On Time-Varying Parameter Models and their Applications to High-Frequency Financial Data

Giuseppe Buccheri

Scuola Normale Superiore

Classe di Scienze

PhD Thesis

Corso di perfezionamento in Matematica per la Finanza

Advisors:

Prof. Dr. Giacomo Bormetti

Prof. Dr. Fulvio Corsi

Prof. Dr. Fabrizio Lillo

Contents

Acknowledgements

I am indebted to my supervisors, Giacomo Bormetti, Fulvio Corsi and Fabrizio Lillo, who introduced me to the topics covered by this thesis. All the four projects that came out as a result of my research activity as a PhD student at Scuola Normale Superiore benefited enormously from their expert advices. The time spent on numerous meetings with them was very important and much appreciated.

I am also indebted to Rosario Mantegna and Stefano Marmi. I started to collaborate with them well before the beginning of my PhD and they played an important role in my decision to switch from Physics to Quantitative Finance, as well as in my decision to enroll in a PhD program at Scuola Normale Superiore.

Another thank goes to all my colleagues and former members of the Quantitative Finance research group at Scuola Normale. I had fruitful research discussions especially with Giulia Livieri, Davide Pirino and Luca Trapin.

I am grateful to the R&D group at UniCredit S.p.A., who financed my PhD scholarship within the project "Dynamics and Information Research Institute - Quantum Information, Quantum Technologies". Last but not least I thank my family, who always supported the career path that I have chosen.

Pisa, September 2018

Chapter 1

Introduction

Time-varying parameter models are ubiquitous in finance and economics. A well-known property of financial returns is volatility clustering, meaning that large changes tend to be followed by large changes and, similarly, small changes tend to be followed by small changes. This phenomenon is a consequence of the fact that volatility is not constant over time but has non-trivial dynamics. Both ARCH-type and stochastic volatility models (Engle 1982, Bollerslev 1986, Tauchen and Pitts 1983) have been employed to describe time-varying volatilities in financial markets. In a multivariate setting, it is a well documented empirical evidence that financial data exhibit strong changes in their dependence structure, especially in periods of market turmoil. To capture this behaviour, different dynamic correlations and copula models have been proposed (cfr. Engle 2002a, Shephard 2005, Patton 2006). High-frequency financial data feature time-varying volatilities and correlations at the intraday level (Andersen and Bollerslev 1997, Tsay 2005, Bibinger et al. 2014). Empirical evidence of time-varying VAR coefficients and stochastic volatility have also been found in the macroeconomic literature, as shown by Cogley and Sargent (2001), Cogley and Sargent (2005), Stock and Watson (2007), amongst others.

This thesis provides new contributions to the field of time-varying parameter models[1] from both theoretical and empirical perspectives. On the one side, we introduce a general smoothing methodology that is named Score-Driven Smoother (SDS). As it will be shown in Chapter 3, the SDS provides smoothed estimates of time-varying parameters of a large class of nonlinear non-Gaussian models. On the other side, we propose new time-varying parameter models capturing relevant empirical features of high-frequency price dynamics. Econometric inference on such models helps answering key questions arising in the field of high-frequency finance and market microstructure.

The SDS is related to the framework of so-called observation-driven models (Cox 1981), that is, time-varying parameter models where the sources of randomness driving parameter changes are nonlinear functions of past observations. A common example is given by the GARCH model of Bollerslev (1986),

[1] We mainly concentrate on discrete-time models.

where the conditional variance depends on squares of past returns. Observation-driven models are typically viewed as data generating processes. However, they can also be employed as misspecified filters. For instance, Nelson 1992 examined the diffusion limit of GARCH models in a setting where the latter are employed as nonlinear filters for stochastic volatility models. If observation-driven models are regarded as data generating processes, there is no room for smoothing, as all relevant information is encoded on *past* observations. In contrast, if they are employed as filters, smoothing can help improving the estimation of time-varying parameters, as shown by Nelson 1996 in the case of the GARCH. While many observation-driven models have been proposed (e.g. all univariate and multivariate GARCH-type models), the literature lacks of corresponding "observation-driven smoothers", allowing to estimate parameters using *actual* and *future* observations. The SDS fills this gap, in that it provides a general methodology to exploit all observations when estimating time-varying parameters.

In our theoretical and empirical work on high-frequency finance, we address relevant research questions related to high-frequency price dynamics and market microstructure. In particular, we focus on the following key points:

1. It is well-known that high-frequency asset prices are characterized by lead-lag effects, meaning that some assets lead the dynamics of other assets. How can we motivate the existence of these effects from a market microstructure perspective and how can we estimate lead-lag correlations by accounting for asynchronous trading and microstructure noise (e.g. price discreteness, bid-ask bounds, etc)?

2. What are the intraday dynamics of dependencies among high-frequency asset prices and how can we forecast intraday correlations by accounting, again, for asynchronous trading and microstructure noise?

3. Do measurement errors on realized volatility measures play a role in volatility estimation and forecasting? Can we disentangle the effect of measurement errors from the one of non-linearity?

The above questions arise from the increasing availability of high-frequency financial data, which nowadays are recorded at very small time-scales (e.g. milliseconds). On the one hand, this leads to the possibility of examining in detail the intraday dynamics of high-frequency prices, allowing to build new data-driven models for market microstructure analysis. However, intraday data are unbalanced, with trades arriving at irregular and random times. In addition, they are contaminated by observational noise, due to microstructure effects. Standard univariate and multivariate techniques are therefore not effective in extracting the meaningful economic content of the data. On the other hand, since the seminal work of Andersen et al. (2003), high-frequency data have been largely employed to model and forecast volatility through realized measures. Time series of realized volatility exhibit strong nonlinear dynamics and are

characterized by measurement errors. Neglecting both effects may lead to biased estimates and poor volatility forecasts (Corsi et al. 2008, Bollerslev et al. 2016a). We design new time-varying parameter models apt to tackle these modelling issues and provide answers to previous questions.

This thesis is divided in five parts. The first part, in Chapter 2, is a brief overview of time-varying parameter models. It introduces a notation and a nomenclature that will be extensively used throughout the work. Moreover, it lists the main results related to the Kalman filter and recently introduced score-driven models (Creal et al. 2013, Harvey 2013). The latter are a large class of observation-driven models including, but not limited to, the GARCH, the EGARCH model of Nelson (1991), the MEM model of Engle (2002b) and the ACD model of Engle and Russell (1998). Both Kalman filter and score-driven models are the main building blocks of the theoretical and empirical work that is illustrated in the subsequent Chapters.

Chapter 3 deals with the SDS. The main theoretical idea conveying to the formulation of the SDS is the following result, which we prove in Section 3.1.1:

Proposition. *In the steady state, the standard Kalman filter and smoothing recursions for linear Gaussian models can be re-written in a general form that only involves the score and the Fisher information of the Gaussian conditional density.*

By computing scores and information based on a non-Gaussian density, these new recursions can be regarded as the approximate filtering and smoothing recursions for a generic nonlinear non-Gaussian model. The predictive filtering recursion turns out to have the form of score-driven models, which in turn are approximate nonlinear filters. More interestingly, the update filter and smoothing recursions lead to a new score-driven update filter (SDU) and the SDS. The SDU updates filtered estimates once new observations become available while the SDS updates filtered estimates by including the effect of future observations.

In this new framework, we can assess filtering uncertainty and construct confidence bands around filtered and smoothed estimates. Filtering uncertainty is absent in correctly specified observation-driven models, as time-varying parameters are deterministic functions of past observations. However, in misspecified observation-driven filters, time-varying parameters are not fully revealed by past observations and therefore filtering uncertainty matters. In the literature, there are no well-established methods to quantify filtering uncertainty in misspecified observation-driven filters. As a byproduct of our results, we provide a general methodology to construct robust confidence bands reflecting filtering uncertainty in general score-driven models.

Given standard score-driven filter recursions, companion SDU and SDS recursions can be computed in order to update filtered estimates based on actual and future observations. Compared to standard smoothing techniques for nonlinear non-Gaussian models, which are based on computationally demand-

ing simulation-based methods, the SDS is a simple backward recursion following the standard forward filtering recursion of score-driven models.

In our simulation and empirical work, we showed the following:

- The SDU and SDS improve significantly over standard score-driven filtered estimates. In particular, for the SDS, we found gains larger than 30% and lower than 60% in mean square errors when estimating time-varying parameters in a large class of misspecified observation-driven models (GARCH, MEM, time-varying AR(1), t-GAS, Wishart-GARCH).

- Confidence bands around filtered and smoothed estimates correctly quantify the true filtering uncertainty.

- In a nonlinear non-Gaussian setting, the SDS does not lead to an excessive deterioration of the reconstructed signal compared to exact methods. Indeed, comparing the performance of our approximate smoother to that of correctly specified parameter-driven models, we found that losses are small, never exceeding 2.5% on average.

- The SDU and SDS are computationally convenient. For instance, we found that smoothing with the SDS is on average more than 200 times faster than smoothing with efficient importance sampling techniques.

- The SDS and SDU can be easily extended to a framework with multiple time-varying parameters, as they maintain the same simple form as in the univariate case.

- Empirically, it turns out that SDU and SDS estimates of daily conditional covariance matrices are superior to those based on standard score-driven filtered estimates, which are often used in practical applications.

In Chapter 4, we motivate the existence of cross-asset pricing effects observed at high-frequency (Hasbrouck and Seppi 2001, Bernhardt and Taub 2008, Pasquariello and Vega 2015) through a new Multi-asset Lagged Adjustment (MLA) price formation model that generalizes standard univariate microstructure models of lagged price adjustment (Hasbrouck 1996). In the MLA, the price formation process of a given asset is influenced by the price of other assets. Lead-lag correlations naturally arise as a consequence of the multivariate nature of the price formation process. We thus establish a link between the financial literature on cross-asset pricing and the econometric literature on high-frequency lead-lag correlations. The MLA represents the first multivariate extension of well known univariate models of lagged price-adjustment. In this new framework, lead-lag effects can be explained from a clear market microstructure perspective.

Estimation of the MLA is robust to both microstructure effects and asynchronous trading. Indeed, it can be estimated through a Kalman-EM algorithm that allows latent prices being contaminated by additive noise and that easily tackles asynchronicity as a missing value problem. The main advantages of the MLA are the following:

- In contrast to existing estimators of lead-lag correlations, the MLA is robust to spurious correlations arising under asynchronous trading. This is shown in detail in the simulation study in Section 4.2.1.
- The MLA is able to disentangle contemporaneous and lagged correlations. As such, it provides an estimate of the integrated covariance of the efficient martingale process that is robust to microstructure noise, asynchronous trading and that takes into account the presence lead-lag correlations.

The latter is a novel contribution, given that estimators of the integrated covariance typically neglect lead-lag dependencies. We provide strong empirical evidence for the existence of a multi-asset price formation mechanism in equity data. In particular, we find that cross-asset effects surge in periods of high volatility. This empirical result can be explained through the behavior of high-frequency traders, who exploit short living arbitrage opportunities that are likely to appear in periods of high uncertainty.

In Chapter (5), we introduce a multivariate conditional correlation model that can handle noisy and asynchronous observations. As such, the model is suitable for investigating the intraday dynamics of correlations among high-frequency data. This topic is interesting from both theoretical and practical reasons. However, it has received little attention in the econometric literature (Bibinger et al. 2014, Koopman et al. 2015a). We propose to model intraday prices through a multivariate Local Level model with Score-Driven (LLSD) time-varying covariances. In the LLSD, the latent random walk represents the efficient price, whose covariances evolve in time based on the score of the conditional density.

The main advantage of this approach is that, even in presence of time-varying parameters, the LLSD is conditionally Gaussian. The Kalman filter can therefore be applied to write both the log-likelihood and the score in closed form. The consequence is that asynchronous trading can be treated as a typical missing value problem, in a similar fashion to the MLA. This implies that we can avoid synchronizing prices, a procedure that generally leads to significant data reduction, especially at large dimensions. Even at very small time-scales (e.g. 1 second), the dynamics of covariances can thus be reconstructed using all available data. We adopt a parametrization of the correlation matrix based on hyperspherical coordinates that guarantees positive-definite estimates. Similarly to DCC models, large dimensionality is handled by separating the estimation of correlations from individual volatilities. The consistency of this two-step estimation approach is studied through Monte-Carlo simulations.

Empirically, we find the following results:

- Volatilities have the typical U-shape: they are large at the beginning and at the end of the trading

day

- Correlations have an increasing pattern during the day. They are found to be small at the beginning of the day and progressively increase until the last minutes, when they tend to decrease

While the intraday behavior of volatilities is a well-established stylized fact, the intraday pattern of correlations is less known. Our result can be explained by the prevalence of idiosyncratic trading in the first part of the day, leading to large idiosyncratic risk and low correlations. A market factor progressively emerge in the second part of the trading day, with all cross-correlations rapidly increasing.

Remarkably, the LLSD updates filtered estimates once new, unbalanced data become available and thus it can track in real-time the response of market covariances to external information, e.g. macro-news announcements. In order to assess the forecasting ability of the LLSD on real data, we perform an out-of-sample test based on intraday portfolio optimization and show its advantages compared to standard dynamic covariance models.

In the last Chapter, we deal with the problem of modelling and forecasting volatility using high-frequency data. The HAR model of Corsi (2009) is nowadays one of the most popular dynamic specification for realized volatility. However, it assumes volatility being observed with no measurement errors and exhibits several forms of misspecification due to the inherent nonlinear dynamics of volatility. Corsi et al. (2008) and Bollerslev (1986) showed that these two effects can largely jeopardize volatility forecasts. Our aim is to disentangle them and quantify their impact on volatility forecasts. We follow a step by step strategy: we first devise two different HAR extensions that account separately for each effect and then combine the two approaches in a single model that accounts for both measurement errors and nonlinearities.

As a first step, we write the HAR in a linear state-space representation where, consistently with the asymptotic distribution of realized variance, the time-varying variance of the measurement error is driven by realized quarticity. The Kalman filter allows to easily estimate bias-corrected HAR parameters and incorporates the effect of measurement errors through a time-varying Kalman gain. We name this model HAR-Kalman (HARK). The HARK provides more conservative forecasts when current volatility estimates are noisy and generates more responsive forecasts when volatility is estimated with a good accuracy.

In a second step, we exploit the fact that general nonlinear autoregressive models can be represented as linear autoregressive models with time-varying coefficients and introduce an HAR model with time-varying parameters. The dynamics of parameters are driven by the score of the conditional likelihood, thus allowing to write the likelihood in closed form. The resulting model is named score-HAR (SHAR). The SHAR is nonlinear in nature and features iid standardized residuals when estimated on real data.

As a final step, the two approaches are combined in a single model, the score-HAR-Kalman (SHARK)

that accounts for both measurement errors and nonlinear dependencies. "Injecting" nonlinearities in the HARK is possible by modeling time-varying parameters through the score of the conditional likelihood. The SHARK is susceptible of treatment with the Kalman filter and can therefore be estimated through standard maximum likelihood methods.

