realized volatility series of the bond maybe due to a less precise measurement of the daily volatility. In fact, the tick frequency of the bond is already more than two times lower than that of the S&P with an average of about 3.3 ticks per minute for the bond against 7.5 ticks per minute for index future. But the real problem with the tick-by-tick data set of the bond is that due to the strong oscillatory AR(1) structure of the observed price (described in section 2.6), the actual number of ticks containing some real information on the dynamics of the true price process is much lower than the total number of ticks. In fact, a simple two ticks average of the bond prices cleary showed (see figure 2.5 in section 2.6) how most of the tick-by-tick returns are just bouncing of the observed price between the same bid and ask quotes. As a consequence the effective number of useful observations can be many times smaller than what would appear from a simple count of the number of ticks.

Accordingly to that high level of the noise to signal ratio in the bond realized volatility series, the R^2 of the estimation of the Asymmetric-HAR(3) model is much lower, remaining below 25%. Moreover, the noisy estimation of the daily realized volatility could be responsible for the lack of significance of the daily volatility component. The weekly and monthly realized volatility, however, being averages over longer periods, arguably contain less noise and more information on the volatility process and, hence, they receive higher weights from the model

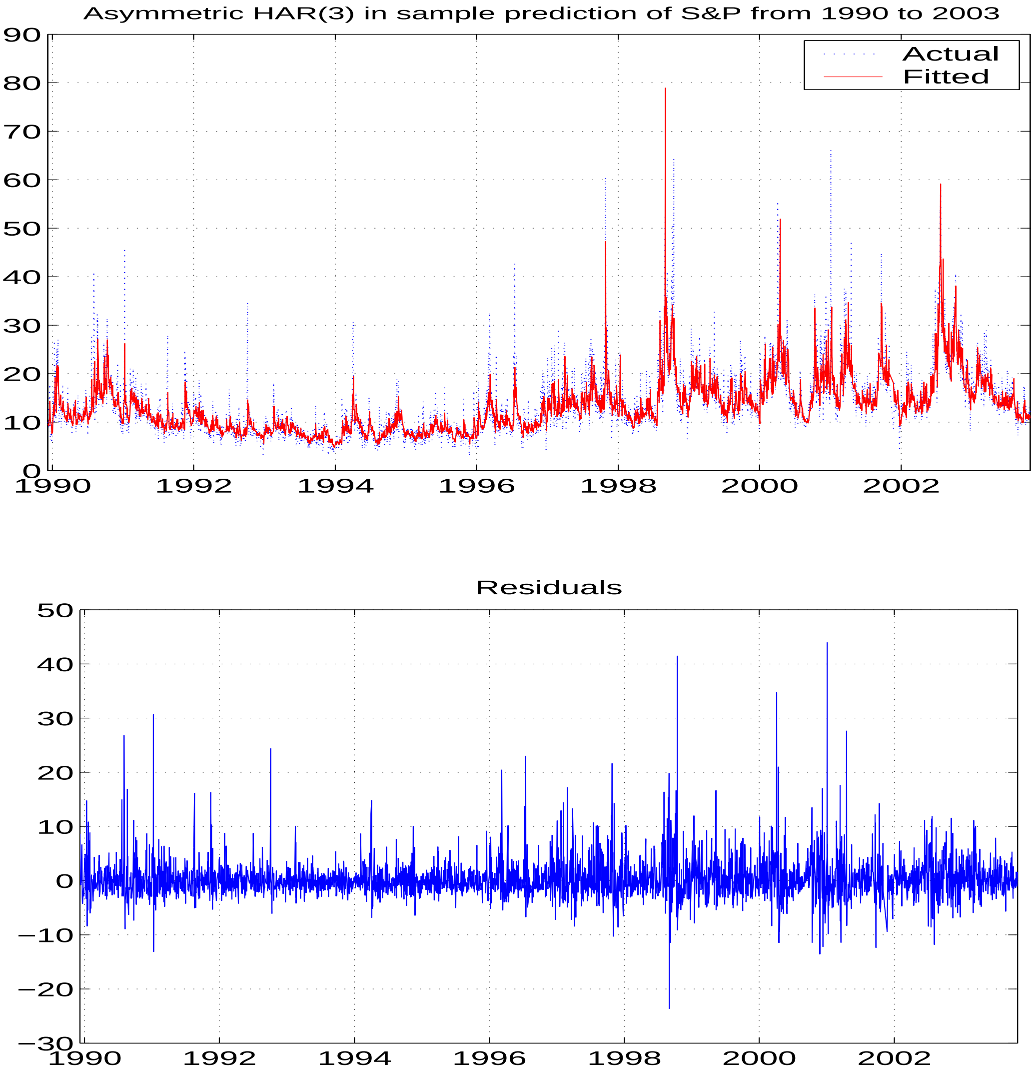


Figure 3.13: Asymmetric-HAR-RV(3) in sample prediction of annualized daily realized volatility of S&P 500 (top panel) and associated residuals (bottom panel) from 1990 to 2003.

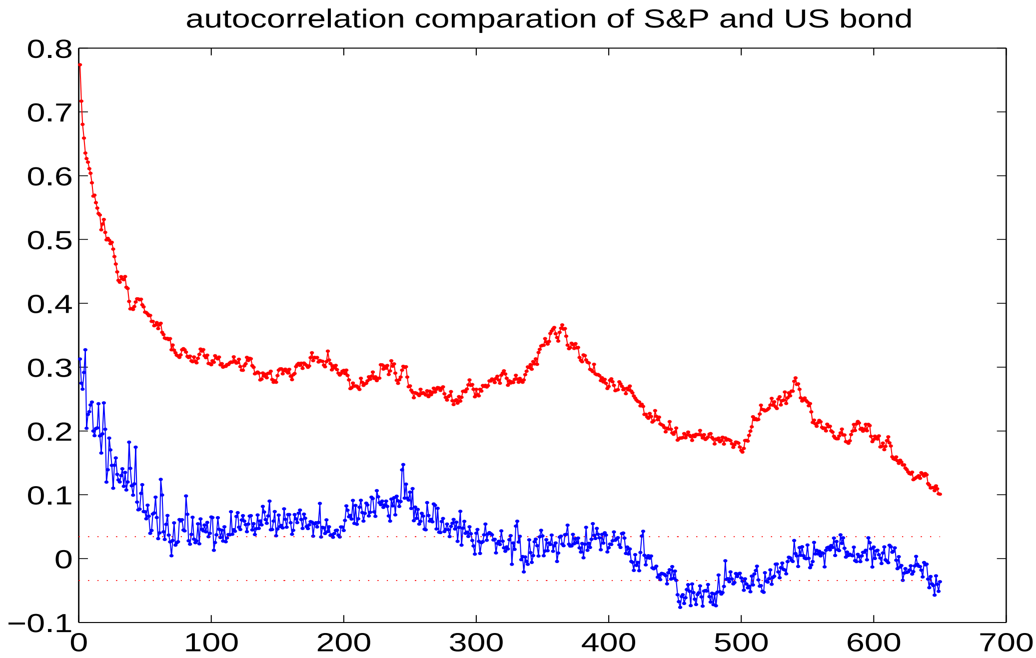


Figure 3.14: Comparation of the autocorrelation function of the S&P 500 (upper line) and US bond (lower line) realized volatility computed on the full sample 1990-2003.

(with the monthly weight even higher than the weekly one).

Interestingly, the three negative returns remain again significant at the 5% level for all the horizons, confirming, also for the US bond (Table 3.9) , the existence of heterogeneous components in the leverage effect.

3.9 Conclusions

The additive volatility cascade inspired by the Heterogeneous Market Hypothesis leads to a simple AR-type model in the realized volatility which has the feature of considering volatilities realized over different time intervals. We term this model, Heterogeneous Autoregressive model of the Realized Volatility (HAR-RV). The new HAR-RV model seems to successfully achieve the purpose of modelling the long memory behavior of volatility in a very simple and parsimonious way. In spite of the simplicity of its structure and estimation, the HAR-RV model shows remarkably good in sample and out of sample forecasting performance. Moreover, by extending the Heterogeneous Market Hypothesis idea to the leverage effect, we propose an asymmetric version of the HAR-RV model which considers asymmetric responses

Asymmetric-HAR-RV(3) ESTIMATED ON THE US BOND FUTURE
Included observations: 3391 after adjusting endpoints
Newey-West Standard Errors and Covariance (lag=5)
R-squared = 0.2464

Variable	Coefficient	t - Statistic	Probability
constant	0.328425	4.991717	0.000001
$\mathcal{RV}^{(d)}$	-0.000052	-0.001973	0.998426
$\mathcal{RV}^{(w)}$	0.374297	7.650972	0.000000
$\mathcal{RV}^{(m)}$	0.412192	8.240982	0.000000
$r_+^{(d)}$	0.003799	2.861930	0.004237
$r_-^{(d)}$	-0.006080	-4.547667	0.000006
$r_+^{(w)}$	0.002356	0.705351	0.480640
$r_-^{(w)}$	-0.006422	-2.026643	0.042777
$r_+^{(m)}$	0.016449	2.642759	0.008261
$r_-^{(m)}$	-0.009393	-1.974038	0.048459

IN SAMPLE MINCER-ZARNOWITZ REGRESSION

	b_0	b_1	R^2
Asymmetric HAR(3)	-0.027279 (-0.1471, 0.0926)	1.013589 (-0.1471, 0.0926)	0.2472

Table 3.9: In sample estimation results of the Newey-West adjusted heteroscedastic-serial consistent least-squares regression of Asymmetric-HAR-RV(3) model for the US bond from 1990 to 2003.