Our empirical analysis provides strong evidence that the effects captured by the (S)HAR(K) are relevant for volatility forecasting. In particular, we show that measurement errors are important at small and intermediate sampling frequencies and the corresponding forecast gains slightly increase with the forecast horizon. Time-varying parameters provide statistically significant improvements that are independent on the sampling frequency and slightly increase with the forecast horizon.

Up to our knowledge, this methodology is the first achieving the goal of disentangling the effects of measurement errors and nonlinear dependencies. This is a significant contribution, considered that existing approaches dealing with measurement errors on realized volatility may lead to spurious effects as a result of mixing the two effects.

Chapter 2

Time-varying parameter models

2.1 Overview

In its most general form, a discrete time, time-varying parameter model can be written as:

$$y_t|\theta_t \sim p(y_t|\theta_t, \Psi) \tag{2.1.1}$$

$$\theta_t = f(\alpha_t, \Psi) \tag{2.1.2}$$

$$\alpha_t = g(\alpha_{t-1}, \alpha_{t-2}, \ldots, y_{t-1}, y_{t-2}, \ldots, \eta_t, \Psi) \tag{2.1.3}$$

where $y_t \in \mathbb{R}^n$ is a vector of observations with conditional density function given by $p(y_t|\theta_t, \Psi)$, $\theta_t \in \mathbb{R}^q$ is known as "signal" and $\alpha_t \in \mathbb{R}^m$ is a vector of time-varying parameters. The latter can depend on lagged values of both α_t, y_t and on a vector $\eta_t \in \mathbb{R}^m$ of idiosyncratic innovations. All the static parameters are collected in a vector that is denoted by Ψ. Eq. (2.1.1) is known as "measurement equation" while eq. (2.1.3) is known as "transition equation".

Based on the specific form of the function $g(\cdot)$, Cox (1981) distinguished between parameter-driven and observation-driven models. In the first class of models, α_t depends only on its lagged values and on idiosyncratic innovations η_t, namely:

$$\alpha_t = g(\alpha_{t-1}, \alpha_{t-2}, \ldots, \eta_t, \Psi) \tag{2.1.4}$$

meaning that the dynamics of α_t are driven by the random shocks η_t, which are independent on observations y_t.

The simplest example of a parameter-driven model is the linear Gaussian state-space model, which reads:

$$y_t|\theta_t \sim N(\theta_t, H) \tag{2.1.5}$$

$$\theta_t = Z\alpha_t \tag{2.1.6}$$

$$\alpha_{t+1} = c + T\alpha_t + \eta_t, \quad \eta_t \sim N(0, Q) \tag{2.1.7}$$

where $Z \in \mathbb{R}^{n\times m}$, $H \in \mathbb{R}^{n\times n}$, $T \in \mathbb{R}^{m\times m}$ and $Q \in \mathbb{R}^{m\times m}$ are system matrices. The specification with $n = m$, $Z = I_n$, $T = I_n$, $c = 0$ is known as "local-level" model and will be of particular interest in Chapter 5. The Kalman filter, which will be described in detail in Section 2.2, provides optimal estimates of α_t and allows to compute the likelihood in closed form.

Most parameter-driven models are nonlinear and/or non-Gaussian. For instance, a scalar stochastic volatility (SV) model reads:

$$y_t|\sigma_t \sim N(0, \sigma_t^2) \tag{2.1.8}$$

$$\sigma_t^2 = \exp(\alpha_t) \tag{2.1.9}$$

$$\alpha_{t+1} = c + \phi\alpha_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \tag{2.1.10}$$

where y_t are typically log-returns and the signal σ_t^2 is related to the time-varying parameter α_t through an exponential link function. The SV model has been largely applied in financial econometrics, cfr. Tauchen and Pitts (1983), Ghysels et al. (1996) and Shephard (2005).

Except for linear Gaussian models and discrete-state hidden Markov models, no closed form solutions for the exact likelihood of a parameter-driven model are generally available. The reason is that, in order to compute the likelihood, one has to integrate over the unobserved state variables:

$$L(\Psi) = \int p(\alpha_1, \ldots, \alpha_N, y_1, \ldots, y_N, \Psi) \prod_{i=1}^{N} d\alpha_i \tag{2.1.11}$$

where $p(\alpha_1, \ldots, \alpha_N, y_1, \ldots, y_N, \Psi)$ is the joint probability density function of observations and time-varying parameters. The above expression is an high-dimensional integral which requires computationally intense methods such as Monte-Carlo integration. In some cases, approximate filtering and smoothing methods are available, such us the QML method of Harvey et al. (1994) for the SV model. For a detailed analysis on the estimation of nonlinear non-Gaussian state-space models, we refer the reader to Durbin and Koopman (2012).

Observation-driven models are an alternative class of time-varying parameter models. In observation-driven models, time-varying parameters are a deterministic function of past observations:

$$\alpha_t = g(\mathcal{Y}_{t-1}, \Psi) \tag{2.1.12}$$

where $\mathcal{Y}_{t-1} = \{y_1, \ldots, y_{t-1}\}$ denotes the set of past observations and $g(\cdot)$ is typically a nonlinear function. The law of motion (2.1.12) implies that parameters are random, since they are a function of past observations. However, conditionally on $\mathcal{Y}_{t-1}$, they are known with no uncertainty. The GARCH model of Bollerslev (1986) is probably the most popular example of observation-driven models. It reads:

$$y_t|\sigma_t \sim N(0, \sigma_t^2) \tag{2.1.13}$$

$$\sigma_{t+1}^2 = \omega + \alpha y_t^2 + \beta\sigma_t^2 \tag{2.1.14}$$

where y_t are log-returns. The time-varying variance σ_t^2 therefore depends recursively on squared of past returns. An immediate consequence of eq. (2.1.14) is that the conditional log-likelihood is:

$$\log p(y_{t+1}|\mathcal{Y}_t, \Psi) = -\frac{1}{2}\log(2\pi) - \log \sigma_{t+1} - \frac{y_{t+1}^2}{2\sigma_{t+1}^2} \tag{2.1.15}$$

where σ_{t+1} is given by the update rule (2.1.14). The full log-likelihood function can thus be written in closed form:

$$\log p(\Psi) = \sum_{t=1}^{N} \log p(y_t|\mathcal{Y}_{t-1}, \Psi) \tag{2.1.16}$$

and can be maximized with respect to $\Psi = \{\omega, \alpha, \beta\}$. Other examples of observation-driven models include the DCC model of Engle (2002a), the multiplicative error model (MEM) of Engle (2002b) and Engle and Gallo (2006) and the ACD model of Engle and Russell (1998).

There is a clear trade-off between parameter-driven and observation-driven models. On the one hand, parameter-driven models are more flexible, since the dynamics of time-varying parameters are driven by independent sources of randomness. In contrast, time-varying parameters are completely determined by past observations in observation-driven models. Also, statistical properties such us stationarity and ergodicity are generally more difficult to analyze for observation-driven models, as there are complex feedbacks effects between observations and time-varying parameters (see e.g. Blasques et al. (2014a). On the other hand, since the likelihood function can be written-down in closed form, observation-driven models are easy to estimate. The maximization of the likelihood is typically performed through standard quasi-Newton algorithms. In contrast, the estimation of parameter-driven models requires the use of computationally intensive simulation-based techniques.

2.2 The Kalman filter

In this Section we report the main results related to the Kalman filter, which will be extensively used throughout the thesis. We closely follow both Harvey (1991) and Durbin and Koopman (2012). Let us consider the linear Gaussian state-space model introduced in eq. (2.1.5). We can re-write it as:

$$y_t = Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, H) \tag{2.2.1}$$

$$\alpha_{t+1} = c + T\alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q) \tag{2.2.2}$$

We are interested in updating our knowledge of the underlying state variable α_t when a new observation y_t becomes available and in predicting α_{t+1} based on the last observations $y_1, \ldots, y_t$. Thus, we define:

$$a_{t|t} = \text{E}[\alpha_t|\mathcal{Y}_t], \quad P_{t|t} = \text{Var}[\alpha_t|\mathcal{Y}_t] \tag{2.2.3}$$

$$a_{t+1} = \text{E}[\alpha_{t+1}|\mathcal{Y}_t], \quad P_{t+1} = \text{Var}[\alpha_{t+1}|\mathcal{Y}_t] \tag{2.2.4}$$

The linearity assumption, together with normality, leads to the following result:

$$\alpha_t|\mathcal{Y}_t \sim N(a_{t|t}, P_{t|t}), \quad \alpha_{t+1}|\mathcal{Y}_t \sim N(a_{t+1}, P_{t+1}) \tag{2.2.5}$$

The Kalman filter allows to compute recursively $a_{t|t}$, $P_{t|t}$, a_{t+1} and P_{t+1}. Assuming $\alpha_1 \sim \mathrm{N}(a_1, P_1)$ and that a_1 and P_1 are known, for $t = 1, \ldots, N$ we have:

$$v_t = y_t - Za_t \tag{2.2.6}$$

$$a_{t|t} = a_t + P_t Z' F_t^{-1} v_t \tag{2.2.7}$$

$$a_{t+1} = c + Ta_t + K_t v_t \tag{2.2.8}$$

and

$$F_t = ZP_tZ' + H \tag{2.2.9}$$

$$P_{t|t} = P_t - P_t Z' F_t^{-1} Z P_t \tag{2.2.10}$$

$$P_{t+1} = TP_t(T - K_t Z)' + Q \tag{2.2.11}$$

where $K_t = TP_tZ'F_t^{-1}$ is the *Kalman gain*. The conditional mean $a_{t|t}$ is known as update filter while a_{t+1} is known as predictive filter. The two conditional covariance matrices $P_{t|t}$ and P_{t+1} allow to construct confidence bounds around $a_{t|t}$ and a_{t+1}, respectively. Note that, even in case the normality assumption is dropped, Kalman filter estimates are minimum variance linear unbiased (MVLU). Generalization of the above recursions to the case where a_1 and P_1 are unknown are discussed in detail by both Harvey (1991) and Durbin and Koopman (2012).

The problem of estimating α_t when all observations are available is known as smoothing. Let us define:

$$\hat{\alpha}_t = \mathrm{E}[\alpha_t|\mathcal{Y}_N], \quad \hat{P}_t = \mathrm{Var}[\alpha_t|\mathcal{Y}_N] \tag{2.2.12}$$

for $N > t$. Similarly to the filtering case, the conditional distribution of α_t given $\mathcal{Y}_N$ is normal:

$$\alpha_t|\mathcal{Y}_N \sim N(\hat{\alpha}_t, \hat{P}_t) \tag{2.2.13}$$

Both $\hat{\alpha}_t$ and $\hat{P}_t$ can be computed through the following backward recursions:

$$r_{t-1} = Z' F_t^{-1} v_t + L_t' r_t \tag{2.2.14}$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \tag{2.2.15}$$

and

$$N_{t-1} = Z' F_t^{-1} Z + L_t' N_t L_t \tag{2.2.16}$$

$$\hat{P}_t = P_t - P_t N_{t-1} P_t \tag{2.2.17}$$

where $L_t = T - K_t Z$, $r_N = 0$, $N_N = 0$ and $t = N, \dots, 1$. Note that a_t, P_t, v_t and F_t are provided by the filtering recursions. Thus, one can run the backward smoothing recursions after having performed the forward filtering recursions.

The matrices P_t, $P_{t|t}$, $\hat{P}_t$ characterize the variance of the conditional distribution of the state variable. They can therefore be used to construct confidence bands around filtered and smoothed estimates. These confidence bands reflect *filtering uncertainty*, i.e. the fact that the state variable is not completely determined by observations. As such, they are different from confidence bands for correctly specified observation-driven model, which reflect uncertainty on maximum likelihood estimates. In the language of Blasques et al. (2016), the latter is known as *parameter uncertainty*.

The linear Gaussian model in eq. (2.2.1), (2.2.2) is time-invariant, as system matrices are constant over time. In this case, the Kalman recursion for P_{t+1} converges to a constant matrix $\bar{P}$, which is the solution of the following matrix Riccati equation:

$$\bar{P} = T\bar{P}T' - T\bar{P}Z'\bar{F}^{-1}Z\bar{P}T' + Q \tag{2.2.18}$$

where $\bar{F} = Z\bar{P}Z' + H$. This result is useful from a computational point of view, as one can avoid computing recursions for $P_{t|t}$, P_{t+1} and $\hat{P}_t$ once convergence to $\bar{P}$ has been reached[1]. The solution of eq. (2.2.18) is known as *steady state*.

The log-likelihood of the linear-Gaussian model can be computed in the prediction error form, namely:

$$\log p(y_t|\mathcal{Y}_{t-1}, \Psi) = \text{const} - \frac{1}{2}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \tag{2.2.19}$$

where v_t and F_t are an output of the Kalman filter recursions and Ψ denotes the set of system matrices to be estimated. The full likelihood function can therefore be written as:

$$\log p(\Psi) = \sum_{t=1}^{N} \log p(y_t|\mathcal{Y}_{t-1}, \Psi) \tag{2.2.20}$$

and can be maximized with respect to Ψ through a quasi-Newton algorithm. For high-dimensional problems, the use of gradient-based algorithms become unfeasible. In this case, estimation is more conveniently performed through the expectation-maximization (EM) algorithm of A. P. Dempster (1977). The latter alternates between an estimation step, performed through the Kalman filter and smoother, and a maximization step where closed-form update formulas for the new values of Ψ can be recovered. The EM algorithm for linear-Gaussian state-space models is discussed e.g. by Shumway and Stoffer (1982).

A relevant advantage of the Kalman filter and smoother recursions for linear Gaussian state-space models is that they hold even in case observations are missing. We distinguish the case where, for some

[1] Convergence is typically very fast, with the matrix norm between P_t and $\bar{P}$ becoming very small after the first few steps.

t, *all* the n observations are missing and the case where some, but not all observations, are missing. In the first case, the Kalman recursions remain valid, provided what one sets $Z = 0$ at time t. The filtering recursions therefore become:

$$a_{t|t} = a_t, \quad P_{t|t} = P_t \tag{2.2.21}$$

$$a_{t+1} = Ta_t, \quad P_{t+1} = TP_tT + Q \tag{2.2.22}$$

while the smoothing recursions reduce to:

$$r_{t-1} = T'r_t, \quad N_{t-1} = T'N_tT \tag{2.2.23}$$

In the second case, one can define a selection matrix W_t such that $y_t^* = W_t y_t$ is the vector of values that are observed at time t. The observation equation becomes:

$$y_t^* = Z_t^* \alpha_t + \epsilon_t^*, \quad \epsilon_t^* \sim N(0, H^*) \tag{2.2.24}$$

where $Z_t^* = W_t Z$, $\epsilon_t^* = W_t \epsilon_t$ and $H^* = W_t H W_t'$. The Kalman filter and smoothing recursions (2.2.6)-(2.2.11) and (2.2.14)-(2.2.15) are then applied to the new observation equation (2.2.24). Note that the model is no longer time-invariant, as the dimensions of Z_t and H_t change over time.

2.3 Score-driven models

Score-driven models, also known as Generalized Autoregressive Score (GAS) models or Dynamic Conditional Score (DCS) models, have been introduced by Creal et al. (2013) and Harvey (2013). They are a general class of observation-driven models where time-varying parameters are driven by the score of the conditional likelihood.

Let us denote by $f_t \in \mathbb{R}^k$ the vector of time-varying parameters and by $p(y_t|f_t, \Psi)$ the conditional density of the observations $y_t \in \mathbb{R}^n$. The transition equation in a score-driven model has the following form:

$$f_{t+1} = \omega + AS_t\nabla_t + Bf_t \tag{2.3.1}$$

where $\nabla_t \in \mathbb{R}^k$ is the score of the conditional density:

$$\nabla_t = \frac{\partial \log p(y_t|f_t, \Psi)}{\partial f_t} \tag{2.3.2}$$

while S_t is a scaling matrix possibly depending on f_t. The vector $\omega \in \mathbb{R}^k$ and the two matrices $A, B \in \mathbb{R}^{k\times k}$ are static parameters. Creal et al. (2013) discussed several choices of S_t based on inverse powers of the information matrix $\mathcal{I}_{t|t-1} \in \mathbb{R}^{k\times k}$:

$$S_t = (\mathcal{I}_{t|t-1})^{-\alpha} = (\mathrm{E}_{t-1}[\nabla_t \nabla_t'])^{-\alpha}, \quad \alpha \in [0, 1] \tag{2.3.3}$$

Common choices of α are $\alpha = 0, 1/2, 1$. There are several intuitive motivations that justify the use of the score. For instance, the update equation (2.3.1) has similar form to a step of the Newton-Raphson algorithm used for numerical optimization. This means that the score updates f_t in the steepest ascent direction, improving the model's local fit in terms of density at time t. Harvey (2013) provides other possible interpretations. In Chapter 3, we will see that (2.3.1) is equivalent to the Kalman filter prediction step in eq. (2.2.8) where one replaces the Gaussian score with the score of the conditional density $p(y_t|f_t, \Psi)$.

Score-driven models have been extensively used in the financial econometric literature. Creal et al. (2011) developed a multivariate dynamic model for volatilities and correlations using fat tailed distributions. Harvey and Luati (2014) described a new framework for filtering with heavy tails while Oh and Patton (2017) introduced high-dimensional factor copula models based on score-driven dynamics for systemic risk assessment. Score-driven models have also been employed to model realized (co)variance measures. Notably examples are given by the realized Wishart-GARCH model of Hansen et al. (2016) and the Heavy GAS tF model of Opschoor et al. (2017).

Compared to other observation-driven models, score-driven models are locally optimal from an information theoretic perspective, as shown by Blasques et al. (2015). The asymptotic properties of the maximum likelihood estimator for score-driven models have been studied by Harvey (2013), Blasques et al. (2014a), Blasques et al. (2014b) and Blasques et al. (2017) while conditions for stationarity and ergodicity for univariate models have been analyzed by Blasques et al. (2014d). Koopman et al. (2016) showed that misspecified score-driven models have similar forecasting performance as correctly specified parameter-driven models.

Score-driven models can be used to generate a wide variety of new models with interesting dynamic features. The key ingredient is that the full shape of the observation density $p(y_t|f_t, \Psi)$ is taken into account not only in the estimation of the static parameters but also in the dynamic equation of time-varying parameters. An example explaining this fact can be made by considering an univariate normal density with a time-varying variance:

$$y_t|\sigma_t \sim N(0, \sigma_t^2) \tag{2.3.4}$$

It is immediate to see that, by setting $f_t = \sigma_t^2$ and $S_t = \mathcal{I}_{t|t-1}^{-1}$, eq. (2.3.1) reduces to:

$$f_{t+1} = \omega + A(y_t^2 - f_t) + Bf_t \tag{2.3.5}$$

that is, a GARCH(1,1) model. However, if one assumes a t-distribution with ν degrees of freedom, i.e.

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim t_\nu \tag{2.3.6}$$

eq. (2.3.1) does not reduce to the t-GARCH(1,1) model of Bollerslev (1987). Instead, one obtains the

following law of motion (see e.g. Creal et al. 2013 or Harvey 2013):

$$f_{t+1} = \omega + A(1 + 3\nu^{-1}) \left[\frac{1 + \nu^{-1}}{(1 + 2\nu^{-1})(1 + \nu^{-1}/(1 - 2\nu^{-1})y_t^2/f_t)} y_t^2 - f_t \right] + B f_t \tag{2.3.7}$$

The updating scheme in eq. (2.3.7) is more complex than the one in eq. (2.3.5). The GARCH(1,1) provides large changes in volatility when large returns occur. However, in presence of fat-tailed distributions, the latter are more likely to be due to outliers rather than shocks in volatility. The role of the factor multiplying y_t^2 in eq. (2.3.7) is to undermine volatility estimates in presence of outliers. This updating scheme is pertinent when dealing with fat-tailed densities and provides robustness against outliers.

2.4 Conditionally Gaussian models

A linear Gaussian model with time-varying system matrices reads:

$$y_t = Z_t \alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, H_t) \tag{2.4.1}$$

$$\alpha_{t+1} = c_t + T_t \alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q_t) \tag{2.4.2}$$

Econometric inference on such a model relies on the law of motion of the time-varying parameters Z_t, H_t, c_t, T_t and Q_t. If system matrices follow a dynamic process with independent sources of randomness, then model (2.4.1) (2.4.2), together with the new dynamic equations for the system matrices, becomes a nonlinear non-Gaussian parameter-driven model. The latter can be estimated through simulation-based techniques. As an example, Stock and Watson (2007) used Markov Chain Monte Carlo (MCMC) methods to estimate their local level model with stochastic variances.

Another possibility is that system matrices depend on past observations, namely:

$$y_t = Z_t(\mathcal{Y}_{t-1})\alpha_t + \epsilon_t, \quad \epsilon_t | \mathcal{Y}_{t-1} \sim \text{NID}(0, H_t) \tag{2.4.3}$$

$$\alpha_{t+1} = c_t(\mathcal{Y}_{t-1}) + T_t(\mathcal{Y}_{t-1})\alpha_t + \eta_t, \quad \eta_t | \mathcal{Y}_{t-1} \sim \text{NID}(0, Q_t) \tag{2.4.4}$$

These kinds of models are known as conditionally Gaussian (Harvey 1991). The advantage of this formulation is that, conditionally on past information, the model is linear and Gaussian, with deterministic time-varying parameters. Thus, the Kalman filter can be applied exactly as in the time-invariant case. In particular, the Kalman filtering recursions become:

$$v_t = y_t - Z_t a_t \tag{2.4.5}$$

$$a_{t|t} = a_t + P_t Z_t' F_t^{-1} v_t \tag{2.4.6}$$

$$a_{t+1} = c + T_t a_t + K_t v_t \tag{2.4.7}$$

and

$$F_t = Z_t P_t Z_t' + H_t \tag{2.4.8}$$

$$P_{t|t} = P_t - P_t Z_t' F_t^{-1} Z_t P_t \tag{2.4.9}$$

$$P_{t+1} = T_t P_t (T_t - K_t Z)' + Q_t \tag{2.4.10}$$

where $K_t = T_t P_t Z_t' F_t^{-1}$. Similarly, Kalman smoother recursions are written in terms of the time-varying system matrices, which are known one-step ahead:

$$r_{t-1} = Z_t' F_t^{-1} v_t + L_t' r_t \tag{2.4.11}$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \tag{2.4.12}$$

and

$$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t \tag{2.4.13}$$

$$\hat{P}_t = P_t - P_t N_{t-1} P_t \tag{2.4.14}$$

where $L_t = T_t - K_t Z_t$, $r_N = 0$, $N_N = 0$ and $t = N, \ldots, 1$. The conditional log-likelihood can be written in the error decomposition form:

$$\log p(y_t | \mathcal{Y}_{t-1}, \Psi) = \text{const} - \frac{1}{2} \left(\log |F_t| + v_t' F_t^{-1} v_t \right) \tag{2.4.15}$$

where v_t and F_t are now obtained through the time-varying Kalman filter recursions (2.4.5)-(2.4.10).

The vector Ψ collects the set of static parameters governing the dynamics of the system matrices and can be estimated by numerically maximizing the log-likelihood function. Note that the Kalman filter is no longer a linear function of past observations. Therefore, conditionally Gaussian models feature nonlinear dynamics while at the same time being susceptible of treatment with the Kalman filter. This is a relevant advantage from an inferential point of view, as estimation can be performed through standard maximum likelihood methods rather than computationally demanding simulation-based techniques. Note also that, in the framework of Cox (1981), conditionally Gaussian models can be regarded as mixed parameter and observation-driven models. Indeed, while the time-varying state vector follows a dynamic model with its own innovations, the dynamics of system matrices are driven by past observations.

In order to provide a complete specification of the model, one needs to augment eq. (2.4.3), (2.4.4) with the law of motion of the system matrices. Since the conditional log-likelihood can be written-down in closed form, score-driven models provide a natural, observation-driven framework to model the dynamics of system matrices. Conditionally Gaussian models with score-driven, time-varying parameters were introduced by Creal et al. (2008) and have been described in their full generality by Delle Monache et al. (2016), who dubbed these class of models as "adaptive state-space" models. Let $f_t \in \mathbb{R}^k$ collect all

the time-varying parameters in Z_t, H_t, c_t, T_t and Q_t. In order to implement the update equation (2.3.1), one needs to compute the score ∇_t and the information matrix $\mathcal{I}_{t|t-1}$ from the conditional log-likelihood (2.4.15). Delle Monache et al. (2016) showed the following result[2]:

$$\nabla_t = -\frac{1}{2}\left[\dot{F}_t'(I_n \otimes F_t^{-1})\text{vec}(I_n - v_t v_t' F_t^{-1}) + 2\dot{v}_t' F_t^{-1} v_t\right] \tag{2.4.16}$$

$$\mathcal{I}_{t|t-1} = \frac{1}{2}\left[\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})\dot{F}_t + 2\dot{v}_t' F_t^{-1}\dot{v}_t\right] \tag{2.4.17}$$

where $\dot{F}_t = \partial \text{vec}(F_t)/\partial f_t'$ and $\dot{v}_t = \partial v_t/\partial f_t'$ denote derivatives of F_t and v_t with respect to f_t.

The prediction error v_t and the covariance matrix F_t appearing in eq. (2.4.16), (2.4.17) are an output of the Kalman filter recursions. Instead, $\dot{v}_t$ and $\dot{F}_t$ can be computed by deriving eq. (2.4.5), (2.4.8) with respect to f_t. As a result, one obtains a new filter which runs in parallel with the standard Kalman filter recursions, thus allowing to compute ∇_t and $\mathcal{I}_{t|t-1}$ at each time step. The parallel filter recursions for the "full" model (2.4.3), (2.4.4) are reported in Delle Monache et al. (2016). In Chapter 5, we will compute the parallel filtering recursions for a multivariate local level model with time-varying covariance matrices while in Chapter 6 we will examine an HAR (Corsi 2009) plus noise model with time-varying coefficients and stochastic variance.

The methodology described above provides a unified filter for both the state vector α_t and the time-varying system matrices. Extension to the case with missing observations is readily available by writing the observation equation as in eq. (2.2.24) and will be relevant for our application to high-frequency data in Chapter 5. Several applications of conditionally Gaussian models to macroeconomic data can be found in Delle Monache et al. (2015) and Delle Monache et al. (2016).

[2]The notation used here is described in detail in appendix (A.1)

Chapter 3

Smoothing with score-driven models

Almost all results in this chapter previously appeared in Buccheri et al. (2017a).

As discussed in Chapter 1, observation-driven models are typically regarded as data generating processes. As such, all relevant information is encoded on past observations and there is no room for using actual and future observations when estimating time-varying parameters. However, they can also be viewed as predictive filters, as time-varying parameters are one-step-ahead predictable. This idea was largely exploited by Daniel B. Nelson, who explored the asymptotic properties of conditional covariances of a misspecified GARCH under the assumption that the data generating process is a diffusion[1]; see Nelson (1992), Nelson and Foster (1994), Nelson and Foster (1995) and Nelson (1996). In particular, Nelson (1996) showed how to efficiently use information in both lagged and led GARCH residuals to estimate the unobserved states of stochastic volatility models. Despite many observation-driven models have been proposed in the econometric literature, little attention has been paid to the problem of smoothing within this class of models in case they are employed as misspecified filters rather than data generating processes.

We aim at filling this gap by introducing a smoothing method for score-driven models of Creal et al. (2013) and Harvey (2013), which represent a general class of observation-driven models. We show that, in the steady state, Kalman filter and smoothing recursions for linear Gaussian models can be re-written in terms of the score of the conditional density, the Fisher information matrix and a set of static parameters. In particular, the predictive filtering recursion turns out to have the form of score-driven models. The latter can therefore be viewed as approximate filters for nonlinear non-Gaussian models. The performances of these filters have been examined by Koopman et al. (2016),

[1]The interpretation of GARCH processes as filters is well described in this statement by Nelson (1992): *"Note that our use of the term 'estimate' corresponds to its use in the filtering literature rather than the statistics literature; that is, an ARCH model with (given) fixed parameters produces 'estimates' of the true underlying conditional covariance matrix at each point in time in the same sense that a Kalman filter produces 'estimates' of unobserved state variables in a linear system".*

who showed that misspecified score-driven models provide similar forecasting performances as correctly specified parameter-driven models. Based on the same logic, we build a new class of approximate nonlinear smoothers that have similar form to Kalman backward smoothing recursions but employ the score of the non-Gaussian density. The resulting smoothing method is very general, as it can be applied to any observation density, in a similar fashion to score-driven models. We name the newly proposed methodology Score-Driven Smoother (SDS). Similarly, we introduce a Score-Driven Update (SDU) filter, allowing to update predictive filtered estimates once new observations become available.

Smoothing with the SDS requires performing a backward recursion following the standard score-driven forward recursion to filter time-varying parameters. While going backward, the SDS updates filtered estimates by including the effect of actual and future observations and leads to a more efficient reconstruction of time-varying parameters. In our experiments, we have found that, compared to filtered estimates, the SDS provides gains up to 63% in mean square errors, for a wide class of data generating processes. Considered that the likelihood of observation-driven models can be typically written down in closed form, smoothing with the SDS is particularly advantageous from a computational point of view. In contrast, the classical theory of filtering and smoothing for nonlinear non-Gaussian models requires the use of computationally demanding simulation-based techniques (Durbin and Koopman 2012). Another relevant advantage of the SDS over traditional simulation-based methods is that extension to a setting with multiple time-varying parameters is immediate, as it maintains the same simple form as in the univariate case.

This general framework allows to construct confidence bands around filtered and smoothed estimates. In observation-driven models, confidence bands are typically needed because static parameters are replaced by their maximum likelihood estimates. In the language of Blasques et al. (2016), this is known as parameter uncertainty. However, if observation-driven models are employed as filters, the latent state variables are not completely revealed by past observations. Thus, also filtering uncertainty has to be considered when building confidence bands. While confidence bands reflecting parameter uncertainty can be built through the methods developed by Blasques et al. (2016), it is less clear how one can take into account filtering uncertainty in observation-driven models. Zamojski (2016) proposed a bootstrap based method to construct in-sample confidence bands for the GARCH. As acknowledged by the author, this method leads to underestimate filtering uncertainty and provides narrow confidence bands. We show that, as a byproduct of our results, one can build both in-sample and out-of-sample confidence bands accounting for filtering uncertainty in score-driven models. In this Chapter, we examine in detail the construction of confidence bands in the case of the GARCH model. A general and systematic treatment of filtering uncertainty in score-driven models is provided by Buccheri et al. (2018c).