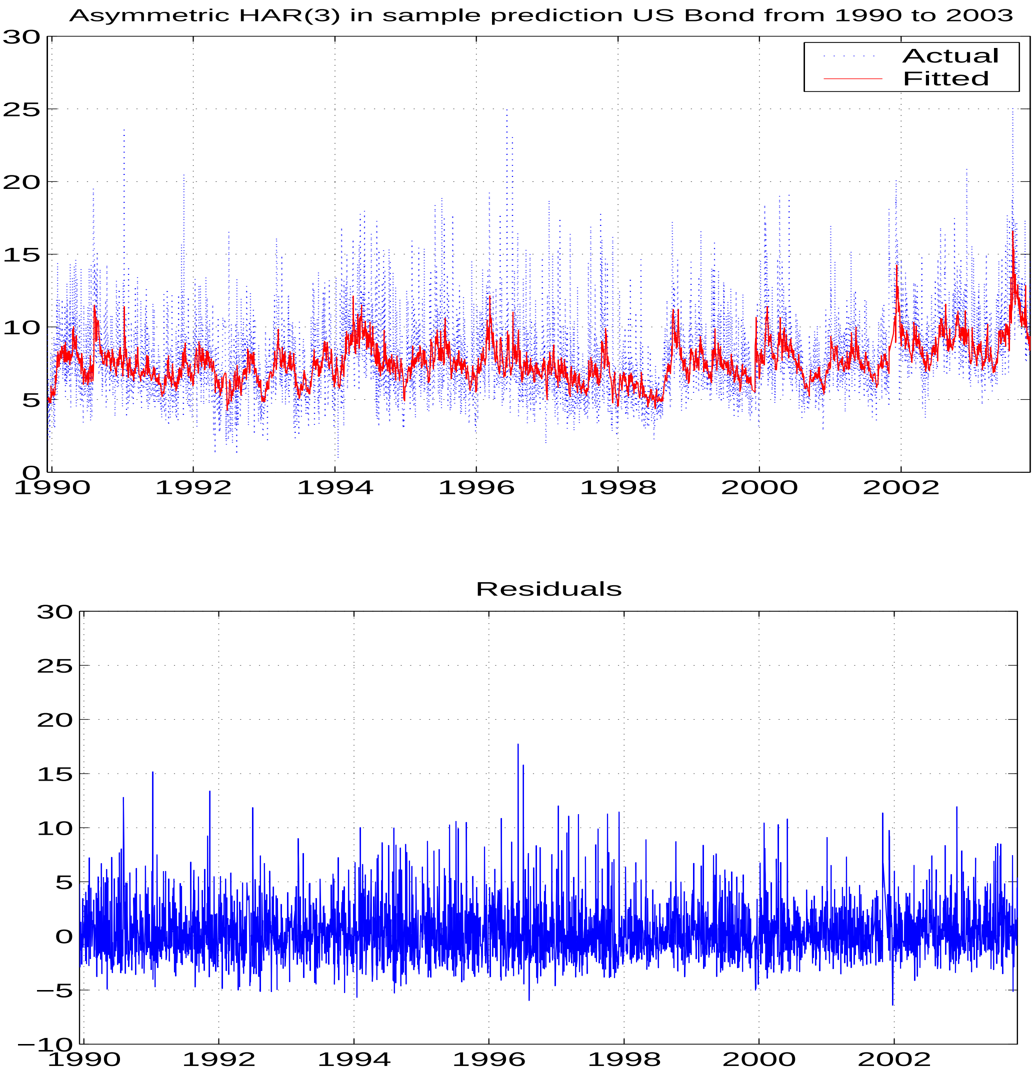


Figure 3.15: Asymmetric-HAR-RV(3) in sample prediction of annualized daily realized volatility of US bond (top panel) and associated residuals (bottom panel) from 1990 to 2003.

of the realized volatility not only to previous daily returns but also to past weekly and monthly returns. The significance (with the expected sign) of the negative returns at all the three frequencies unveils the presence of an heterogeneous component structure arising from the activities and reactions of trading subjects operating at different frequencies, also in the leverage effect. This novel empirical finding seems to further confirm the validity of the Heterogeneous Market Hypothesis.

Chapter 4

Realized Correlations

4.1 Introduction

Asset returns cross correlations are pivotal to many prominent financial problems such as asset allocation, risk management and option pricing. As for the realized volatility approach, the idea of employing high frequency data in the computation of covariances and correlations between two assets leads to the analogous concept of *realized covariance* (or covariation) and *realized correlation* .

However, also the standard realized covariance measures are highly sensitive to the presence of market microstructure effects inducing a bias which increases with the sampling frequency. Therefore, as for the volatility, the frequency of the cross product returns used in the computation of the standard realized covariance is usually reduced to avoid the impact of microstructure effects. Instead, in this chapter, we propose a realized correlation measure constructed from realized volatility and realized covariance both computed using all the tick-by-tick data available.

4.2 Realized Covariance tick-by-tick

The standard way to compute the realized covariance is to first choose a time interval, construct an artificially regularly spaced time series by means of some interpolation scheme and then take the contemporaneous sample covariance of those regularly spaced returns. But simulations and empirical studies indicate that such covariance measure presents a bias toward zero which rapidly increases with the reduction of the time length of the fix interval chosen. As for the realized volatility, the presence of market microstructure can induce significant bias in standard realized covariance measure. However, the microstructure effects responsible for this bias are different. In fact, bid-ask bouncing, which is the major source of bias for the realized volatility, will just increase the variance of the covariance estimator but it will not induce any bias. On the contrary, the so called non-synchronous trading effect (Lo and MacKinlay 1990) strongly affects the estimation of the realized covariance and correlation. In fact, since the sampling from the underlying stochastic process is different for different assets, assuming that two time series are sampled simultaneously when, indeed, the sampling is nonsynchrounos gives rise to the non-synchronous trading effect. As a result, covariances and correlations measured with high frequency data will posses a bias toward zero which increases as the sampling frequency increases. This effect of a dramatic drop of the absolute value of correlations among stocks when increasing the sampling frequency was first reported by Epps (1979) and hence called "the Epps effect". Since then, the Epps effect has been confirmed on real data and simulations by many other authors, such as Dacorogna and Lundin (1999) Renó (2003) and Martens (2004), among others.

Existing empirical work on realized covariance usually compute the sample covariance based on the 5 or 30 minutes return interval. Such frequencies are heuristically chosen to try to avoid the bias and market microstructure effects. In some cases, a numbers of leads and lags covariance are added to reduce the remaining bias. However this type of correction will increase the variance of the estimator. Though the optimal choice of the frequency of the returns and the number of lead and lags would substantially lowers the RMSE compared to the heuristic choices, this optimal values are unknown in empirical application (Martens 2004).

Intuitively, the reason why the non-synchronous trading effect biases the usual covariance estimator based on fixed interval returns can be seen as twofold. First, the absence of trading on one asset in a certain interval produce a zero returns for that interval and then artificially imposes a zero value to the cross product of returns inducing a bias toward zero in the realized covariance (which, in its standard version, is simply the sum of those cross products).

Secondly, the construction of a regular grid, depending on the frequency of tick arrivals, affects the computation of the realized covariance. For the more liquid assets with higher average arrival rates, the last tick falling in a certain grid interval is typically much closer to the end point of the grid compared to that of a less liquid asset. Any difference in the time stamps between these last ticks in grid for the two assets, will correspond to a portion of the cross product returns which will not be accounted for in the computation of the covariance. This is because for the more liquid asset, the (unobserved) returns corresponding to this time difference will be imputed to the current grid interval while for the less liquid asset such portion of returns will be ascribed to the next grid interval, so that the two will be no longer matched and their contribution to the cross products sum will be lost. This lost portion of covariance in each interval also induces a downward bias in the realized covariance computed with a regular grid; a bias which will also increase with the number of intervals and hence with the frequency.

Contrary to these approaches the simple realized covariance estimator adopted here does not resort on the construction of a regular grid. Instead, following the general statistical principles which tells us to never "throw data away", we analyse an unbiased realized covariance measure directly built on the raw tick-by-tick data series (Corsi 2005). In this way, both source of bias (the presence of zero returns and the lost portion of cross products) of the realized covariance, are avoided.

Under the assumption of no true leads and lags cross-covariance, an unbiased covariance estimator can be computed by simply summing all the cross products of returns which have a non zero overlapping of their respective time span. In other words, a given tick-by-tick return on one asset is multiplied with any other tick-by-tick returns of the other asset which has a non zero overlap in time, i.e. which share (even for a very small fraction) the same time interval.

Analytically, this tick-by-tick Realized Covariance estimator can be defined as

$$RC_t = \sum_{s=1}^{M_{i,t}} \sum_{q=1}^{M_{j,t}} r_{i,s}\, r_{j,q}\, I(\tau_{q,s} > 0) \tag{4.1}$$

with $I(\cdot)$ the indicator function and

$$\tau_{q,s} = \max(0, \min(n_{s+1}, n_{q+1} - \max(n_s, n_q)) \tag{4.2}$$

being the overlap in time between any two tick-by-tick returns $r_{i,s}$ and $r_{j,q}$.

The simplest way to intuitively show the unbiasedness of this estimator, is by assuming an underling discrete time process, with arbitrary clock time interval δ. In this setting, the expectation of the cross product of two overlapping tick-by-tick returns $r_{i,s},\ r_{j,q}$ can be expressed as a linear combination of the cross-covariances $\gamma(h) = \mathrm{Cov}(r_{i,s},\ r_{j,s-h\cdot\delta})$. But, being all the cross-covariances with $h \neq 0$ equal to zero, it reduces to $\mathbb{E}\left[r_{i,s}\, r_{j,q}\right] = \tau_{q,s}\gamma(0)$. Therefore, the expectation of the tick-by-tick covariance estimator RC_t, is

$$\mathbb{E}\left[RC_t\right] = \left[\sum_{s=1}^{M_{i,t}} \sum_{q=1}^{M_{j,t}} r_{i,s}\, r_{j,q}\, I(\tau_{q,s} > 0)\right] = \sum_{s=1}^{M_{i,t}} \sum_{q=1}^{M_{j,t}} \tau_{q,s}\gamma(0). \tag{4.3}$$

and given that the sum of all the overlapping intervals $\tau_{q,s}$ in a day is the whole trading day itself, we can conclude that RC_t is an unbiased estimator of the daily covariance.

Hence, loosely speaking, this estimator is unbiased because no portion of covariance will be lost while the portion of cross product which does not overlap will have zero mean. Moreover, avoiding the noise and the discarding of price observations caused by the regular grid interpolation, will considerably reduce its variance. However, in the presence of a fix ammount of market microstructure noise, this estimator will not be consistent because, although unbiased, his variance will diverge as the number of observations tends to infinity.

4.3 Simulations

In this section we evaluate the performance of different covariance estimators in a simulation environment based on the Lo and MacKinlay's (1990) non-synchronous trading model. In this model the true return of any asset i is given by a single factor model. Considering only two assets, the two return series are then given by

$$r_{i,t} = \mu_i + \beta_i f_t + \epsilon_{i,t} \qquad i = 1, 2 \tag{4.4}$$

where β_i is the factor loadings of asset i, $\epsilon_{i,t}$ represents the idiosyncratic noise of asset i and f_t is the zero mean common factor.

Under the assumptions that the idiosyncratic noise $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are mutually uncorrelated and both uncorrelated with the common factor f_t, the true covariance between the two assets is simply

$$\sigma_{1,2} = \beta_1 \beta_2 \sigma^2_{f,t} \tag{4.5}$$

where $\sigma^2_{f,t}$ is the variance of the common factor f_t.

In the Lo and MacKinlay's model the common factor f is assumed to be a simple homoskedastic process and, hence, the variance of f is a constant σ^2_f. As a consequence, also the true covariance remains constant. In the version adopted here, however, in order to give more dynamics and realism to the DGP, the common factor f is assumed to follow the stochastic volatility model of Heston (1993) so that, also the true covariance will dynamically change through time.

Therefore, the dynamics of the common factor is given by the following continuous time process

$$\begin{aligned} f(t) &= (\mu - v(t)/2)dt + \sigma_f(t)dB(t) && (4.6) \\ dv(t) &= k(\alpha - v(t))dt + \gamma v(t)^{1/2} dW(t) && (4.7) \end{aligned}$$

where $v = \sigma_f^2$ and the initial value $v(0)$ is drawn from the unconditional Gamma distribution of v.

The values of the parameters are chosen so to have a process with zero mean, expected annualized volatility of 15% and satisfying the Feller's condition $2k\alpha = \gamma^2$. Thus, the following annualized values for the parameter are chosen: $\mu = 0$, $k = 8$, $\alpha = 0.0225$, $\gamma = 0.5$ and a correlation coefficient between the two Brownian motions of $\rho = -0.5$. Those parameter values, will remain constant throughout the simulations. The Heston model for the factor and the true returns process of the two assets will be simulated at the usual Euler clock of one second.

In the Lo and MacKinlay's model the prices are assumed to be observed with a certain probability $1-\pi_i$, where π_i is the so called non-trading probability. We found more convenient to express the frequency of the price observations in terms of the corresponding average intertrade duration between ticks[1] τ_i.

Each time a price is observed we simulate market microstructure noise by randomly adding or subtracting half of the spread to the true price. The size of the spread is chosen so to obtain an average level of the noise to signal ratio of the observed returns process equal to one.

In addition to the proposed tick-by-tick estimator, the other covariance measures included for comparison in the simulation are:

- The standard realized covariance computed with an interpolated regular grid of 1 minute returns.
- The standard realized covariance computed with a fix returns time interval of 5 minutes.
- The Scholes and William (1977) covariance estimator, which add to the contemporaneous sample covariance of fix interval returns, one lead and lag cross covariance. To improve the performance of this estimator we chose the frequency of the fix interval returns which seems to provide the best results given the observation frequency of the two assets.
- The estimator proposed by Cohen et al. (1983), which is a simple generalization of the Scholes and Williams estimator where more than one leads and lags are considered. Here, as in Bollerslev and Zhang (2003) we compute the Cohen et al. estimators with 12 leads and lags and at the frequency which seems to be the best performing given the corresponding simulation set up.

[1] For example a non-trading probability of 90% corresponds to an exponential distribution of the intertrade duration with a mean value of 10 seconds.

- The Lo and MacKinlay's estimator given by

$$\hat{\sigma}_{1,2} = \frac{1 - \hat{\pi}_1 \hat{\pi}_2}{(1 - \hat{\pi}_1)(1 - \hat{\pi}_2)} \text{Cov}\left[r^s_{1,t}, r^s_{2,t}\right] \tag{4.8}$$

  where $\text{Cov}\left[r^s_{1,t}, r^s_{2,t}\right]$ is the covariance between the observed 1 second returns $r^s_{i,t}$. Contrary to the highly noisy non-trading probability estimation proposed by Lo and MacKinlay's where $\hat{\pi}_1 = \text{Cov}\left[r^s_{1,t}, r^s_{2,t+1}\right] / \text{Cov}\left[r^s_{1,t}, r^s_{2,t}\right]$ and analogously for $\hat{\pi}_2$, we estimate those probabilities by simply counting the observed number of ticks in each day and dividing it by the total number of seconds in the day.

We first simulate 25,000 paths at a moderate observation frequency of $\tau_1 = 30$ seconds and $\tau_2 = 1$ minute. With the factor loadings $\beta_1 = 0.8$ and $\beta_2 = 1.25$, and an average value for the volatility of the common factor of 15% per annum, the true covariance is, on average, 2.25% per annum, which together with the volatilities of the idiosyncratic noises $\sigma_{\epsilon,1} = 0.16$ and $\sigma_{\epsilon,2} = 0.16535$ implied an average correlation of 45%. Given the relatively low frequency of the two series we compute the Scholes and Williams estimator with 3 minutes returns and the Cohen et al., which is able to correct for higher level of bias, at the 20 seconds interval (i.e. at the average frequency of the more liquid asset). The results are reported in figure 4.1 and in the left panel of table 4.1.

With those observation frequencies the 1 minute realized covariance is highly biased: on average it would correspond to a 20% correlation against the true value of 45%. Under these conditions, the 5 minutes realized covariance gives better results both in terms of dispersion and in terms of bias (though a significant bias still exists being the implied correlation equal to 37.5%). Despite the direct estimation of the non-trading probabilities the Lo and MacKinlay's estimator (though unbiased) is extremely inaccurate probably because of the significant presence of market microstructure noise. With the carefully chosen frequency both the Scholes and Williams and the Cohen et al. estimators are almost unbiased and reasonably accurate. However, the best estimator is clearly the proposed one (termed "All-Ticks" in figures and tables) with no bias and the smallest dispersion.

We repeat the simulation with a higher observation frequency for the two assets, choosing $\tau_1 = 5$ seconds and $\tau_2 = 10$ seconds. Now, the return frequency for the Scholes and Williams estimator is chosen at 30 seconds and that of Cohen et al. at 5 seconds. Figure 4.2 and the right panel of table 4.1 report the results.

The 1 minute realized covariance, though less disperse now, is still significantly bias with an implied average correlation of about 39%. The 5 minutes realized covariance, instead, is unbiased but with a large variance. The higher number of price observations seems to be of

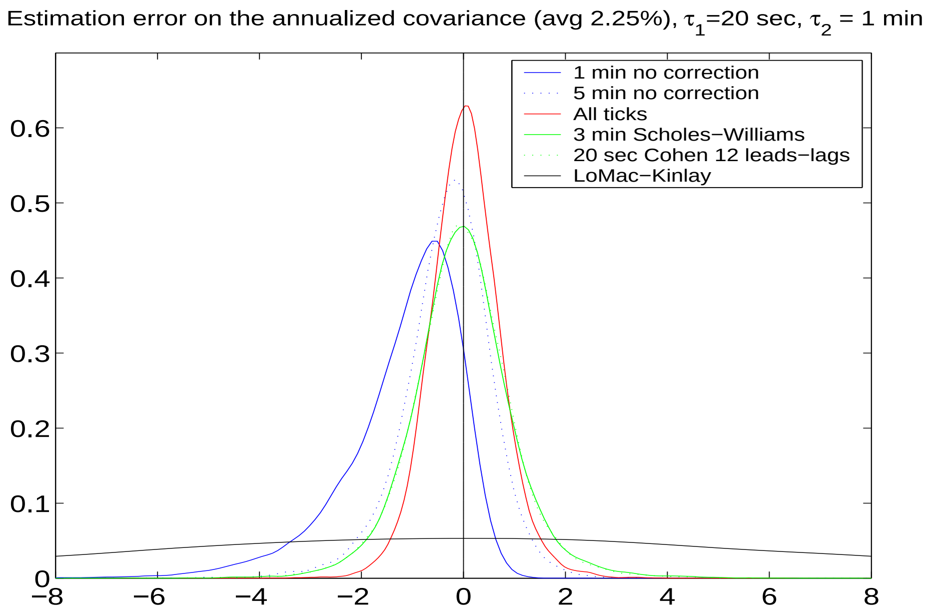


Figure 4.1: Comparison of the pdf of the covariance estimation errors with noise to signal one and average observation frequencies $\tau_1 = 30$ seconds and $\tau_2 = 1$ minute.

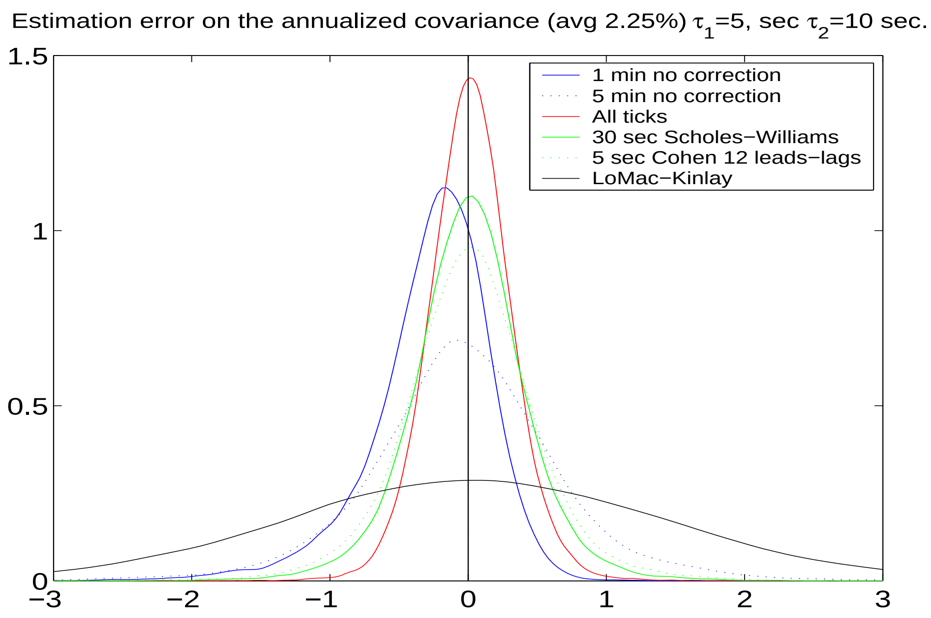


Figure 4.2: Comparison of the pdf of the covariance estimation errors with noise to signal one and average observation frequencies $\tau_1 = 5$ seconds and $\tau_2 = 10$ seconds.