For any score-driven model, one can devise companion SDS and SDU recursions. In particular,

the SDS is useful for off-line signal reconstruction and analysis, while the SDU can be used for on-line updating of time-varying parameters. We examine in detail the companion SDS and SDU recursions of popular observation-driven models, namely the GARCH, the MEM model of Engle (2002b) and Engle and Gallo (2006) and an AR(1) model with a time-varying autoregressive coefficient. In oder to show the effectiveness of the proposed methodology in a setting with multiple time-varying parameters, we consider the t-GAS model of Creal et al. (2011) and the Wishart-GARCH model of Hansen et al. (2016).

A related smoothing technique for a dynamic Student's t location model was introduced by Harvey (2013), who replaced prediction errors in the Kalman smoothing recursions with a martingale difference that is proportional to the score of the t distribution. An application of this smoother can be found in Caivano et al. (2016). The main difference with our approach is that we write the Kalman recursions for the mean of time-invariant linear Gaussian models in a general form that only depends on the score and the Fisher information matrix of the observation density. The resulting smoothing recursions are different by those obtained by Harvey (2013) and are easily applicable to a generic score-driven model by replacing the Gaussian density with the observation density at hand. The SDS is also related to the "approximation via mode estimation" technique described by Durbin and Koopman (2000) and Durbin and Koopman (2012). These authors proved that one can find a sequence of approximating linear Gaussian models enabling the computation of the conditional mode of a non-Gaussian model via a Newton-Raphson algorithm. The main difference with our methodology is that the SDS requires a unique, nonlinear recursion rather than a sequence of Kalman recursions for approximating linear Gaussian models. In addition, in our methodology, the filter coincides with well-known observation-driven model (e.g. GARCH, MEM, ACD, etc) while the approximation via mode estimation technique uses a sequence of filters that are not easily interpretable as dynamic models.

By performing extensive Monte Carlo simulations of nonlinear non-Gaussian state-space models, we compare the performance of the SDS to that of correctly specified parameter-driven models. In particular, we consider two stochastic volatility models and a stochastic intensity models. Importance sampling methods allow to evaluate the full likelihood of these models. Compared to correctly specified models, the losses incurred by the SDS are very small in all the simulated scenarios and are always lower, on average, than 2.5% in mean square errors. Computational times are decisively in favour of the SDS. For the models used in the simulation study, we found that smoothing with the SDS is on average 215 times faster than smoothing with efficient importance sampling techniques. The advantages of the proposed method are also shown on empirical data. Using realized covariance as a proxy of latent covariance, we show that SDU and SDS covariance estimates obtained through the dynamic t-GAS model fitted on Russel 3000 stock returns are superior to standard filtered score-driven estimates. The analysis allows to examine the informational content of present and future log-returns from a dynamic

covariance modelling perspective.

3.1 Theoretical framework

In this Section, we discuss in detail the main theoretical results conveying to the formulation of our approximate, nonlinear smoothing technique. We start by showing that, in the steady state, the classical Kalman filter and smoothing recursions for linear Gaussian models can be re-written in an alternative form that only involves the score of the conditional likelihood, the Fisher information matrix and a set of static parameters. Abstracting from the linear Gaussian setting, these recursions can be viewed as the approximate filtering and smoothing recursions for a non-Gaussian model by computing scores and information based on the non-Gaussian density. We then show that filtering uncertainty in score-driven models can be evaluated as an immediate byproduct of our results.

3.1.1 Kalman recursions: a more general representation

In Appendix A.2 we prove the following:

Proposition 1. *In the steady state, eq. (2.2.7), (2.2.8), (2.2.14), (2.2.15) can be written as:*

$$a_{t|t} = a_t + T^{-1}R\nabla_t \tag{3.1.1}$$

$$a_{t+1} = c + Ta_t + R\nabla_t \tag{3.1.2}$$

and

$$r_{t-1} = \nabla_t + (T - R\mathcal{I})'r_t \tag{3.1.3}$$

$$\hat{\alpha}_t = a_t + T^{-1}Rr_{t-1} \tag{3.1.4}$$

where $\nabla_t = \frac{\partial \log p(y_t|\mathcal{F}_{t-1})}{\partial a_t}$, $\mathcal{I} = Z'\bar{F}^{-1}Z$, $\bar{F} = Z'\bar{P}Z + H$, $R = T\bar{P}$ *and* $\bar{P}$ *is the steady state variance matrix which is the solution of the matrix Riccati equation:*

$$\bar{P} = T\bar{P}T' - T\bar{P}Z'\bar{F}^{-1}Z\bar{P}T' + Q \tag{3.1.5}$$

Note that a steady state solution exists whenever the system matrices are constant (Harvey 1991, Durbin and Koopman 2012). In this case, the variance matrix P_t converges to $\bar{P}$ after few time steps. The new Kalman recursions for the mean are re-parameterized in terms of the score ∇_t and the Fisher information matrix $\mathcal{I}$. This representation is equivalent to the one in equations (2.2.7), (2.2.8) and (2.2.14), (2.2.15). However, it is more general, as it only relies on the measurement density $p(y_t|\mathcal{F}_{t-1})$. In principle, the forward recursions (3.1.1), (3.1.2) and the backward recursions (3.1.3), (3.1.4) can be applied to any parameter-driven model for which a measurement density $p(y_t|\mathcal{F}_{t-1})$ is defined.

3.1.2 SDS recursions

The predictive filter (3.1.2) has an autoregressive structure and is driven by the score of the conditional likelihood, i.e. it has the form of score-driven models of Creal et al. 2013 and Harvey 2013. Thus, if one looks at score-driven models as filters, it turns out that the score-driven filter (SDF hereafter) is optimal in case of linear Gaussian models. In case of nonlinear non-Gaussian models, the SDF can be regarded as an approximate nonlinear filter. The main difference with the Kalman filter is that the Gaussian score is replaced by the score of the true conditional density, thus providing robustness to non-Gaussianity. As shown by Koopman et al. (2016), score-driven filters have similar predictive accuracy as correctly specified nonlinear non-Gaussian models, while at the same time providing significant computational gains. Indeed, the likelihood can be written in closed form and standard quasi-Newton techniques can be employed for optimization.

Based on the same principle, we introduce an approximate nonlinear smoother allowing to estimate time-varying parameters using all available observations. In case of linear Gaussian models, the Kalman smoother is a minimum variance linear unbiased estimator (MVLUE) of the state. Thus, we define our smoother in such a way that it coincides with the latter in this specific case. In case of nonlinear non-Gaussian models, it maintains the same simple form of Kalman backward smoothing recursions but replaces the Gaussian score with the one of the non-Gaussian density.

Let us assume that observations $y_t \in \mathbb{R}^p$, $t = 1, \dots, n$, are generated by the following observation density:

$$y_t | f_t \sim p(y_t | f_t, \Theta) \tag{3.1.6}$$

where $f_t \in \mathbb{R}^k$ is a vector of time-varying parameters and Θ is a vector of static parameters. We generalize the filtering and smoothing recursions (3.1.1)-(3.1.4) for the measurement density $p(y_t|f_t, \Theta)$ as:

$$f_{t|t} = f_t + B^{-1} A \nabla_t \tag{3.1.7}$$

$$f_{t+1} = \omega + A \nabla_t + B f_t \tag{3.1.8}$$

$t = 1, \dots, n$ and:

$$r_{t-1} = \nabla_t + (B - A\mathcal{I}_{t|t-1})' r_t \tag{3.1.9}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{3.1.10}$$

where $r_n = 0$ and $t = n, \dots, 1$. The predictive filter in Eq. (3.1.8) has the same form of score-driven models. The term ∇_t is now the score of the measurement density $p(y_t|f_t, \Theta)$, namely:

$$\nabla_t = \frac{\partial \log p(y_t | f_t, \Theta)}{\partial f_t} \tag{3.1.11}$$

while $\mathcal{I}_{t|t-1} = \mathrm{E}_{t-1}[\nabla_t \nabla_t']$ is the information matrix, which may be time-varying. The vector $\omega \in \mathbb{R}^k$ and the two matrices $A, B \in \mathbb{R}^{k \times k}$ are static parameters included in Θ. They are estimated by maximizing the log-likelihood, namely:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \sum_{t=1}^{n} \log p(y_t | f_t, \Theta) \tag{3.1.12}$$

Thus, one can run the backward smoothing recursions (3.1.9), (3.1.10) after computing the forward filtering recursions (3.1.7), (3.1.8), in a similar fashion to Kalman filter and smoothing recursions. Note that the above recursions are nonlinear, as the score of a non-Gaussian density is typically nonlinear in the observations. The filter $f_{t|t}$ in eq. (3.1.7) allows to update the current estimate f_t once a new observation y_t becomes available. While going backward, the smoothing recursions (3.1.9), (3.1.10) update the two filters f_t and $f_{t|t}$ using all available observations. Smoothed estimates $\hat{f}_t$ are generally less noisy than filtered estimates $f_{t|t}$, f_t and provide a more accurate reconstruction of the time-varying parameters.

As discussed in Section 2.3, it is a standard practice in score-driven models replacing the score ∇_t with the scaled score $s_t = S_t \nabla_t$. The role of the scaling matrix S_t is to take into account the curvature of the log-likelihood function. The filtering and smoothing recursions (3.1.7)-(3.1.10) are obtained if one sets S_t equal to the identity matrix. When using a scaled score s_t, the filtering recursions (3.1.7), (3.1.8) become:

$$f_{t|t} = f_t + B^{-1} A s_t \tag{3.1.13}$$

$$f_{t+1} = \omega + A s_t + B f_t \tag{3.1.14}$$

Since the score is now scaled by S_t, the term $A\mathcal{I}_{t|t-1}$ in eq. (3.1.9) must take into account the new normalization. Thus, we replace A with AS_t. As a result, we obtain the general backward smoothing recursions:

$$r_{t-1} = s_t + (B - AS_t \mathcal{I}_{t|t-1})' r_t \tag{3.1.15}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{3.1.16}$$

Note that the second equation is unaffected, as the term r_{t-1} already corrects for the scaling. For instance, if $S_t = \mathcal{I}_{t|t-1}^{-1}$, we obtain:

$$r_{t-1} = s_t + (B - A)' r_t \tag{3.1.17}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{3.1.18}$$

that is, the information matrix $\mathcal{I}_{t|t-1}$ disappears because its effect is already taken into account when

scaling the score. If $S_t = \mathcal{I}_{t|t-1}^{-1/2}$, we get:

$$r_{t-1} = s_t + (B - A\mathcal{I}_{t|t-1}^{1/2})' r_t \tag{3.1.19}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{3.1.20}$$

From a computational point of view, the backward recursions (3.1.15), (3.1.16) are simple since s_t and $\mathcal{I}_{t|t-1}$ are typically available from the forward filtering recursion. We term the approximate smoother obtained through recursions (3.1.15), (3.1.16) as Score-Driven Smoother (SDS). Basically, for any score-driven model, one can devise a companion SDS recursion that only requires the s_t, $\mathcal{I}_{t|t-1}$ and the static parameters, as estimated through the SDF. Note that the forward recursion (3.1.13) is the analogue of recursion (2.2.7) in the Kalman filter and allows to update SDF estimates once a new observation y_t becomes available. We denote the approximate Score-Driven Update filter (3.1.13) by SDU. The proposed methodology can thus be schematically represented through the following procedure:

1. Estimation of static parameters:

$$\tilde{\Theta} = \underset{\Theta}{\operatorname{argmax}} \sum_{t=1}^{n} \log p(y_t | f_t, \Theta)$$

2. Forward predictive and update filter:

$$f_{t+1} = \tilde{\omega} + \tilde{A} s_t + \tilde{B} f_t$$

$$f_{t|t} = f_t + \tilde{B}^{-1} \tilde{A} s_t$$

3. Backward smoother:

$$r_{t-1} = s_t + (\tilde{B} - \tilde{A} S_t \mathcal{I}_{t|t-1})' r_t$$

$$\hat{f}_t = f_t + \tilde{B}^{-1} \tilde{A} r_{t-1}$$

3.1.3 Filtering uncertainty

The general framework developed in Section 3.1.2 also allows to construct in-sample and out-of-sample confidence bands around filtered and smoothed estimates. As underlined by Blasques et al. (2016), confidence bands can reflect both parameter and filtering uncertainty. Parameter uncertainty is related to the fact that static parameters are replaced by their maximum likelihood estimates. Both observation-driven and parameter-driven models are affected by parameter uncertainty. In observation-driven models, confidence bands reflecting parameter uncertainty can be constructed through the methods developed by Blasques et al. (2016). Filtering uncertainty is related to the fact that time-varying parameters are not completely revealed by observations. As such, it is absent in observation-driven models, where time-varying parameters are deterministic functions of past observations. However, if observation-driven

models are regarded as filters, one is interested in constructing confidence bands around filtered and smoothed estimates reflecting the conditional distribution of the underlying state variable.

In linear Gaussian models, filtering uncertainty can be assessed through the variance matrices P_{t+1}, $P_{t|t}$, $\hat{P}_t$ introduced in Section 2.2, which provide the conditional variance of the unobserved state variable. It is instead less clear how one can quantify filtering uncertainty in misspecified observation-driven models. Zamojski (2016) proposed a bootstrap based method for assessing filtering uncertainty in GARCH filters. Confidence bands constructed through this technique tend to underestimate filtering uncertainty, because they are based on bootstraps of the filter rather than the underlying state variable. In addition, the method of Zamojski (2016) does not allow to construct out-of-sample confidence bands, which are often needed in practical applications.

In our framework, in-sample and out-sample confidence bands can be constructed by exploiting the relation between Kalman filter recursions and score-driven recursions. In Section 3.1.1, we have shown that the steady state variance matrix $\bar{P}$ can be expressed as:

$$\bar{P} = T^{-1}R \tag{3.1.21}$$

In the score-driven framework, the analogue of P_{t+1}, which we denote by J_{t+1}, is then given by:

$$J_{t+1} = B^{-1}AS_t \tag{3.1.22}$$

where the scaling matrix S_t is introduced to take into account different normalizations of the score. From eq. (2.2.10), (A.2.2), the analogue of $P_{t|t}$ is:

$$J_{t|t} = J_t - J_t\mathcal{I}_{t|t-1}J_t \tag{3.1.23}$$

Similarly, the analogue of $\hat{P}_t$, from eq. (3.1.24), (2.2.10), is:

$$N_{t-1} = \mathcal{I}_{t|t-1} + (B - AS_t\mathcal{I}_{t-1})'N_t(B - AS_t\mathcal{I}_{t-1}) \tag{3.1.24}$$

$$\hat{J}_t = J_t - J_tN_{t-1}J_t \tag{3.1.25}$$

with $N_n = 0$ and $t = n, \dots, 1$.