COVARIANCE ESTIMATION RESULTS

	$\tau_1 = 30$ sec, $\tau_2 = 1$ min			$\tau_1 = 5$ sec, $\tau_2 = 10$ min		
	bias	std	RMSE	bias	std	RMSE
1 min no correction	-1.2438	1.1592	1.7002	-0.2921	0.4352	0.5241
5 min no correction	-0.3786	0.8783	0.9564	-0.0599	0.7233	0.7258
All-Ticks	0.0095	0.6930	0.6931	0.0057	0.2988	0.2988
Scholes-Williams	-0.0685	1.0148	1.0171	-0.0168	0.4229	0.4233
Cohen 12 leads-lags	-0.0795	0.9925	0.9957	-0.0121	0.4919	0.4920
LoMac-Kinlay	-0.0074	7.8983	7.8983	0.0143	1.4013	1.4014

Table 4.1: The table reports the mean, standard deviation and RMSE of the estimation errors on the annualized covariance (on average 2.25%) obtained at different observation frequencies for the two assets and a noise to signal ratio equal to one.

little help for the performance of the Lo and MacKinlay's estimator in the presence of market microstructure noise. As before, at the chosen frequencies both the Scholes and Williams and the Cohen et al. estimators are almost unbiased, with the second one being slightly more precise. But, again, the tick-by-tick covariance estimator remain unbiased and the most precise among the estimators considered.

Summarizing the results of this simulation study, the simple tick-by-tick estimator proposed, results to be the best performing for both choices of trading frequencies of the two assets. Surprisingly, it also performs favorably compared to the Scholes and Williams and the Cohen et al. estimators even if their return frequency has been chosen according to the simulation settings to give the best results. The proposed tick-by-tick estimator, however, does not require any choice of return frequency or interpolation scheme since it can be directly applied to the raw tick-returns series of any two assets, always providing unbiased results.

4.4 Empirical application

We apply the proposed tick-by-tick covariance estimator to the bivariate series of S&P 500 and 30 years US Treasury Bond futures from 1990 to 2003.

Unfortunately, since the time stamps of the data in our disposal, are rounded at the 1 minute level, the proposed estimator can not be directly implemented in such a simple way but it requires a slightly different scheme. In fact, the rounding of the seconds to the minute, precludes the knowledge of the correct time ordering among the ticks of the two series inside the 1 minute interval, which is necessary for the application of the tick-by-tick estimator. To overcome this problem, we construct the tick-by-tick estimator by simply considering only the firsts and lasts ticks of each 1 minute interval. This version will hence be termed "First-Last" tick-by-tick estimator.

Using a subsample of the total number of ticks employed by the "All-Ticks" estimator, we expect the "First-Last" to be less efficient. In order to evaluate this loss of efficiency of the modified tick-by-tick estimator on this type of data, we perform a simulation study which tries to reproduce as much as possible the econometric properties of the two empirical series, i.e. the parameters of the simulation will be chosen to match as closely as possible the empirical observation frequencies, level of volatilities, noise structure and intensities and so on.

Therefore, with asset 1 mimicking the S&P and asset 2 the US bond, the following configuration of the parameters are chosen: $\tau_1 = 8$ seconds, $\tau_2 = 18$ seconds, an average annualized volatility of about 20% for asset 1 and 10% for asset 2 and with a correlation of 30% . As describe in section 2.6 the noise structure is closely reproduced by assuming more complex ARMA structures for the market microstructure component: an MA(2) for the asset 1 (S&P) with $\theta_1 = 0.85$, $\theta_2 = 0.25$ and a noise to signal of 0.45, and a strong oscillatory AR(1) with $\phi_1 = -0.6$ and noise to signal of 0.6 for the asset 2 (US bond).

Figure 4.3 and table 4.2 reports the results of the 25,000 simulations. As expected, the First-Last estimator results to be less precise than the All-Ticks, however this loss of efficiency due to the lower number of ticks employed, is contained and the First-Last remain the best performing measure compared to the other covariance estimators.

Applying to the S&P 500 and US bond series the First-Last estimator we obtain the realized covariance time series shown in figure 4.4. To better appreciate the remarkable difference between the tick-by-tick realized covariance and the standard cross product of daily returns (the usual proxy for daily covariance in standard multivariate volatility models) both measures are plotted together on the same scale.

Combining the First-Last covariance measure together with the Fitted DST volatility estimator we are now able to obtain a realized correlation measure where both the volatilities and the covariance are computed from tick-by-tick data. Figure 4.5 shows the time series of

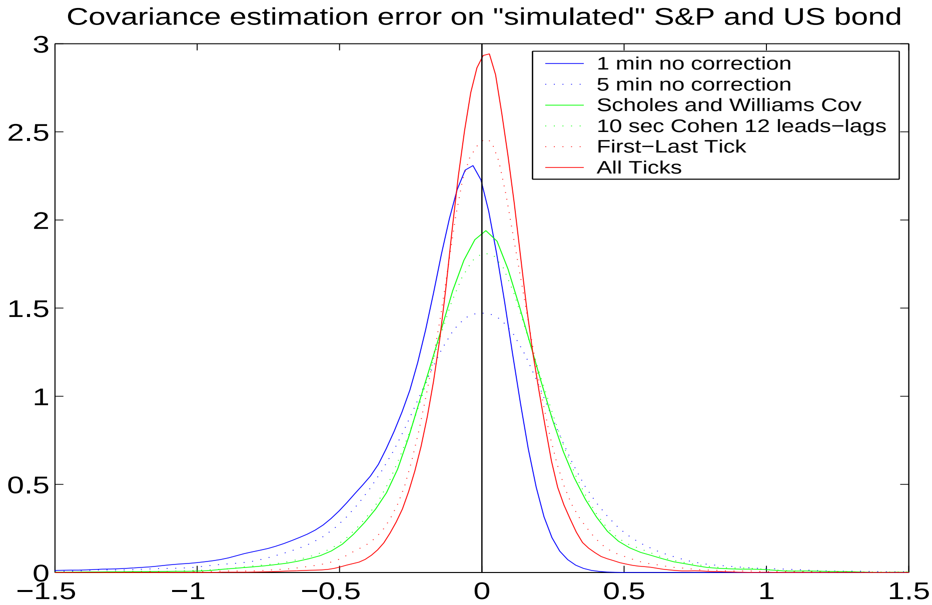


Figure 4.3: Comparison of the pdf of the covariance estimation errors for a simulation set up which reproduces the statistical properties of the S&P 500 and US bond future data.

3,391 daily tick-by-tick correlations from 1990 to 2003[2].

Simply looking at the correlation dynamics, there seems to be an important change of regime around the end of October 1997. Before that date, the correlation between the two series seems to oscillate around a positive stable value of about 20% until '94 and around 40% from '94 to '97, while after end of '97 the correlation start to exhibit a stronger dynamics and becomes predominantly negative.

This structural change in the dynamics of the correlation between S&P and US bond is also apparent from the different behaviour of the autocorrelation function computed on the '90-'97 sample and the '98-'03 one (figure 4.6). In the first sample the level of the autocorrelation for the firsts 100 lags is much lower and the decay of the firsts 50 lags is much faster. After the end of '97 the memory of the process, in particular the short and medium ones, sharply increases. It is also interesting to note how the structural change affects the global autocorrelation function computed on the full sample inducing an artificially high persistence in the autocorrelation coefficients. Nonetheless, even without this structural

[2]In two days (out of 3,391) the estimated realized correlation resulted to be outside the [-1, 1] correlation boundary, in those two cases we set the correlation absolute value to 0.9999.

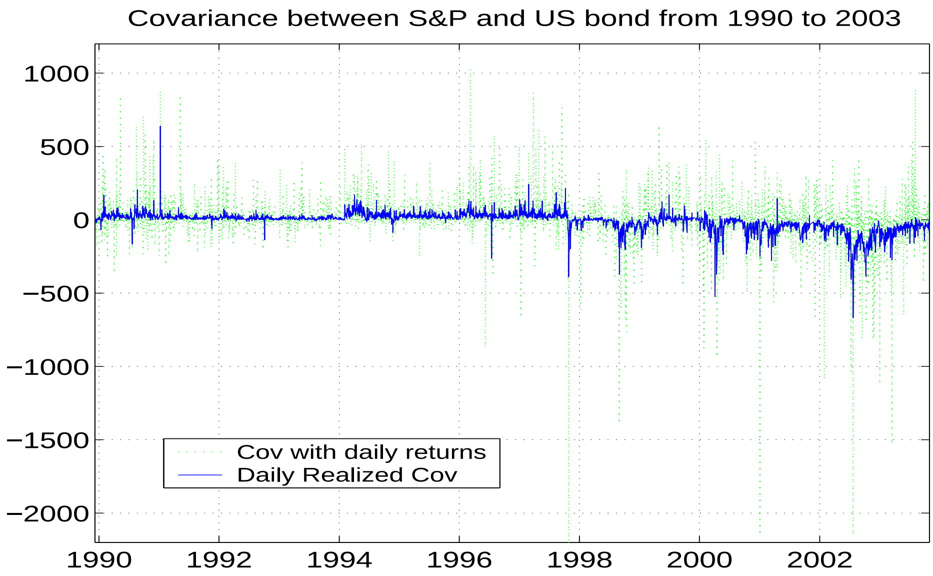


Figure 4.4: Time series of tick-by-tick realized covariance and the daily cross product returns of S&P 500 - US bond from 1990 to 2003.

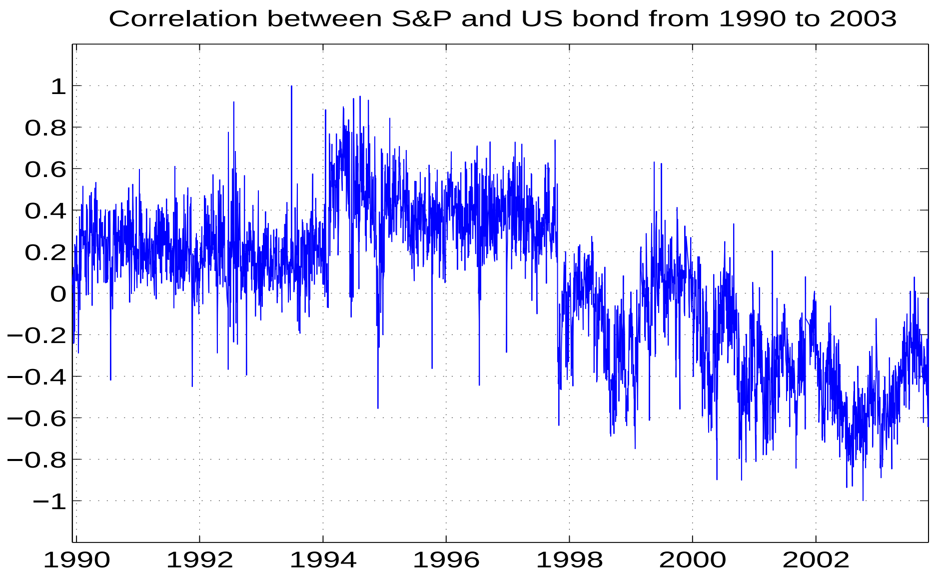


Figure 4.5: Time series of tick-by-tick realized correlation of S&P 500 - US bond from 1990 to 2003.

COVARIANCE ESTIMATION RESULTS

	bias	std	RMSE
1 min no correction	-0.1746	0.2729	0.3240
5 min no correction	-0.0382	0.3385	0.3406
Scholes and Williams Cov	-0.0076	0.2625	0.2626
10 sec Cohen 12 leads-lags	-0.0061	0.2751	0.2752
First-Last Tick	-0.0009	0.2018	0.2018
All Ticks	0.0019	0.1580	0.1581

Table 4.2: The table report the mean, standard deviation and RMSE of the estimation errors on the annualized covariance for a simulation set up which reproduces the statistical properties of the S&P 500 and US bond future data.

break effect, the autocorrelation function of the realized correlation remain highly persistent as shown by the two separated sub sample autocorrelations.

4.4.1 Modelling Realized Correlation with the HAR model

This empirical evidence on the high degree of persistence of correlations, suggests that the parsimonious HAR model could also be successfully applied to model the time series of realized correlations. Therefore, we first estimate the realized correlation HAR model (HAR-RC) on the full sample. Table 4.3 reports the results of this in sample estimation and 1 day ahead prediction on the full period showing the remarkable R^2 and the high significance level of the three heterogeneous components also for the correlation process. Then, in order to study the time evolution of the weights of the three different market components, we estimate the HAR-RC(3) model on a moving window of 1000 days (making 1 day ahead prediction at each step). Looking at the bottom panel of figure 4.7 shows how from the beginning of the sample (which is now 1994 because the first 4 years are used in the burn in of the 1000 days rolling window) the weight of the daily component steadily increases until 2002 going from about 10% (the smallest one of the three) to the about 40% (the highest one). Such increment is only partially compensated by a decline in the weekly component (from 40% to 30%), while the monthly one remains substantial the same in the two sub samples. The growth of the daily component weight could be responsible for the increase in the short period memory observed

HAR-RC(3) ESTIMATED ON THE S&P - US BOND CORRELATION
Included observations: 3391
Newey-West Standard Errors and Covariance (lag=5)
R-squared = 0.8141

Variable	Coefficient	t - Statistic	Probability
constant	0.000193	0.070422	0.943862
$RC^{(d)}$	0.266740	9.744576	0.000000
$RC^{(w)}$	0.364659	8.904585	0.000000
$RC^{(m)}$	0.348939	9.185706	0.000000

IN SAMPLE MINCER-ZARNOWITZ REGRESSION

	b_0	b_1	R^2
HAR-RC(3)	-0.000025 (-0.0053, 0.0053)	1.000008 (0.9839 1.0161)	0.8143

Table 4.3: In sample estimation results of the Newey-West adjusted least-squares regression of HAR-RC(3) model for the S&P - US Bond realized correlation from 1990 to 2003.

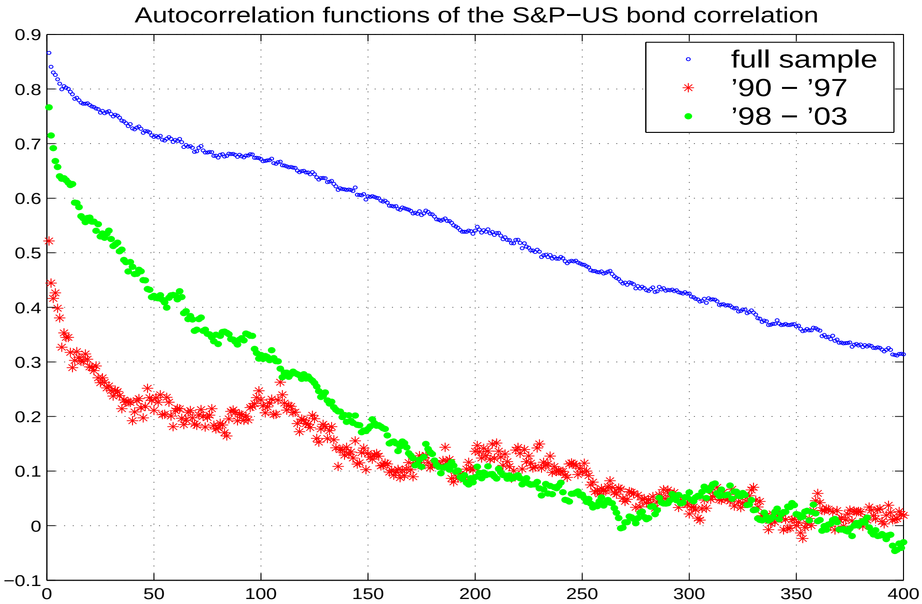


Figure 4.6: Autocorrelation functions of the S&P 500 US bond correlation for the full sample 1990-2003 and the two sub samples 1990-1997 and 1998-2003.

in the autocorrelation function of the second part of the sample. This analysis shows how the identification of the different market components and the study of their dynamics, which is made possible by the HAR model, can help in explain (and maybe also predict) interesting properties and dynamics of financial data.

4.5 Realized Correlation matrix

Since the realized estimators described so far are consistent even in the presence of market microstructure noise, the resulting covariance or correlation matrix will be, asymptotically, definite positive.

However, in small sample, when the number of ticks is much smaller than the one required for asymptotic convergence, realized quantities might also carry a significant amount of measurement error. Measurement errors on the elements of the covariance or correlation matrix implies that the largest sample eigenvalues of the matrix are biased upwards, while the smallest ones are biased downwards. Then in some cases the smallest eigenvalues can become negative so that the matrix will no longer be definite positive.

A general and convenient way to address this problem would be through the so called