Confidence bands can be computed as quantiles of the conditional distribution of the state variable. For a general state-space model, the latter is non-Gaussian and is not known analytically. Assuming a Gaussian density generally leads to underestimate filtering uncertainty, as the true conditional density is typically fat-tailed. In order to construct robust confidence bands, we use a more flexible density determined by matching location and scale parameters with those of the normal density. This method is described in its full generality by Buccheri et al. (2018c). In Section 3.2.2, we show an application to the GARCH and assess the performance of robust confidence bands in a simulation study.

3.2 Examples of SDS recursions

In this Section we provide several examples of SDS estimates. As a first step, we focus on two volatility models that are quite popular in the econometric literature, namely the GARCH model of Bollerslev (1986) and the multiplicative error model (MEM) of Engle (2002b) and Engle and Gallo (2006). These are score-driven models which are susceptible of treatment within our framework. As a third example, we present an AR(1) model with a score-driven autoregressive coefficient. The time-varying autoregressive coefficient allows to capture temporal variations in persistence, as well as nonlinear dependencies (Blasques et al. 2014c). Autoregressive models with time-varying coefficients have been employed for instance by Delle Monache and Petrella (2017). In Chapter 6 we will use time-varying autoregressive coefficients to capture nonlinearities in volatility dynamics.

One of the advantages of the SDS recursions (3.1.15), (3.1.16) is that they maintain the same simple form when $f_t \in \mathbb{R}^k$, $k > 1$ is a vector containing multiple time-varying parameters. In this multivariate setting, the use of simulation-based techniques would be highly computationally demanding. In order to test the SDS in a multivariate setting, we consider the t-GAS model of Creal et al. (2011) and the Wishart-GARCH model of Hansen et al. (2016). The former is a conditional correlation model for heavy-tail returns while the latter is a joint model for the dynamics of daily returns and realized covariance matrices. In these models, the number of time-varying parameters grows as the square of the number of assets and therefore they provide an interesting multivariate framework in which to assess the performance of the SDS.

1. GARCH-SDS

Consider the model:

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \text{NID}(0,1) \tag{3.2.1}$$

The conditional density is thus:

$$p(y_t|\sigma_t^2) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{y_t^2}{2\sigma_t^2}} \tag{3.2.2}$$

Setting $f_t = \sigma_t^2$ and $S_t = \mathcal{I}_{t|t-1}^{-1}$, equations (3.1.13), (3.1.14) reduce to:

$$f_{t|t} = f_t + B^{-1}A(y_t^2 - f_t) \tag{3.2.3}$$

$$f_{t+1} = \omega + A(y_t^2 - f_t) + Bf_t \tag{3.2.4}$$

In particular, the predictive filter (3.2.4) is the standard GARCH(1,1) model. The smoothing recursions (3.1.15), (3.1.16) reduce to:

$$r_{t-1} = y_t^2 - f_t + (B - A)'r_t \tag{3.2.5}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{3.2.6}$$

$t = n, \ldots, 1$.

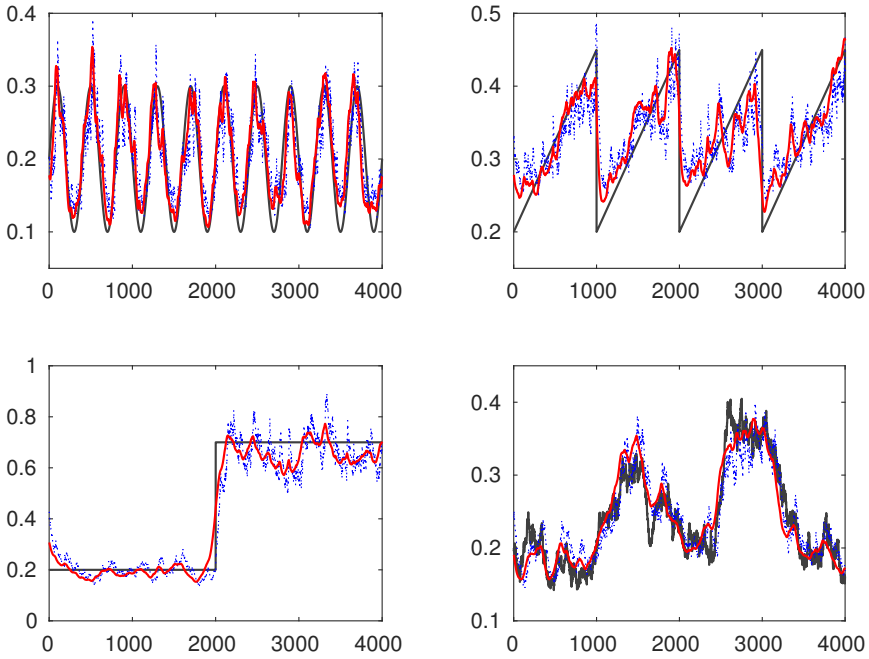


Figure 3.2.1: Comparison among simulated (black lines), filtered (blue dotted lines) and smoothed (red lines) variance σ_t^2 of GARCH(1,1) model.

2. MEM-SDS

Consider the model:

$$y_t = \mu_t \epsilon_t \tag{3.2.7}$$

where ϵ_t has a Gamma distribution with density $p(\epsilon_t|\alpha) = \Gamma(\alpha)^{-1}\epsilon_t^{\alpha-1}\alpha^\alpha e^{-\alpha\epsilon_t}$. The conditional density is thus given by:

$$p(y_t|\mu_t, \alpha) = \Gamma(\alpha)^{-1} y_t^{\alpha-1}\alpha^\alpha \mu_t^{-\alpha} e^{-\alpha\frac{y_t}{\mu_t}} \tag{3.2.8}$$

Setting $f_t = \mu_t$ and $S_t = \mathcal{I}_{t|t-1}^{-1}$, equations (3.1.13), (3.1.14) reduce to:

$$f_{t|t} = f_t + B^{-1}A(y_t - f_t) \tag{3.2.9}$$

$$f_{t+1} = \omega + A(y_t - f_t) + Bf_t \tag{3.2.10}$$

In particular, the predictive filter (3.2.10) is the standard MEM(1,1) model. The smoothing recursions (3.1.15), (3.1.16) reduce to:

$$r_{t-1} = y_t - f_t + (B - A)'r_t \tag{3.2.11}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{3.2.12}$$

$t = n, \ldots, 1$.

3. AR(1)-SDS

Consider the model:

$$y_t = c + \alpha_t y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, q^2) \tag{3.2.13}$$

The conditional density is thus given by:

$$p(y_t|\alpha_t) = \frac{1}{\sqrt{2\pi}q} \exp\left[-\frac{1}{2}\left(\frac{y_t - c - \alpha_t y_{t-1}}{q}\right)^2\right] \tag{3.2.14}$$

Setting $f_t = \alpha_t$ and $S_t = \mathcal{I}_{t|t-1}^{-1}$, equations (3.1.13), (3.1.14) reduce to:

$$f_{t|t} = f_t + B^{-1}A\left(\frac{y_t - c - f_t y_{t-1}}{y_{t-1}}\right) \tag{3.2.15}$$

$$f_{t+1} = \omega + A\left(\frac{y_t - c - f_t y_{t-1}}{y_{t-1}}\right) + Bf_t \tag{3.2.16}$$

while the smoothing recursions (3.1.15), (3.1.16) reduce to:

$$r_{t-1} = \left(\frac{y_t - c - f_t y_{t-1}}{y_{t-1}}\right) + (B - A)'r_t \tag{3.2.17}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{3.2.18}$$

$t = n, \ldots, 1$.

4. t-GAS-SDS

Let $r_t \in \mathbb{R}^p$ denote a vector of demeaned daily log-returns. Consider the following observation density:

$$p(r_t|V_t, \nu) = \frac{\Gamma((\nu + p)/2)}{\Gamma(\nu/2)[(\nu - 2)\pi]^{p/2}|V_t|^{1/2}}\left[1 + \frac{r_t' V_t^{-1} r_t}{(\nu - 2)}\right]^{-\frac{\nu+p}{2}} \tag{3.2.19}$$

where $V_t \in \mathbb{R}^{p\times p}$ is a time-varying covariance matrix and $\nu > 2$ is the number of degrees of freedom. Note that $p(r_t|V_t, \nu)$ is a normalized Student t distribution such that $\mathrm{Cov}(r_t|V_t, \nu) = V_t$. Applying the filtering eq. (3.1.14) leads to the t-GAS model of Creal et al. (2011). Closed form formulas for the score and information matrix are reported in Creal et al. (2011). These authors also proposed two parameterizations of V_t leading to positive-definite estimates. The first is similar to the one used in the DCC model of Engle (2002a), while the second is based of hyperspherical coordinates. In the two parameterizations, the number of time-varying parameters is $k = p + p(p + 1)/2$ and $k = p(p + 1)/2$, respectively.

5. Wishart-GARCH-SDS

Let us assume that, in addition to daily log-returns r_t, we can compute realized measures from the intraday returns of the p assets. Let $X_t \in \mathbb{R}^{p\times p}$ denote a positive definite estimate of the realized

covariance matrix. Let also $\mathcal{F}_t$ denote the σ-field generated by r_t and X_t. The observation density in the Wishart-GARCH model is:

$$r_t|\mathcal{F}_{t-1} \sim N_k(0, V_t) \tag{3.2.20}$$

$$X_t|\mathcal{F}_{t-1} \sim W_k(V_t/\nu, \nu) \tag{3.2.21}$$

where $N_k(0, V_t)$ is a multivariate zero-mean normal distribution with covariance matrix V_t and $W_k(V_t/\nu, \nu)$ is a Wishart distribution with mean V_t and degrees of freedom $\nu \geq p$. Assuming that r_t and X_t are conditionally independent given $\mathcal{F}_{t-1}$, the conditional log-likelihood can be written as:

$$\log L(r_t, X_t|V_t, \mathcal{F}_{t-1}) = \log L(r_t|V_t, \mathcal{F}_{t-1}) + \log L(X_t|V_t, \mathcal{F}_{t-1}) \tag{3.2.22}$$

where:

$$\log L(r_t|V_t, \mathcal{F}_{t-1}) = \frac{1}{2}d_r(p) - \frac{1}{2}\log|V_t| - \frac{1}{2}\text{tr}(V_t^{-1}r_t r_t') \tag{3.2.23}$$

$$\log L(X_t|V_t, \mathcal{F}_{t-1}) = \frac{1}{2}d_X(p) + \frac{\nu - p - 1}{2}\log|X_t| - \frac{\nu}{2}\log|V_t| - \frac{\nu}{2}\text{tr}(V_t^{-1}X_t) \tag{3.2.24}$$

Here $d_r(p) = -p\log(2\pi)$, $d_X(p, \nu) = \nu p \log(\nu/2) - 2\log\Gamma_p(\nu/2)$ and Γ_p is the multivariate Gamma function of order p. We denote the vector of time-varying covariances by $f_t = \text{vech}(V_t) \in \mathbb{R}^k$, $k = \frac{p(p+1)}{2}$. The score $\nabla_t = \frac{\partial \log L(r_t, X_t|f_t, \mathcal{F}_{t-1})}{\partial f_t} \in \mathbb{R}^k$ and the information matrix $\mathcal{I}_{t|t-1} = \text{E}_{t-1}[\nabla_t \nabla_t'] \in \mathbb{R}^{k\times k}$ can be computed as reported in Hansen et al. (2016). Opschoor et al. (2017) proposed an alternative specification with a heavy tail distribution for both returns and realized measures. Similar SDS recursions can be recovered for this fat-tail specification using our general framework.

Figures 3.2.1 - 3.2.3 show several examples of SDS estimates from the above models. The time-varying parameters follow both deterministic and stochastic patterns and are generated as described in the next paragraph.

3.2.1 Comparison of SDF and SDS estimates

In order to show the effectiveness of the proposed methodology, we compare SDF, SDU and SDS estimates. It is natural expecting that SDU and SDS estimates are affected by lower estimation errors, as they use more information when reconstructing time-varying parameters. However, comparing with the SDF allows to provide a quantitative assessment of the benefits of using the SDU and SDS in place of standard score-driven estimates.

We first focus on the univariate models (GARCH, MEM, AR(1)) and simulate $N = 250$ time-series of $n = 4000$ observations with different dynamic patterns for the time-varying parameters. The first 2000 observations are used to estimate the models while the remaining observations are used for testing. Let β_t generically denote the time-varying parameters σ_t^2, μ_t and α_t in the three models. We consider the following data generating processes for β_t:

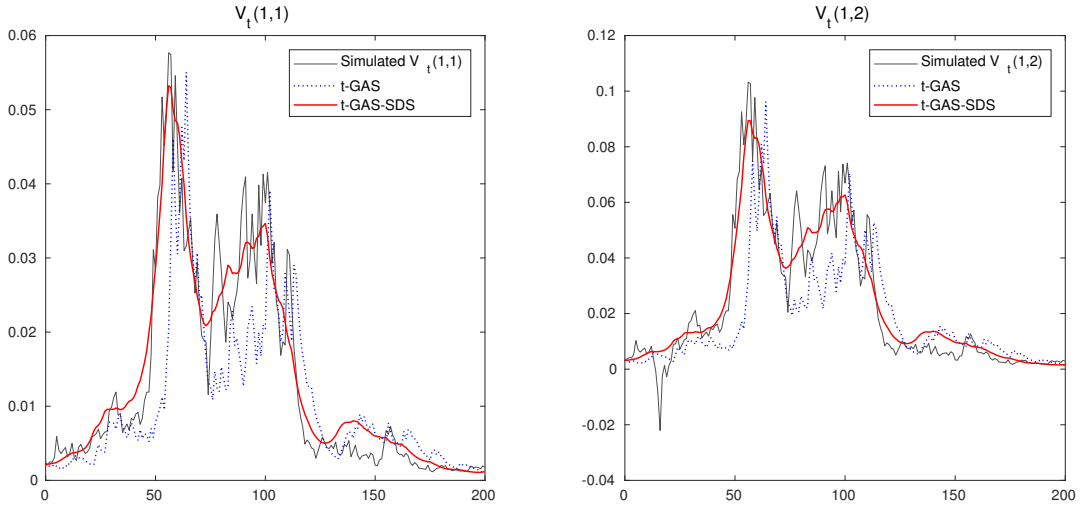


Figure 3.2.2: Comparison among simulated true covariances V_t (black lines), filtered (blue dotted lines) and smoothed (red lines) (co)variances of t-GAS model in the case $k = 5$. We show the variance corresponding to the first asset on the left and the covariance between the first and the second asset on the right.