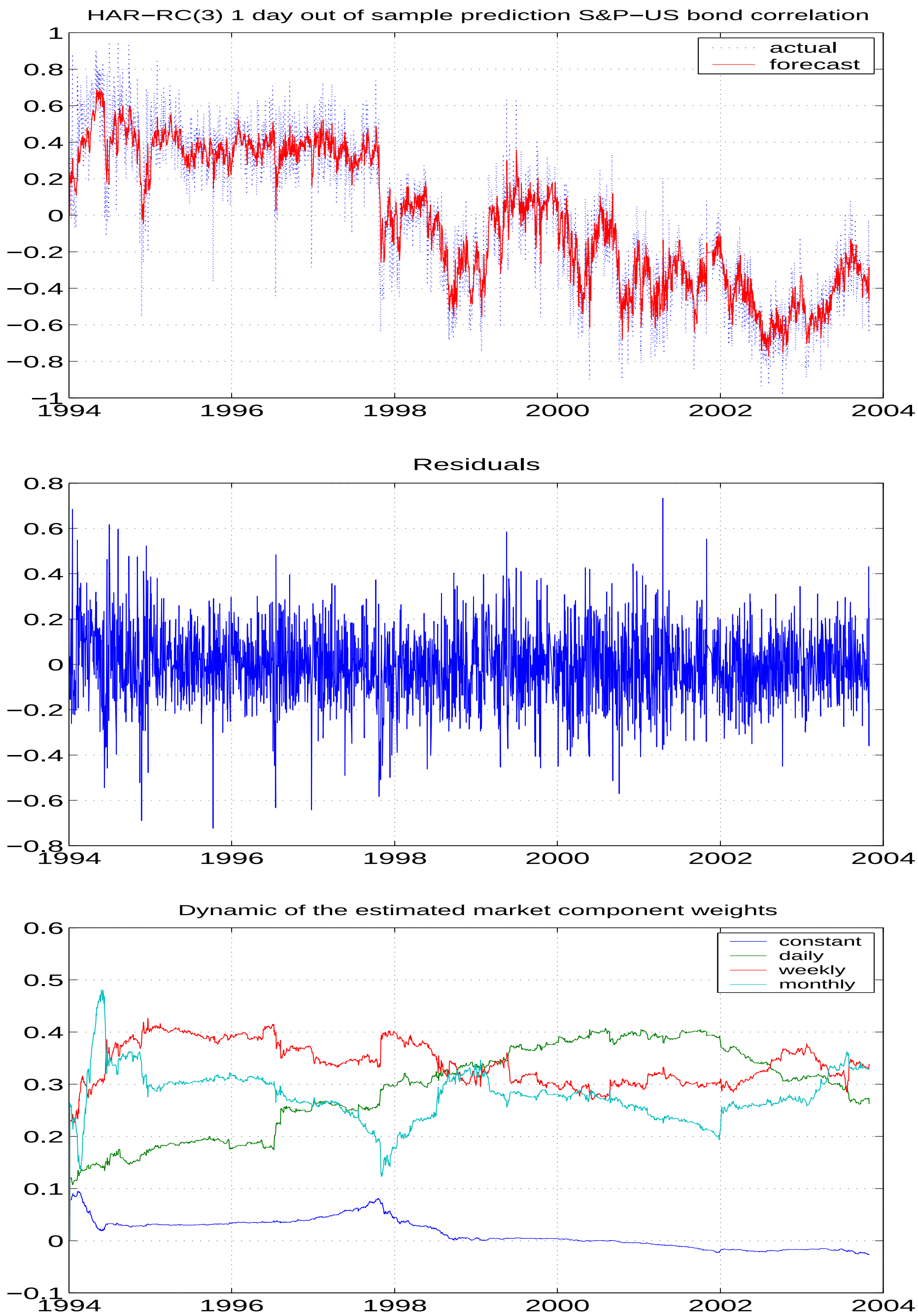


Figure 4.7: Top: comparison of actual (dotted) and out of sample prediction (solid) of the HAR-RC(3) model for daily realized correlations. Middle: residuals. Bottom: time evolution of the regression coeffiecients which, according to the model, represent market component weights.

shrinkage estimator proposed by Frost and Savarino (1986) and Ledoit and Wolf (2003, 2004b) for the covariance matrix and by Audrino and Barone-Adesi (2004) for the conditional correlation matrix. The problem of the sample eigenvalues being more dispersed than true ones, can in fact be addressed by shrinking the sample eigenvalues towards their cross-sectional mean. This eigenvalue shrinkage can be achieved by a simple linear combination of the sample covariance or correlation matrix with that of a structured model.

Moreover, the shrinkage idea can also be interpreted as a way to balance the trade-off between the bias of the highly structured matrix and the variance of the sample estimator. This balance is achieved by appropriately choosing the weight given to the structured model, the so called shrinkage intensity or shrinkage constant. Hence, by combining these two 'extreme' estimators is possible to reduce the estimation error of the covariance or correlation matrix, obtaining a 'compromise' estimator that performs better than either extremes. Therefore, the application of the shrinkage technique is always beneficial even when the positiveness of the estimated matrix is not a concern.

In implementing the shrinkage procedure, two key choices has to be made: first, toward which structural model to shrink, i.e. choosing the shrinkage target and second, how intensely, that is, determining the value of the shrinkage constant.

The shrinkage target should fulfill two different requirements: posses small or no estimation error (that is, having a lot of structure) but also reflects important characteristics of the unknown quantity being estimated. Ledoit and Wolf (2004b) suggest the identity matrix, Ledoit and Wolf (2004a) and Audrino and Barone-Adesi (2004) recommend the constant correlation matrix[3], while Frost and Savarino (1986) and Ledoit and Wolf (2003) propose the single-factor matrix of Sharpe (1963) as shrinkage target. The single factor model has the advantage to be more flexible and able to capture the main feature of financial data but it requires the knowledge of the nature and dynamics of the factor and the time series estimation of the factor loadings.

Here we propose an alternative shrinking target for the sample realized correlation matrix which try to combine the greater flexibility of the one factor model with the easy of estimation and implementation of the constant correlation and identity matrix. The idea stems directly from the observation that the one factor model applied to standardized residual of the returns of either the assets $\varepsilon_i = r_i/\sigma_i$ and the factor $\varepsilon_f = r_f/\sigma_f$

$$\varepsilon_{i,t} = \beta_i \varepsilon_{f,t} + e_{i,t} \tag{4.9}$$

[3]A covariance matrix where all the pairwise correlations are identical and equal to the cross sectional average of all the sample correlations.

has a variance-covariance matrix of the form

$$F = \begin{bmatrix} 1 & \beta_1\beta_2 & \dots & \beta_1\beta_N \\ \beta_2\beta_1 & 1 & \dots & \beta_2\beta_N \\ \vdots & & \ddots & \vdots \\ \beta_N\beta_1 & \beta_N\beta_2 & \dots & 1 \end{bmatrix}$$

since $\text{Var}(\varepsilon_i) = \beta_i^2 + \sigma_{ei}^2 = 1$ by construction and $\text{Cov}(\varepsilon_i\varepsilon_j) = \beta_i\beta_j$, therefore, it does not depends on the underling factor. This structure suggests the possibility to estimate the vector of factor loadings β by simply minimazing the difference between any generic off diagonal elements of the matrix F with the corresponding elements of the sample realized correlation matrix S, that is:

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{N} \sum_{j \neq i} (\beta_i\beta_j - s_{i,j})^2 \qquad s.t. \qquad | \beta_i | < 1 \qquad \forall i \tag{4.10}$$

with $s_{i,j}$ the generic element of S. Hence, the estimation of the factor loadings is carried out without any knowledge of the nature and dynamic of the underling factor. The minimization algorithm in (4.10) projects the sample correlation matrix into the space spanned by single factor models for standardized residual by searching for the one factor model correlation matrix which is closest (in a mean square sense) to the observed realized correlation matrix at day t. With this approach is then possible to impose a single factor structure to the shrinkage target of the correlation remaining completely agnostic on the identification and estimation of that factor.

The shrinkage target $\hat{F}$ will then be a symmetric matrix with diagonal of ones and the generic off diagonal element equals to $\hat{\beta}_i\hat{\beta}_j$. With this shrinkage target the shrinkage estimator becomes

$$R = \alpha\hat{F} + (1-\alpha)S \tag{4.11}$$

where α represents the shrinkage constant.

The second step is to choose the shrinkage intensity α. Ledoit and Wolf (2003a), minimizing the Frobineus distance between the shrinkage estimator and the true covariance matrix, derive the expression for the optimal shrinkage intensity in the form $\alpha = k/T$ with k being a constant and T the number of observations. They also provide a way to consistently estimate the constant k directly from the data.

Expressing the optimal shrinkage constant in the form of this ratio, makes explicit the dependence of the shrinkage intensity α from T. In fact, the shrinkage intensity depends, among other things, on the estimation error contained in the sample estimator. So, when the number of observations is higher the shrinkage intensity will tend to be smaller. This representation of α is especially convenient in our framework since (contrary to the case considered in Ledoit and Wolf where daily returns are used) the employment of tick-by-tick data makes the values of T changes from asset to asset depending on their liquidity. In general, we will need to shrink more the realized correlation obtained with less observations than those computed from highly liquid assets. Therefore, instead of applying the same shrinkage constant k/T to all the elements of the correlation matrix, we construct a "weightening" matrix Γ having a generic elements $\Gamma_{i,j} = 1/\min(T_i, T_j)$, with T_i the number of ticks for the asset i.

So, the correlation shrinkage estimator with the optimal shrinkage matrix Γ would have the form

$$R = \hat{k}\,\Gamma \circ \hat{F} + \left(\iota\,\iota' - \hat{k}\,\Gamma\right) \circ S \tag{4.12}$$

where ι is a vector of ones of length N and $\circ$ the element by element Hadamard product.

This shrinkage procedure will substantially increase the efficiency of the estimator and reduce the probability of having a negative definite correlation matrix but formally does not guarantee the estimated correlation to be always definite positive. However, even in the rare cases when the optimal shrinkage intensity produces a negative definite matrix, the positivity of the shrinkage target guarantees that it is always possible to obtain a definite positive correlation matrix by appropriately increasing the shrinkage intensity. In such cases, the definite positiveness of the correlation matrix would come at the cost of some efficiency loss.

Preliminary simulations results on the estimation of the realized correlation matrix (table 4.4) show that in the presence of highly liquid assets, when the realized correlation matrix is already precise and the probability to be negative definitive minor, the shrinkage intensity is very small and the difference between the shrinkage estimator and the sample one is negligible. But, when the average numbers of observations per day decreases, the optimal shrinkage becomes very helpful and effective in reducing the estimation error and the probability of getting a negative definite matrix. Moreover, even in such rare cases when the optimal shrinkage still gives a negative definite matrix, the required increase in the shrinkage intensity to obtain a definite positive matrix is usually quite small, so that the efficiency loss remain also small.

Having constructed an appropriate realized correlation matrix we can now extend the univariate HAR model for a single correlation to the multivariate case. The simplest way is

CORRELATION MATRIX ESTIMATION

Number of tick per day	5-min	all ticks	optimal shrinkage
1800-2200	1.1068 (0%)	0.5028 (0%)	0.4957 (0%)
1200-800	1.1374 (0%)	0.7297 (0%)	0.6931 (0%)
400-600	1.2537 (0%)	1.035 (1.2%)	0.883 (0%)
200-300	1.463 (3.4%)	1.503 (22.4%)	1.274 (0.4%)

Table 4.4: Average Frobenious distance from the true correlation matrix and percentage of failings (in parethesis) for estimated correlation matricies computed with 5 minutes returns, all tick-by-tick returns and its optimal shrinkage correction. The simulation is repeated 500 times for 10 assets having noise to signal ratios of 0.6.

to apply the HAR model to the realized correlation matrix so that

$$\hat{R}_{t+1} = C + B^{(d)} \circ R_t^{(d)} + B^{(w)} \circ R_t^{(w)} + B^{(m)} \circ R_t^{(m)} \tag{4.13}$$

where $R_t^{(d)}$, $R_t^{(w)}$, and $R_t^{(m)}$ are respectively the daily, weekly and monthly realized correlation matrix, the C, $B^{(d)}$, $B^{(w)}$ and $B^{(m)}$ are NxN matrix of coefficients and $\circ$ is the Hadamard product. Since we have achieved the positivity of $R_t^{(d)}$ (and consequently that of $R_t^{(w)}$ and $R_t^{(m)}$) through the shrinkage procedure, $\hat{R}_{t+1}$ will be definite positive as long as C , $B^{(d)}$, $B^{(w)}$ and $B^{(m)}$ will also be definite positive. Moreover, to ensure the forecast matrix $\hat{R}_{t+1}$ to be, not only definite positive, but also a true correlation matrix, it is necessary that for each individual correlation equation the corresponding parameters satisfy $C_{i,j} + B_{i,j}^{(d)} + B_{i,j}^{(w)} + B_{i,j}^{(m)} \leq 1$.

Many extensions of this simple model can be envisage: include realized volatilities at different frequencies as explanatory variables for correlations or add matrices of cross-product returns measured over the three different horizons with possibly different coefficients for the positive and negative values to account for asymmetric effects.

4.6 Conclusions and directions for future research

In this chapter we have extended the approach of directly using all the available tick-by-tick data to the realized covariance and realized correlation estimation. As for the realized volatility, the presence of market microstructure (although of a different nature) can induce significant bias in standard realized covariance measure computed with artificially regularly spaced returns. Contrary to these standard approaches we propose a very simple and unbiased realized covariance estimator which does not resort on the construction of a regular grid, but directly and efficiently employs the raw tick-by-tick returns of the two series. Montecarlo simulations show that this simple tick-by-tick covariance posses no bias and the smallest dispersion, resulting to be the best performing measure among the covariance estimators considered in the study. Combining the Firs-Last covariance together with the DST volatility estimator we obtain a realized correlation measure where both the volatilities and the covariances are computed from tick-by-tick data. In the empirical analysis performed on S&P 500 and US bond data we apply the HAR model to the tick-by-tick correlation measure obtaining highly significant coefficients for all the three heterogeneous components and remarkably good out of sample forecasting performance. We then suggest the use of a shrinkage approach with a newly proposed shrinkage target for the construction of a definite positive and more accurate correlation matrix.

These promising results on the volatility and correlations modelling together with its simplicity and easy of estimation, makes the HAR model susceptible to be readily extended to the multivariate setting. Its autoregressive structure would suggest that a natural multivariate extension could be the development of a Vector-HAR analogously to the standard VAR model. However, due to the curse of dimensionality, a general Vector-HAR type of model would quickly become intractable as soon as the dimension of the portfolio increases even to moderate sizes. A standard simplifying assumption made in many multivariate GARCH model (including the recent DCC of Engle 2002) to deal with this problem is to assume that each conditional volatility and pairwise correlation depends only on its own history and not on that of the other series (although this dependence can be fitted with a different set of parameters for each series).

Under this assumption, a relatively parsimonious Multivariate-HAR model could be obtained combining, the Asymmetric-HAR model for the volatilities (3.14) with that for correlations (4.13) through the usual volatility-correlation decomposition of the covariance matrix. In this way, we would obtain a DCC style Multivariate-Asymmetric-HAR (MAHAR) model for realized variations. This MAHAR model would combine the flexibility of a DCC structure with the precision of realized variation measures and the easy of estimation of HAR models since only OLS regressions equation by equation would be required to estimate the whole multivariate model.

Then, following Barone-Adesi, Bourgoin, and Giannopoulos (1998) and Barone-Adesi, Engle and Mancini (2004) the HAR and MAHAR model could be used in the Filtered Historical Simulation method to estimate more accurate portfolio risk measures or compute option prices under incomplete markets framework.

Bibliography

Ahn, D.-H., Boudoukh, J., Richardson, M. and Whitelaw, R.: 1999, Behavioralize this! international evidence on autocorrelation patterns of stock index and futures returns, *Technical Report 99-040*, New York University, Leonard N. Stern School of Business-. available at http://ideas.repec.org/p/fth/nystfi/99-040.html.

Aït-Sahalia, Y., Mykland, P. A. and Zhang, L.: 2005, How often to sample a continuous-time process in the presence of market microstructure noise, *Review of Financial Studies* **18**, 351–416.

Alizadeh, S., Brandt, M. W. and Diebold, F. X.: 2002, Range-based estimation of stochastic volatility models, *Journal of Finance* **57**(3), 1047–90.

Andersen, T. G. and Bollerslev, T.: 1997, Heterogeneous information arrivals and return volatility dynamics: Uncovering the long run in high frequency data, *Journal of Finance* **52**, 975–1005.

Andersen, T. G. and Bollerslev, T.: 1998, Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International Economic Review* **39**, 885–905.

Andersen, T. G., Bollerslev, T. and Diebold, F. X.: 2003, Some like it smooth, and some like it rough: Untangeling continous and jumps component in measuring, modelling, and forecasting asset return volatility. Duke University NC USA.

Andersen, T. G., Bollerslev, T. and Diebold, F. X.: 2005, Parametric and nonparametric volatility measurement, *in* Y. Ait-Sahalia and L. P. Hansen (eds), *Handbook of Financial Econometrics*, North Holland, Amsterdam.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P.: 2001, The distribution of exchange rate volatility, *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P.: 2003, Modeling and forecasting realized volatility, *Econometrica* **71**((2)), 579–625.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Vega, C.: 2003, Micro effects of macro announcements: Real-time price discovery in foreign exchange, *American Economic Review* **93**, 38–62.

Arneodo, A., Muzy, J. and Sornette, D.: 1998, Casual cascade in stock market from the "infrared" to the "ultraviolet", *European Physical Journal B* **2**, 277.

Audrino, F. and Barone-Adesi, G.: 2004, Average conditional correlation and tree structures for multivariate garch models. Working paper, http://ssrn.com/abstract=553821.

Bandi, F. and Russell, J. R.: 2005, Microstructure noise, realized volatility, and optimal sampling. Working Paper, Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O. E., Graversen, S. E., Jacod, J., Podolskij, M. and Shephard, N.: 2005, A central limit theorem for realised power and bipower variations of continuous semimartingales, *in* Y. Kabanov and R. Lipster (eds), *From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev*, Springer. Forthcoming.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N.: 2004, Regular and modified kernel-based estimators of integrated variance: The case with independent noise. Working paper, http://ssrn.com/abstract=620203.

Barndorff-Nielsen, O. E. and Shephard, N.: 2001, Non-gaussian ornstein-uhlembech-based models and some of their uses in financial economics, *Journal of the Royal Statistical Society* **Series B**(63), 167–241.

Barndorff-Nielsen, O. E. and Shephard, N.: 2002a, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society* **Series B**(64), 253–280.

Barndorff-Nielsen, O. E. and Shephard, N.: 2002b, Estimating quadratic variation using realized variance, *Journal of Applied Econometrics* **17**(5), 457–477.

Barndorff-Nielsen, O. E. and Shephard, N.: 2005, How accurate is the asymptotic approximation to the distribution of realized volatility?, *in* D. W. F. Andrews and J. H. Stock (eds), *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg*, Cambridge University Press, pp. 306–331.

Barone-Adesi, G., Bourgoin, F. and Giannopoulos, K.: 1998, Dont look back, *Risk* **11**, 100–103.

Barone-Adesi, G., Engle, R. F. and Mancini, L.: 2004, Garch options in incomplete markets. NCCR-FinRisk Paper; University of Zurich Paper No. 155, http://ssrn.com/abstract=603382.

Blume, M. and Stambaugh, S.: 1983, Biases in computed returns, *Journal of Financial Economics* **12**, 387–404.

Bollen, B. and Inder, B.: 2002, Estimating daily volatility in financial markets utilizing intraday data, *Journal of Empirical Finance* **186**.

Bollerslev, T. and Zhang, B. Y. B.: 2003, Measuring and modeling systematic risk in factor pricing models using high-frequency data, *Journal of Empirical Finance* **10**(5), 533–558.

Breymann, W., Ghashghaie, S. and Talkner, P.: 2000, A stochastic cascade model for fx dynamics, *International Journal of Theoretical and Applied Finance* **3**, 357–360.

Breymann, W., Jegadeesh, N. and Swaminathan, B.: 1993, Investment analysis and the adjustment of stock prices to common information, *Review of Financial Studies* **6**, 799–824.

Breymann, W., Zumbach, G., Dacorogna, M. and Müller, U. A.: 2000, Dynamical deseasonalization in otc and localized exchange-traded markets,, *Technical report*, Internal document WAB.2000-01-31, Olsen & Associates, Seefeldstrasse 233, 8008 Z urich, Switzerland.

Bustos, O. and Yohai, V. J.: 1986, Robust estimates for arma models, *Journal of the American Statistical Association* **81**, 155–68.

Clark, P. K.: 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* **41**(1), 135–55.

Cohen, K., Hawawini, G. A., Maier, S. F., R., R. S. and D., D. W.: 1983, Friction in the trading process and the estimation of systematic risk, *Journal of Financial Economics* **12**, 263–278.

Cohen, K., S., M., R., S. and D., W.: 1981, Transaction costs, order placement strategy, and existence of the bid-ask spread, *Journal of Political Economy* **89**, 287–305.

Comte, F. and Renault, E.: 2001, Long memory in continuous time stochastic volatility models, *Mathematical Finance* **8**, 291–323.

Conrad, J. and Kaul, G.: 1988, Time varying expected returns, *Journal of Business* **61**, 409–425.

Corsi, F.: 2003, Simple long memory models of realized volatility. Manuscript, USI.

Corsi, F.: 2005, Realized correlation tick-by-tick. Manuscript University of Lugano.

Corsi, F., Kretschmer, U., Mittnik, S. and Pigorsch, C.: 2005, The volatility of volatility. Unpublished Manuscript.

Corsi, F., Zumbach, G., Müller, U. A. and Dacorogna, M.: 2001, Consistent high-precision volatility from high-frequency data, *Economic Notes* **30**(2), 183–204.

Curci, G. and Corsi, F.: 2003, A discrete sine transform approach for realized volatility measurement. NCCR FINRISK Working Paper No. 44.