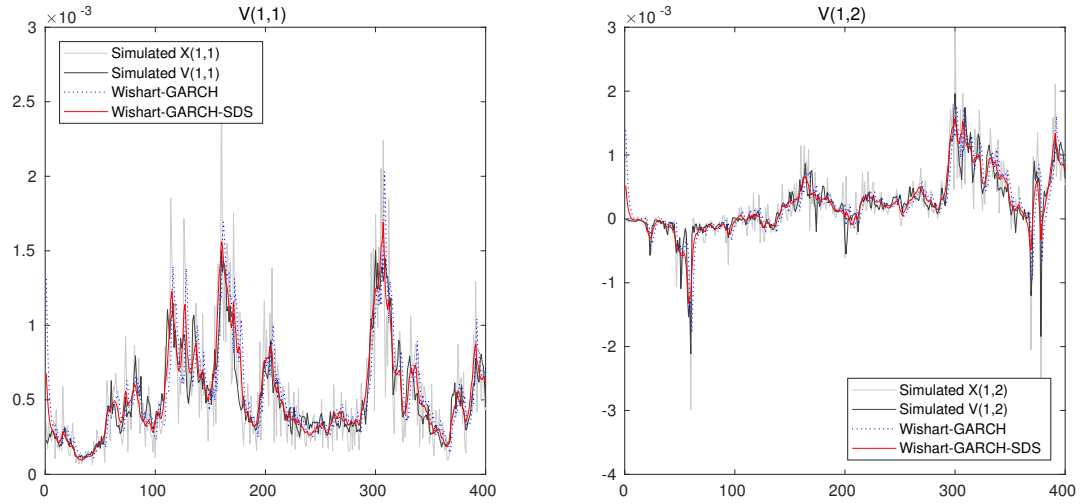


Figure 3.2.3: Comparison among simulated observations of X_t (grey lines), simulated true covariances V_t (black lines), filtered (blue dotted lines) and smoothed (red lines) (co)variances of realized Wishart-GARCH model in the case $k = 5$. We show the variance corresponding to the first asset on the left and the covariance between the first and the second asset on the right.

1. Slow sine: $\beta_1 + \frac{1}{2}\sin(\frac{2\pi t}{N})$
2. Fast sine: $\beta_1 + \frac{1}{2}\sin(\frac{20\pi t}{N})$
3. Ramp: $\beta_1 + \frac{1}{N} \mod (t, \frac{N}{\omega})$
4. Step: $\beta_1 + \frac{1}{2}(t > \frac{N}{2})$
5. Model: $\beta_{t+1} = c + \varphi\beta_t + \xi_t, \quad \xi_t \sim \mathrm{N}(0, \sigma^2)$

$t = 1, \ldots, n$. We set $\omega = 4$, $k = 1$, $c = 0.01$, $\varphi = 0.98$, $\sigma^2 = 0.5$. For some of these dynamic specifications, figure 3.2.1 shows examples of filtered and smoothed estimates of time-varying parameters obtained through the GARCH. As expected, SDS estimates are less noisy than filtered estimates and provide a more accurate reconstruction of time-varying parameters.

	Slow sine	Fast sine	Ramp	Step	Model
			MSE		
			GARCH		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9958	0.9879	0.9888	0.9889	0.9912
SDS	0.4231	0.5398	0.5494	0.5029	0.5289
			MEM		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9928	0.9718	0.9769	0.9794	0.9791
SDS	0.3874	0.4300	0.5239	0.5140	0.5132
			AR(1)		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9954	0.9849	0.9935	0.9966	0.9991
SDS	0.5657	0.5202	0.6257	0.6303	0.6441
			MAE		
			GARCH		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9974	0.9902	0.9975	0.9966	0.9948
SDS	0.6279	0.7017	0.7479	0.7314	0.7351
			MEM		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9957	0.9813	0.9968	0.9962	0.9885
SDS	0.5990	0.6318	0.7370	0.7497	0.7178
			AR(1)		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9991	0.9891	0.9978	0.9988	0.9995
SDS	0.7074	0.7010	0.7940	0.7934	0.7705

Table 3.2.1: Average MSE and MAE of filtered SDU and smoothed SDS estimates relative to standard filtered SDF estimates.

Table 3.2.1 shows average MSE and MAE of SDF, SDU and SDS estimates, for all the patterns considered above. The SDU updates f_t once a new observation y_t becomes available. This translates into a slight improvement over standard filtered estimates. The SDS, using all available observations, significantly improves on SDF estimates, with relative gains larger than 35% and lower than 62% in mean square errors.

We now consider the two multivariate models, namely the t-GAS and the Wishart-GARCH. We compare SDF estimates to SDU and SDS in a simulation setting where $N = 250$ time series of $n = 2000$ daily realized covariance matrices and log-returns are generated as described in Appendix A.3. The aim of the experiment is to estimate the true covariance matrix V_t from observations of daily returns r_t in the t-GAS model and from observations of both daily returns r_t and realized covariance matrices X_t in the Wishart-GARCH model. We consider three scenarios where the number of assets is $p = 5, 10, 20$ respectively, and thus we have $k = 15, 55, 210$ time-varying covariances[2].

For $p = 5$, figures 3.2.2 and 3.2.3 compare SDF and SDS estimates of the $V_t(1,1)$ and $V_t(1,2)$ elements of the simulated covariance matrix in the t-GAS and Wishart-GARCH models, respectively. As in the previous univariate cases, smoothed estimates provide a better reconstruction of the time-varying covariances. Note that, compared to the t-GAS model, the Wishart-GARCH provides estimates which are closer to the simulated V_t, as they are obtained by conditioning on a larger information set.

In order to quantify estimation errors, we use the root mean square error (RMSE) and the quasi-likelihood (Qlike), which are robust loss measures for covariance estimates (Patton 2011). These are defined in Appendix A.3. Table 3.2.2 shows relative RMSE and Qlike gains of SDU and SDS estimates over SDF. We first note that SDU and SDS provide significantly lower RMSE. In the t-GAS model, the relative gain of the SDU is roughly equal to 3%, while the one of SDS is larger than 14% and lower than 19%. In the Wishart-GARCH model, the relative gain of SDU is larger than 7% and lower than 13%, while the one of the SDS is larger than 13% and lower than 19%. It is interesting to note that SDU gains are significantly larger in the Wishart-GARCH model. This is due to the fact that today's realized covariance X_t is a highly informative proxy of V_t, thus leading to drastic RMSE reduction when included in the information set. In contrast, daily returns are less informative and thus it is necessary to include all available observations to achieve significant RMSE reduction in the t-GAS model. If one looks at the Qlike loss, relative gains of SDU and SDS are moderate compared to RMSE but they are statistically significant. Even in this case, SDU gains are larger in the Wishart-GARCH model due to the highly informative content of realized covariance measures.

p	5	10	20
		RMSE	
		t-GAS	
SDF	1.0000	1.0000	1.0000
SDU	0.9723	0.9654	0.9666
SDS	0.8165	0.8202	0.8596
		Wishart-GARCH	
SDF	1.0000	1.0000	1.0000
SDU	0.8733	0.8821	0.9330
SDS	0.8164	0.8070	0.8742
		Qlike	
		t-GAS	
SDF	1.0000	1.0000	1.0000
SDU	0.9989	0.9965	0.9922
SDS	0.9821	0.9888	0.9898
		Wishart-GARCH	
SDF	1.0000	1.0000	1.0000
SDU	0.9858	0.9909	0.9901
SDS	0.9658	0.9700	0.9712

Table 3.2.2: Average root mean square error (RMSE) and quasi-likelihood (Qlike) of SDU and SDS estimates relative to SDF estimates of t-GAS and Wishart-GARCH model

Nominal c.l.	90%	95%	99%
		In-sample	
		Normal	
SDF	0.8356	0.9026	0.9703
SDU	0.8346	0.9022	0.9701
SDS	0.8303	0.8980	0.9680
		Robust	
SDF	0.8968	0.9523	0.9930
SDU	0.8963	0.9520	0.9929
SDS	0.8921	0.9489	0.9918
		Out-of-sample	
		Normal	
SDF	0.8304	0.8975	0.9669
SDU	0.8296	0.8971	0.9667
SDS	0.8268	0.8951	0.9661
		Robust	
SDF	0.8917	0.9483	0.9913
SDU	0.8912	0.9481	0.9912
SDS	0.8890	0.9467	0.9909

Table 3.2.3: Average coverage of in-sample and out-of-sample confidence bands obtained by assuming a normal conditional density and the fat-tailed density described in Section 3.2.2. We report results for three different nominal confidence levels, namely 90%, 95%, 99%.

.

3.2.2 Confidence bands

In Section (3.1.3), we have seen that an estimate of the conditional variance of the state variable is given by the variance matrices J_{t+1}, J_t and $\hat{J}_t$ defined in eq. (3.1.22), (3.1.23), (3.1.25). As in the Kalman filter, one can use these variances to construct confidence bands around filtered and smoothed estimates. However, the conditional density of the state variable is typically fat-tailed and cannot be written in closed form. Assuming normality generally provides narrow confidence bands and thus underestimates filtering uncertainty.

Robust in-sample and out-of-sample confidence bands can be constructed by computing quantiles of a more flexible distribution determined by matching location and scale parameters. We illustrate here an application of this technique in the case of the GARCH. A detailed and systematic treatment can be found in Buccheri et al. (2018c).

Let us consider the following stochastic volatility model:

$$y_t = e^{\frac{\theta_t}{2}}\epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0,1) \tag{3.2.25}$$

$$\theta_{t+1} = \gamma \ + \phi\theta_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0,\sigma_\eta^2) \tag{3.2.26}$$

We are interested in computing quantiles of the conditional density of e^{θ_t}. Filtered and smoothed estimates of the latent log-variance θ_t are recovered by computing the score-driven recursions for the following observation density:

$$p(y_t|f_t) = \frac{1}{\sqrt{2\pi e^{f_t}}} e^{-\frac{y_t^2}{2e^{f_t}}}$$

As an outcome of this procedure, we also obtain the conditional variances J_{t+1}, J_t and $\hat{J}_t$. Let $F_{\theta_t|\mathcal{F}_{t_1}}(\theta) = P(\theta_t < \theta|\mathcal{F}_{t_1})$, $1 < t_1 < n$, be the conditional distribution function of θ_t. The quantile function of e^{θ_t} is then given by:

$$F^{-1}_{e^{\theta_t}|\mathcal{F}_{t_1}} = \exp\left(F^{-1}_{\theta_t|\mathcal{F}_{t_1}}\right) \tag{3.2.27}$$

As a first approximation, we compute quantiles by assuming $\theta_t|\mathcal{F}_{t_1} \sim N(f_{t|t_1}, J_{t|t_1})$. For $t_1 = t-1$, $t_1 = t$, $t_1 = n$ we obtain $N(f_t, J_t)$, $N(f_{t|t}, J_{t|t})$, $N(\hat{f}_t, \hat{J}_t)$, respectively. These conditional densities depend on parameters which are an output of the score-driven recursions and thus confidence bands can be easily computed through eq. (3.2.27) using the Gaussian quantile function. We then assume $\theta_t|\mathcal{F}_{t_1} \sim t(f_{t|t_1}, J_{t|t_1}, \nu)$, i.e. a Student's t-distribution with location $f_{t|t_1}$, scale $J_{t|t_1}$ and ν degrees of freedom. If $\nu \to \infty$, we recover the Gaussian confidence bands. However, if ν is finite, confidence bands will be larger and provide a better approximation to the true filtering uncertainty. In this example, the parameter ν is chosen by fitting a t distribution on the residuals of an AR(1) model estimated on $\hat{f}_t$. More sophisticated techniques are developed in Buccheri et al. (2018c).

[2]We implement the t-GAS model using hyperspherical coordinates, and thus we have $k = p(p+1)/2)$ time-varying covariances.

In order to test the quality of confidence bands, we generate 1000 time series of $n = 4000$ observations of the stochastic volatility model (3.2.25), (3.2.26). The values of static parameters are chosen in order to be similar to those obtained when estimating the model on real financial returns: $\gamma = 0.001$, $\phi = 0.98$, $\sigma_\eta^2 = 0.02$. Figure (3.2.4) shows one of the simulated patterns, together with 95% confidence bands for filtered and smoothed estimates computed through the method described above.

We estimate the GARCH in the sub-sample comprising the first 2000 observations and construct in-sample confidence bands. In the remaining sub-sample of 2000 observations, out-of-sample bands are constructed using previous parameter estimates. Both Gaussian and robust confidence bands are built at 90%, 95%, 99% nominal confidence levels. We compare the nominal confidence level to the coverage, defined as the fraction of times the true variance path e^{θ_t} is inside the confidence bands. Table (3.2.3) shows average coverages for in-sample and out-of-sample SDF, SDU and SDS confidence bands. As expected, confidence bands constructed by assuming a Gaussian density provide an average coverage which is significantly lower than the nominal confidence level, meaning that they underestimate filtering uncertainty. In contrast, the average coverage of robust confidence bands is very close to the nominal coverage. Similar results are found when changing the variance σ_η^2 in the latent process. In particular, for larger values of σ_η^2, the quality of Gaussian confidence bands further deteriorates, while robust bands still provide a good matching to the nominal level. A systematic treatment of the technique described here and, more generally, of filtering uncertainty in observation-driven models can be found in Buccheri et al. (2018c).

3.3 Monte Carlo analysis

In this Section we perform extensive Monte Carlo simulations to test the performance of the SDS under different dynamic specifications for the time-varying parameters. Since we interpret the SDS as an approximate smoother for nonlinear non-Gaussian models, we compare its performance to that of correctly specified parameter-driven models. The main idea is to examine the extent to which the approximation leads to similar results as correctly specified parameter-driven models. In this case, the use of the SDS would be particularly advantageous from a computational point of view, as the likelihood of score-driven models can be written in closed form and smoothing can be performed through a simple backward recursion. This analysis is similar in spirit to that of Koopman et al. (2016), who compared score-driven models to correctly specified parameter-driven models and found that the two classes of models have similar predictive accuracy, with very small average losses. We find a similar result for the SDS.