Dacorogna, M. and Lundin, M.: 1999, Correlation of high-frequency financial time series, *in* P. Lequeux (ed.), *Financial Markets Tick by Tick*, Wiley & Sons.

Dacorogna, M. M., Genay, R., Müller, U. A., Olsen, R. B. and Pictet, O. V.: 2001, *An Introduction to High-Frequency Finance*, Academic Press.

Dacorogna, M., Müller, U., Dav, R., Olsen, R. and Pictet, O.: 1998, Modelling short-term volatility with garch and harch models, *In "Nonlinear Modelling of High Frequency Finacial Time Series"* pp. 161–76. edited by C. Dunis and B. Zhou, John Wiley, Chichester.

Dacorogna, M., Müller, U., Nagler, R. J., Olsen, R. B. and Pictet, O. V.: 1993, A geographical model for the daily and weekly seasonal volatility in the fx market, *Journal of International Money and Finance* **12**(4), 413–438.

Ding, Z., Granger, C. and Engle, R.: 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* pp. 83–106.

Engle, R.: 2002, Dynamic conditional correlation - a simple class of multivariate garch models, *Journal of Business and Economic Statistics* **17**(5).

Epps, T.: 1979, Comovements in stock prices in the very short run, *Journal of the American Statistical Association* **74**(366), 291–298.

French, K. R. and Roll, R.: 1986, Stock return variances : The arrival of information and the reaction of traders, *Journal of Financial Economics* **17**(1), 5–26.

Frost, P. A. and Savarino, J. E.: 1986, An empirical bayes approach to portfolio selection, *Journal of Financial and Quantitative Analysis* **21**, 293–305.

Gencay, R. and Selcuk, F.: 2004, Asymmetry of information flow between volatilities across time scales, *Econometric Society 2004 North American Winter Meetings 90*, Econometric Society.

Ghashghaie, S., Breymann, W., Peinke, J., Talkner, P. and Dodge, Y.: 1996, Turbolent cascades in foreign exchange markets, *Nature* **381**, 767–770.

G.Hawawini: 1980, Intertemporal cross dependence in securities daily returns and the short-term intervalling effect on systematic risk, *Journal of Financial and Quantitative Analysis* **15**, 139–149.

Glosten, L.: 1994, Is the electronic open limit-order book inevitable?, *Journal of Finance* **49**, 1127–1161.

Glottlieb, G. and Kalay, A.: 1985, Implication of the discreteness of observed stock prices, *Journal of Finance* **40**, 135–153.

Goodhart, C. A. E.: 1989, news and the foreign exchange market, *Technical report*, Proceedings of the Manchester Statistical Society.

Goodhart, C. A. E.: 1991, Every minute counts in financial markets, *Journal of International Money and Finance* **10**, 23–52.

Granger, C.: 1980, Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* **14**, 227–38.

Hamilton, J.: 1994, *Time Series Analysis*, Princeton: Princeton University Press.

Hansen, P. R. and Lunde, A.: 2004, An unbiased measure of realized variance. Working Paper, Department of Economics, Stanford University.

Hansen, P. R. and Lunde, A.: 2005, Realized variance and market microstructure noise, *Journal of Business and Economic Statistics* . Forthcoming.

Harris, L.: 1990, Estimation of stock price variances and serial covariances from discrete observations, *Journal of Financial and Quantitative Analysis* **25**(3), 291–306.

Harvey, A. C.: 1989, *Forecasting, Structural Time Series Models, and the Kalman Filter*, Cambridge University Press, Cambridge.

Hasbrouck, J.: 1993, Assessing the quality of a security market: a new approach to transaction-cost measurement, *Review of Financial Studies* **6**(1), 191–212.

Hasbrouck, J.: 1996, Modelling market microstructure time series, *in* G. S. Maddala and C. R. Rao (eds), *Statistical Methods in Fianance*, Vol. 14 of *Handbook of Statistics*, North Holland, Amsterdam, pp. 647–692.

Hasbrouck, J.: 1999, The dynamichs of discrete bid and ask quotes, *Journal of Finance* **54**, 2109–42.

Hasbrouck, J. and Ho, T.: 1987, Order arrival, quote behavior, and the return-generating process, *Journal of Finance* **42**, 1035–48.

Heston, S.: 1993, A closed-form solution for options with stochastic volatility with applications to bonds and currency options, *Review of Financial Studies* **6**, 327–343.

Holden, A. and Subrahamanyam, A.: 1992, Long-lived private information and imperfect competition, *Journal of Finance* **47**, 247–270.

LeBaron, B.: 2001, Stochastic volatility as a simple generator of financial power-laws and long memory, *Working Paper* **30**.

Ledoit, O. and Wolf, M.: 2003, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* **10**(5), 603–621.

Ledoit, O. and Wolf, M.: 2004a, Honey, i shrunk the sample covariance matrix, *Journal of Portfolio Management* **31**(1).

Ledoit, O. and Wolf, M.: 2004b, A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis* **88**(2), 365–411.

Lo, A. and MacKinlay, A.: 1990, An econometric analysis of nonsynchronous trading, *Journal of Econometrics* **45**, 181–212.

Lo, A. W. and MacKinlay, A. C.: 1988, Stock market prices do not follow random walks: evidence from a simple specification test, *The Review of Financial Studies* **1**, 41–66.

Lobato, I. and Savin, N.: 1998, Real and spurious long-memory properties of stock market data, *Journal of Business and Economics Statistics* **16**, 261–283.

Lux, T.: 1996, Long-term stochastic dependence in financial prices: evidence from the german stock market, *Applied Economics Letters* **3**, 701–6.

Lux, T. and Marchesi, M.: 1999, Scaling and criticality in a stochastic multi-agent model of financial market, *Nature* **397**, 498–500.

Lynch, P. E.: 2000, *Adaptive models for speculative turbulence*, PhD thesis, Department of Electrical and Electronic Engineering, University of Manchester Institute of Science and Technology. Manchester, M60. UK.

Maheu, J. M. and McCurdy, T. H.: 2002, Nonlinear features of realized fx volatility, *Review of Economics and Statistics* . forthcoming.

Martens, M.: 2004, Estimating unbiased and precise realized covariances. Social Science Research Network Electronic Library.

Martens, M. and Zein, J.: 2002, Predicting financial volatility: High-frequency time-series forecasts vis--vis implied volatility. Manuscript, Econometric Institute and Department of Finance, Erasmus University, Rotterdam.

Merton, R. C.: 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* **8**, 323–61.

Mills, T.: 1996, Stylized facts of the temporal and distributional properties of daily ft-se returns, *Applied Financial Economics* **7**, 599–604.

Müller, U., Dacorogna, M., Dav, R., Olsen, R., Pictet, O. and von Weizsacker, J.: 1997, Volatilities of different time resolutions - analysing the dynamics of market components, *Journal of Empirical Finance* **4**, 213–39.

Müller, U., Dacorogna, M., Dav, R., Pictet, O., Olsen, R. and Ward, J.: 1993, Fractals and intrinsic time - a challenge to econometricians. XXXIXth International AEA Conference on Real Time Econometrics, 14-15 Oct 1993, Luxembourg.

Muzy, J., Delour, J. and Bacry, E.: 2000, Modelling fluctuations of financial time series: from cascade process to stochastic volatility model, *Preprint cond-mat/0005400* .

Oomen, R.: 2006, Properties of realized variance under alternative sampling schemes, *Journal of Business and Economic Statistics* . Forthcoming.

Oomen, R. C.: 2002, Modelling realized variance when returns are serially correlated. Working paper, Warwick Business School.

Oomen, R. C.: 2005, Properties of bias-corrected realized variance under alternative sampling schemes, *Journal of Financial Econometrics* **3**(4), 555–577.

Parkinson, M.: 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business* **53**, 61–65.

Pasquini, M. and Serva, M.: 2000, Clustering of volatility as a multiscale phenomenon, *European Physical Journal B* **16**, 195.

Peters, E.: 1994, *Fractal Market Analysis*, A Wiley Finance Edition. John Wiley & Sons, New York.

Renó, R.: 2003, A closer look at the epps effect, *International Journal of Theoretical and Applied Finance* **6**(1), 87–102.

Roll, R.: 1984, A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance* **39**, 1127–1139.

Schmitt, F., Schertzer, D. and Lovejoy, S.: 1999, Multifractal analysis of foreign exchange data, *Applied Stochastic Models and Data Analysis* **15**, 29–53.

Scholes, M. and Williams, J.: 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics* **5**, 181–212.

Schwert, G. W.: 1998, Stock market volatility: Ten years after the crash, *Technical report*, Brookings-Wharton Papers on Financial Services.

Taylor, S. J. and Xu, X.: 1997, The incremental volatility information in one million foreign exchange quotations, *Journal of Empirical Finance* **4**, 317–340.

Thomakos, D. D. and Wang, T.: 2002, Realized volatility in the futures markets, *Journal of Empirical Finance* . forthcoming.

Vandewalle, N. and Ausloos, M.: 1998a, Multi-affine analysis of typical currency exchange rates, *European Physical Journal B* **4**, 257–61.

Vandewalle, N. and Ausloos, M.: 1998b, Spasrseness and roughness of foreign exchange rates, *International Journal of Modern Physics* **9**, 711–9.

Zhang, L.: 2004, Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Discussion Paper, Carnegie-Mellon University.

Zhang, L., Aït-Sahalia, Y. and Mykland, P. A.: 2005, A tale of two time scales: determining integrated volatility with noisy high frequency data, *Journal of the American Statistical Association* (100), 1394–1411.

Zhou, B.: 1996, High-frequency data and volatility in foreign-exchange rates, *Journal of Business & Economic Statistics* **14**(1), 45–52.

Zumbach, G. O., Corsi, F. and Trapletti, A.: 2002, Efficient estimation of volatility using high frequency data, *Working Paper, Olsen & Associates* .

Zumbach, G. O. and Lynch, P.: 2001, Heterogeneous volatility cascade in financial markets, *Forthcoming in Physica A* .