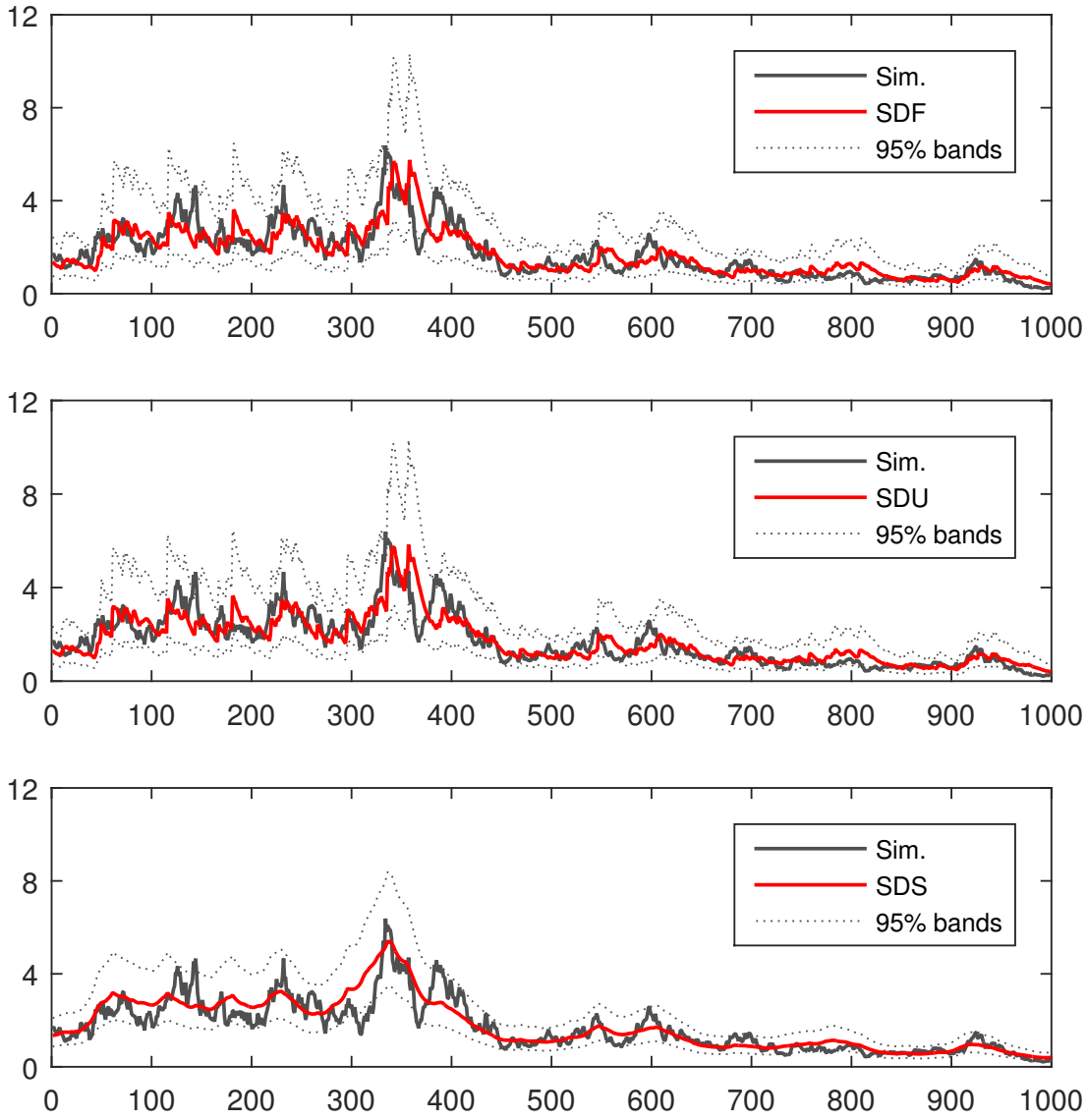


Figure 3.2.4: Simulated volatility and SDF, SDU, SDS estimates together with 95% in-sample confidence bands computed through the technique described in Section 3.2.2.

3.3.1 Linear non-Gaussian models

We first consider an AR(1) model with a t-distributed measurement error:

$$y_t = \alpha_t + \epsilon_t, \quad \epsilon_t \sim t(0, \sigma_\epsilon^2, \nu) \tag{3.3.1}$$

$$\alpha_{t+1} = c + \phi\alpha_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2) \tag{3.3.2}$$

We choose $c = 0.01$ and $\phi = 0.95$. The signal-to-noise ratio is defined as $\delta = \frac{\sigma_\eta^2}{\sigma_\epsilon^2}$. The corresponding observation driven model is a t-location model (Harvey 2013) with measurement density:

$$p(y_t|f_t;\varphi,\beta) = \frac{\Gamma[(\beta+1)/2]}{\Gamma(\beta/2)\varphi\sqrt{\pi\beta}}\left[1 + \frac{(y_t - f_t)^2}{\beta\varphi^2}\right]^{-(\beta+1)/2} \tag{3.3.3}$$

Setting $S_t = \mathcal{I}_{t|t-1}^{-1}$, eq. (3.1.14) reduces to:

$$f_{t+1} = \omega + A(\beta+3)\frac{y_t - f_t}{\beta + \left(\frac{y_t - f_t}{\varphi}\right)^2} + Bf_t \tag{3.3.4}$$

while the smoothing recursions (3.1.15), (3.1.16) reduce to:

$$r_{t-1} = (\beta+3)\frac{y_t - f_t}{\beta + \left(\frac{y_t - f_t}{\varphi}\right)^2} + (B - A)'r_t \tag{3.3.5}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{3.3.6}$$

$t = N, \ldots, 1$. We compare standard Kalman filtered and smoothed estimates to SDF, SDU and SDS estimates. Similarly to previous simulation studies, we generate 1000 time series of 4000 observations and use the first subsample of 2000 observations for estimation and the remaining observations for testing. Table 3.3.1 shows relative MSE and MAE for different values of ν. Note that SDF, SDU and SDS provide better estimates than standard Kalman filter and smoother. In particular, we observe large differences for low values of ν, where the t-distribution strongly deviates from the Gaussian, and for low values of δ, at which accounting for the non-normality of the measurement error becomes more important. Note also that gains of SDS over Kalman smoother estimates are larger than gains of SDF over the Kalman filter for low ν and δ. These results confirm the ability of the SDS to provide robust smoothed estimates of time-varying parameters, to the same extent as the SDF provides robust filtered estimates of time-varying parameters in presence of a non-Gaussian prediction density.

3.3.2 Nonlinear non-Gaussian models

We now examine the behavior of the SDS in presence of nonlinear non-Gaussian parameter-driven models. In particular, we consider the following three specifications, which are quite popular in the econometric literature:

	SDF − KF(p)			SDU − KF(u)			SDS − KS		
δ	0.1	1	10	0.1	1	10	0.1	1	10
	$\nu = 3$								
MSE	0.8610	0.9522	0.9991	0.8074	0.8357	1.0223	0.8093	0.8876	0.9618
MAE	0.9389	0.9859	1.0036	0.9110	0.9439	1.0150	0.9128	0.9634	1.0169
	$\nu = 5$								
MSE	0.9552	0.9912	1.0032	0.9407	0.9681	1.0032	0.9376	0.9880	1.0058
MAE	0.9792	0.9973	0.9999	0.9720	0.9886	1.0032	0.9698	0.9949	1.0112
	$\nu = 8$								
MSE	0.9877	0.9981	1.0029	0.9895	0.9928	1.0107	0.9844	0.9954	1.0117
MAE	0.9939	0.9992	1.0039	0.9927	0.9973	1.0070	0.9917	0.9982	1.0136

Table 3.3.1: Average MSE and MAE of SDF, SDU, SDS relative to Kalman filtered and smoothed estimates of AR(1) model plus non-Gaussian noise.

1. Stochastic volatility model with Gaussian measurement density:

$$y_t = e^{\frac{\theta_t}{2}}\epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0,1)$$
$$\theta_{t+1} = \gamma + \phi\theta_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0,\sigma_\eta^2)$$

2. Stochastic volatility with non-Gaussian measurement density:

$$y_t = e^{\frac{\theta_t}{2}}\epsilon_t, \quad \epsilon_t \sim t(0,1,\nu)$$
$$\theta_{t+1} = \gamma \ + \phi\theta_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0,\sigma_\eta^2)$$

3. Stochastic intensity model with Poisson measurement density:

$$p(y_t|\lambda_t) = \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!}, \quad \theta_t = \log\lambda_t$$
$$\theta_{t+1} = \gamma + \phi\theta_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0,\sigma_\eta^2)$$

In order to estimate the two stochastic volatility models, Harvey et al. (1994) proposed a Quasi Maximum Likelihood method (QML) based on a Gaussian quasi-likelihood. Normality is assumed for the linearized model. The latter is thus susceptible of treatment with the Kalman filter and the method can be viewed as providing approximate filtered and smoothed estimates. It is therefore interesting to compare the performance of the QML to that of the SDS.

The three models above are estimated through importance sampling (IS hereafter). As discussed by Durbin and Koopman (2000), IS methods are simple and effective for problems in time series analysis. They are based on independent samples rather than Markov chains, thus enabling to obtain accurate estimates of Monte-Carlo variances. Sandmann and Koopman (1998) devised an IS technique to evaluate the full likelihood function of stochastic volatility models. We estimate models 1 and 2 by employing

the same IS approach but use the recently developed "Numerically Accelerated Importance Sampling" (NAIS) technique of Koopman et al. (2015b) to choose the parameters of the importance density. This method has been shown to provide several efficiency gains compared to other IS approaches. The stochastic intensity model can also be estimated through IS, as described e.g. by Durbin and Koopman (1997). Similarly to the previous cases, we choose the parameters of the importance density through the NAIS. More details on IS techniques for nonlinear non-Gaussian state-space models can be found in Durbin and Koopman (2012).

We set the measurement densities of the corresponding observation-driven models as indicated below.

1. For the two stochastic volatility models:

$$p(y_t|f_t) = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\Gamma\left(\frac{\beta}{2}\right)\sqrt{\pi\beta e^{f_t}}}\left[1+\frac{1}{\beta}\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2\right]^{-\frac{\beta+1}{2}} \tag{3.3.7}$$

2. For the stochastic intensity model:

$$p(y_t|f_t) = e^{-e^{f_t}}\frac{e^{f_t y_t}}{y_t!} \tag{3.3.8}$$

The use of a t distribution for the first model is motivated by the fact that even Gaussian stochastic volatility models are able to generate a predictive density with fat-tails and over-dispersion (Carnero et al. 2004). Thus, in order for the observation-driven model to capture these dynamic features, we adopt a more flexible specification for the measurement density. This is in line with Koopman et al. (2016), who used more flexible densities for observation-driven counterparts of parameter-driven models. Note that the measurement density (3.3.7) is similar to the one leading to the Beta-t-EGARCH model of Harvey and Chakravarty (2008).

In the first case, setting $S_t = \mathcal{I}_{t|t-1}^{-1}$, the filtering recursion (3.1.14) reduces to:

$$f_{t+1} = \omega + A\frac{\beta+3}{\beta}\left[\frac{\beta+1}{\beta}\frac{\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2}{1+\frac{1}{\beta}\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2}-1\right]+Bf_t \tag{3.3.9}$$

while the smoothing recursions (3.1.15), (3.1.16) reduce to:

$$r_{t-1} = \frac{\beta+3}{\beta}\left[\frac{\beta+1}{\beta}\frac{\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2}{1+\frac{1}{\beta}\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2}-1\right]+(B-A)'r_t \tag{3.3.10}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{3.3.11}$$

In the case of the measurement density in eq. (3.3.8), setting $S_t = \mathcal{I}_{t|t-1}^{-1}$, the filtering recursion (3.1.14) reduces to:

$$f_{t+1} = \omega + A(e^{-f_t}y_t - 1) + Bf_t \tag{3.3.12}$$

while the smoothing recursions (3.1.15), (3.1.16) reduce to:

$$r_{t-1} = e^{-f_t} y_t - 1 + (B - A)' r_t \tag{3.3.13}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{3.3.14}$$

Estimation of parameter-driven models is performed through the NAIS with $S = 200$ simulations. We also introduce control variables as described by Koopman et al. (2015b). Smoothed estimates are computed with $G = 200$ simulations. Larger values of S and G do not lead to significant improvements of parameter-driven model estimates compared to SDS estimates. The simulation setting is the same as in the previous experiment, with 1000 simulated time series of 4000 observations. Figure 3.3.1 compares smoothed estimates obtained through both IS and the SDS for the three different models at hand. A simple visual inspection shows that the estimates provided by the two methods are very close.

In order to examine in more detail differences between SDS and IS estimates, tables 3.3.2, 3.3.3, 3.3.4 show the results of Monte Carlo experiments for the three models. In the case of stochastic volatility models, we consider several scenarios by varying the autoregressive coefficient ϕ and the coefficient of variation CV. The latter is defined as in Sandmann and Koopman (1998), namely:

$$CV = \exp\left(\frac{\sigma_\eta^2}{1 - \phi^2}\right) - 1 \tag{3.3.15}$$

CV	0.1	1	5	10	0.1	1	5	10
	MSE				MAE			
	$\phi = 0.98$							
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	0.9988	1.0050	1.0001	1.0162	1.0004	1.0043	1.0017	1.0097
QML	1.4153	1.3880	1.3333	1.3138	1.1797	1.1739	1.1564	1.1475
	$\phi = 0.95$							
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0057	0.9983	0.9988	1.0059	1.0034	1.0023	1.0024	1.0059
QML	1.3131	1.3737	1.3246	1.3168	1.1450	1.1758	1.1567	1.1524
	$\phi = 0.90$							
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0076	0.9956	0.9974	1.0086	1.0044	1.0010	1.0033	1.0093
QML	1.2371	1.3157	1.2893	1.2750	1.1109	1.1508	1.1422	1.1370

Table 3.3.2: Average MSE and MAE of IS, SDS and QML smoothed estimates normalized by IS loss in case of stochastic volatility model with Gaussian measurement density.

Note that CV is related to the variance of the signal innovation. The values of both ϕ, CV and of the remaining parameters are chosen to be close to those estimated on real financial time series, as discussed in Sandmann and Koopman (1998). For the stochastic intensity model, we consider scenarios characterized by different autoregressive coefficients ϕ and different values of the variance σ_η^2 of the

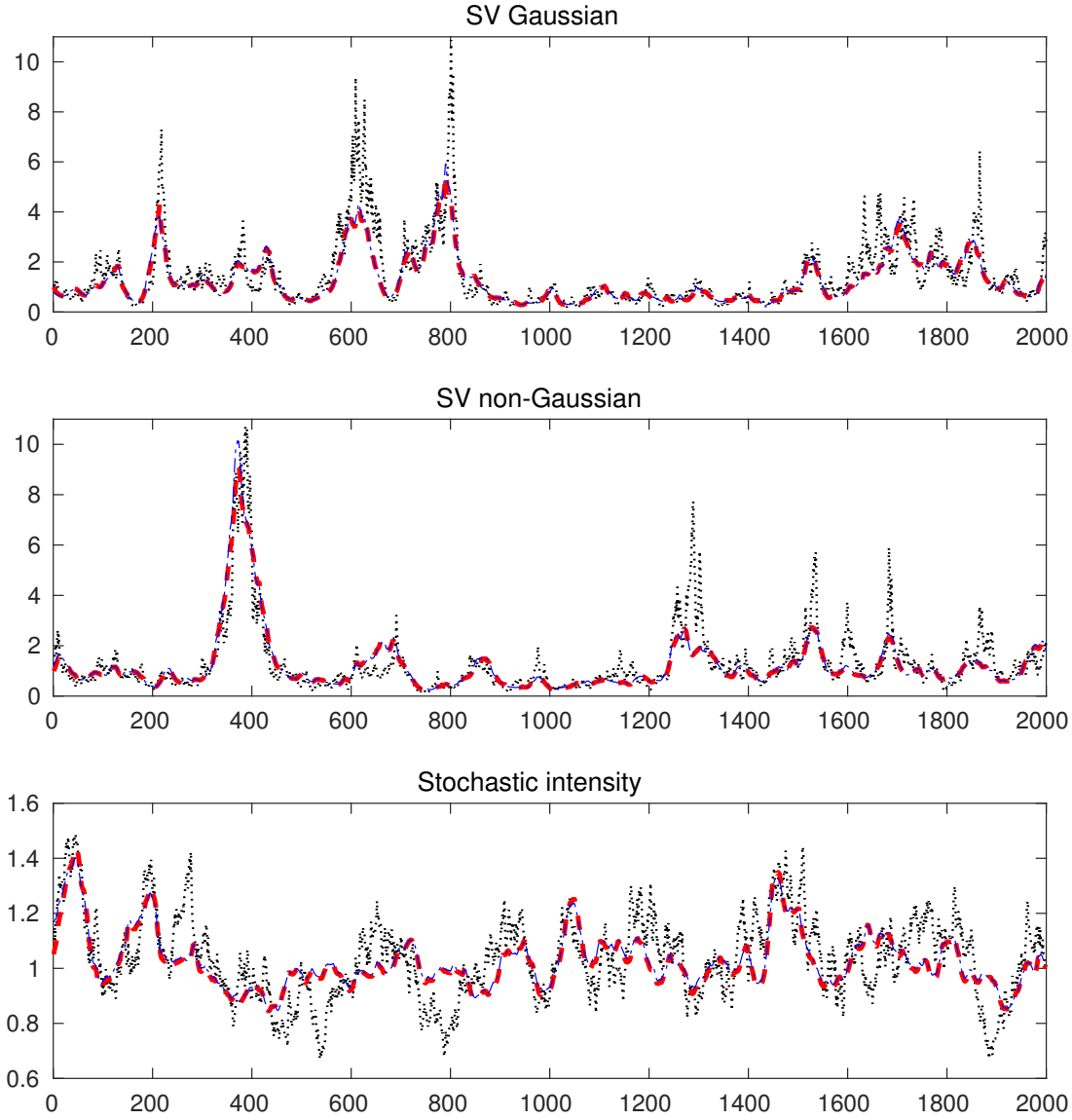


Figure 3.3.1: Comparison among simulated unobserved components (black dotted lines), IS smoothed estimates (red dashed lines) and SDS smoothed estimates (blue dotted and dashed).

signal. In the case of the two stochastic volatility models, the SDS largely outperforms the QML in all scenarios. The performance of the latter tends to worsen as CV decreases, according to the fact that the non-normality of the measurement equation becomes more relevant for low CV. Compared to IS, the relative MSE loss of the SDS is very small. In particular, it is always less than 2% in the Gaussian case, while it is always lower than 2.5% in the non-Gaussian case. Larger losses are observed for large values of CV, where σ_η is large and accounting for the non-normality of observations is less relevant.

CV	0.1	1	5	10	0.1	1	5	10
	MSE				MAE			
	$\phi = 0.98$, $\nu = 3$							
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0026	0.9950	1.0140	1.0169	1.0015	0.9997	1.0077	1.0098
QML	1.3962	1.2553	1.2125	1.1998	1.1735	1.1184	1.1013	1.0939
	$\phi = 0.95$, $\nu = 3$							
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0014	1.0049	1.0121	1.0200	1.0008	1.0031	1.0064	1.0105
QML	1.3058	1.2639	1.2447	1.2246	1.1354	1.1230	1.1158	1.1056
	$\phi = 0.90$, $\nu = 3$							
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0020	1.0033	1.0149	1.0221	1.0016	1.0023	1.0081	1.0117
QML	1.2306	1.2325	1.2262	1.2200	1.1026	1.1075	1.1062	1.1034

Table 3.3.3: Average MSE and MAE of IS, SDS and QML smoothed estimates normalized by IS loss in case of stochastic volatility model with non-Gaussian measurement density.

In the case of the stochastic intensity model, we observe a similar behavior but the relative MSE loss is slightly larger for $\phi = 0.98$ and $\sigma_\eta^2 = 0.01$, where it is found to be around 5%. Overall, average MSE losses are less than 2.5% if one averages across all scenarios. This result is in agreement with what Koopman et al. (2016) found by comparing the prediction performance of score-driven models to that of correctly specified parameter-driven models.

Finally, it is interesting to look at computational times. Table 3.3.5 shows average computational times of IS relative to those of SDS. We report both the time required to estimate the parameters and that required for smoothing. Estimation of static parameters is much faster, as the likelihood can be computed in closed form. In contrast, estimating correctly specified parameter-driven models with $S = 200$ simulations is on average 80 times slower. Note that decreasing S would lead to faster estimates at the expense of reducing efficiency. Smoothing with IS is on average 215 times slower compared to smoothing with the SDS. Note also that, while it is generally difficult to extend the NAIS and other IS methods to a setting with multiple time-varying parameters, the SDS maintains the same form in case f_t is a large vector of time-varying parameters. Thus, using the SDS allows to obtain smoothed estimates which are very close to those of correctly specified parameter-driven models but reducing considerably

$\sigma_\eta^2 \times 100$	0.1	0.5	1	0.1	0.5	1
	MSE			MAE		
	$\phi = 0.98$					
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0149	1.0281	1.0521	1.0067	1.0132	1.0244
	$\phi = 0.95$					
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0203	1.0176	1.0254	1.0097	1.0083	1.0120
	$\phi = 0.90$					
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0310	1.0160	1.0205	1.0142	1.0079	1.0099

Table 3.3.4: Average MSE and MAE of IS and SDS smoothed estimates normalized by IS loss in case of stochastic intensity model with Poisson measurement density.

the computational burden.

	SV Gaussian	SV Fat tail	Intensity
Estimation	45.35	107.90	82.33
Smoothing	179.74	229.79	236.34

Table 3.3.5: Average computational times of IS relative to SDS.

3.4 Empirical illustration

It is interesting to investigate whether the results found in the simulation study in Section 3.2 also hold on empirical data. In particular, we aim to provide a quantitative assessment of the improvement of SDU and SDS estimates over standard score-driven estimates in a problem of empirical relevance. Unlike the simulation study, it is generally difficult to perform such analysis empirically, given that time-varying parameters do not belong to the econometrician's information set and cannot be employed as a benchmark in the loss function. However, when dealing with conditional covariance estimates computed from *daily* log-returns, one can use realized covariance computed from *intraday* log-returns as an accurate proxy of the true latent covariance (Andersen and Bollerslev 1997). Loss functions can therefore be built as if covariances were observed. This empirical analysis will also show the advantages of the SDS methodology in an highly multivariate framework, where the use of simulation-based methods is computationally problematic or even unfeasible.

Our dataset consists of unbalanced 1-minute transaction data of Russel 3000 constituents over the period from 18-11-1999 to 27-09-2013. The total number of assets is 4166. The analysis is performed on the subsample comprising the last $T = 2000$ days, in order to avoid discontinuities due to changes on index composition. We consider trades from 9:30 to 16:00, leading to 390 timestamps per day. Assets

	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
	RMSE$\times 10^7$			
SDF	0.7967*	0.2904	0.2181	0.9222
	1.0000	1.0000	1.0000	1.0000
SDU	0.7967*	0.2806*	0.1997	0.8976*
	0.9733	0.9665	0.9157	0.9732
SDS	0.7445*	0.2791*	0.1732*	0.8101*
	0.9345	0.9611	0.7941	0.8784
	Qlike			
SDF	-34.1291	-37.3352	-32.4399	-37.7819
	1.0000	1.0000	1.0000	1.0000
SDU	-34.2181	-37.4292	-32.6133	-37.9202*
	0.9974	0.9975	0.9947	0.9964
SDS	-34.3482*	-37.5424*	-32.6987*	-37.9751*
	0.9936	0.9945	0.9921	0.9949

Table 3.4.1: Absolute and relative RMSE and Qlike of SDF, SDU, SDS estimates of the t-GAS model for the four randomly selected portfolios with $n = 5$. The asterisk implies that the estimator is included in the model confidence set.

having less than 10 trades per day are excluded in order to avoid poor and ill-conditioned realized covariance estimates. As a final outcome of our filtering procedure, we obtain $N = 1682$ assets.

In order to estimate conditional covariances from daily log-returns, we use the t-GAS model of Creal et al. (2011) described in Section (3.2). Compared to standard conditional covariance models, the main advantage of the t-GAS is that it updates covariances by taking into account the full shape of the Student t observation density and thus provides robustness against outliers (see discussions on Creal et al. 2011). We implement the parameterization based on hyperspherical coordinates, as it generally leads to better estimates. The number of time-varying parameters grows as p^2, where p is the number of assets.

Among the universe of $N = 1682$ assets, we select random groups of $p = 5, 10, 20$ assets. In particular, for each cross-section dimension p, we randomly choose four groups. The analysis is thus performed on 12 different groups of assets. As done in the simulation study, we use the RMSE and Qlike as loss functions. The benchmark used in the loss function is the realized covariance estimator of Barndorff-Nielsen and Shephard (2004) computed at the 5-minutes sampling frequency. The use of other estimators does not alter the outcome of the experiment. For each group of assets, the t-GAS is estimated on the time-series of $T = 2000$ open-to-close log-returns. We thus compute (i) the predictive filter f_t, (ii) the update filter $f_{t|t}$ and (iii) the smoother $\hat{f}_t$. The statistical significance of loss differences is tested through the model confidence set of Hansen et al. (2011) at the 90% confidence level.

Tables 3.4.1, 3.4.2, 3.4.3 show the results of the analysis, for $p = 5$, 10, 20, respectively. We first note

	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
	RMSE$\times 10^5$			
SDF	0.1271	0.2080	0.3105	0.1398
	1.0000	1.0000	1.0000	1.0000
SDU	0.1229*	0.2021*	0.3008	0.1356*
	0.9662	0.9716	0.9688	0.9704
SDS	0.1189*	0.1946*	0.2868*	0.1345*
	0.9354	0.9355	0.9237	0.9623
	Qlike			
SDF	-65.5001	-68.3171	-70.4809	-70.9020
	1.0000	1.0000	1.0000	1.0000
SDU	-65.7213	-68.4716	-70.6554	-71.0882
	0.9974	0.9977	0.9975	0.9974
SDS	-65.8784*	-68.6694*	-70.9327*	-71.2838*
	0.9936	0.9949	0.9936	0.9946

Table 3.4.2: Absolute and relative RMSE and Qlike of SDF, SDU, SDS estimates of the t-GAS model for the four randomly selected portfolios with $n = 10$. The asterisk implies that the estimator is included in the model confidence set.

that covariance estimates constructed through the predictive filter feature larger RMSE and Qlike and are always excluded from the model confidence set. They are in fact less informative, since only past log-returns are used when reconstructing time-varying parameters. Covariance estimates built through $f_{t|t}$ and $\hat{f}_t$ are both included in the model confidence set constructed through the RMSE. If the Qlike is used, only smoothed estimates are included. The latter provide a better reconstruction of realized covariance, suggesting that future observations contain relevant information on today's covariance. Note also that relative gains are similar across different dimensions, meaning that the SDS is not affected by the proliferation of time-varying parameters when the number of assets increases.

The above results suggest that, when extracting latent covariance, the smoothing provided by the t-GAS model is effective in aggregating all available information. Compared to standard score-driven filtered estimates, the update filter $f_{t|t}$ and the smoother $\hat{f}_t$ can thus be regarded as providing a more accurate estimate of latent covariance. As seen in the simulation study, this is true for a large class of dynamic models. We thus use these results to solicit the use of the SDS in place of standard filtered estimates in signal reconstruction analysis.

	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
	RMSE$\times 10^5$			
SDF	0.7061	0.3705	0.2146	0.3255
	1.0000	1.0000	1.0000	1.0000
SDU	0.6950*	0.3634*	0.2099	0.3170*
	0.9843	0.9808	0.9782	0.9739
SDS	0.6709*	0.3565*	0.1998*	0.3046*
	0.9501	0.9622	0.9311	0.9358
	Qlike			
SDF	-134.7951	-128.6277	-140.1625	-137.9562
	1.0000	1.0000	1.0000	1.0000
SDU	-135.1702	-129.0375	-140.5694	-138.2863
	0.9972	0.9968	0.9971	0.9976
SDS	-135.7584*	-129.3180*	-141.2055*	-138.8856*
	0.9929	0.9947	0.9926	0.9933

Table 3.4.3: Absolute and relative RMSE and Qlike of SDF, SDU, SDS estimates of the t-GAS model for the four randomly selected portfolios with $n = 20$. The asterisk implies that the estimator is included in the model confidence set.

Chapter 4

A multi-asset price formation model

Almost all results in this chapter previously appeared in Buccheri et al. (2018b).

The dynamics of high-frequency asset prices are known to be characterized by "lead-lag effects": some assets (the laggers) tend to follow the movements of other assets (the leaders). This phenomenon is of fundamental interest in market microstructure research and in high-frequency financial econometrics. However, while it has received some attention in the empirical finance literature (see e.g. Chan 1992, de Jong and Nijman 1997, Chiao et al. 2004, Huth and Abergel 2014, Dobrev and Schaumburg 2017) and in the statistical literature (Hoffmann et al. 2013, Hayashi and Koike 2016, Hayashi and Koike 2017), there is still a lack of econometric approaches aiming to describe lead-lag effects from a market microstructure perspective. On the one hand, there is no well-established microstructure theory explaining the existence of lead-lag effects. On the other hand, compared to the case of low-frequency (e.g. daily) data, the estimation of both contemporaneous and lagged correlations among assets traded at high-frequency is a more complex task. This is mainly due to the presence of microstructure noise and asynchronous trading, which prevent the use of traditional multivariate techniques[1].

Motivated by the empirical evidence of cross-asset trading (Hasbrouck and Seppi 2001, Bernhardt and Taub 2008, Pasquariello and Vega 2015), i.e. the fact that dealers may rely on the prices of other securities when setting their quotes, we introduce a multi-asset price formation mechanism which generalizes well known univariate microstructure models of lagged price adjustment (see Hasbrouck, 1996 for a review on lagged price adjustment models). We name our model Multi-asset Lagged Adjustment (MLA). The MLA is a micro-founded model where lead-lag correlations among high-frequency returns naturally arise as a result of the multivariate nature of the price formation process. Econometric inference on the MLA allows to test for the presence of lead-lag correlations in the latent price process or, equivalently,

[1] For instance, the so-called Epps effect Epps (1979), i.e. the bias towards zero of (contemporaneous) sample correlations as the sampling frequency increases, is due to asynchronous trading.

for the existence of a non-trivial multi-asset price generation mechanism. Interestingly, by separating the estimation of lead-lag correlations from contemporaneous correlations, we obtain an estimate of the integrated covariance of the efficient martingale price process that is robust to microstructure noise, asynchronous trading and lead-lag effects.

Lagged price adjustment models, also known as partial price adjustment models, were proposed, amongst others, by Hasbrouck and Ho (1987), Amihud and Mendelson (1987) and Damodaran (1993). The theoretical concept underlying these models is that prices do not instantaneously adjust when new information arrives. Instead, due to lagged dissemination of information and price smoothing by market dealers, the adjustment process is delayed. Hasbrouck and Ho (1987) introduced a lagged adjustment price process that allows to describe return autocorrelations at orders greater than one, as observed on real transaction data. We extend this idea to a multivariate framework by viewing the price formation process as a genuine multi-asset process where information related to other assets affects the price discovery process of a given asset. By doing so, we establish a link between the market microstructure literature on lagged price adjustment and that on cross-asset trading.

The concept of cross-asset trading (also known as cross-asset pricing or cross-asset learning) has been extensively exploited by researchers since the seminal work of Caballé and Krishnan (1994), who developed a model of insider trading based on the informational assumption that market makers can learn about one security from observing all order flows in the market. Based on cross-asset learning, Cespa and Focault (2011) developed a transmission mechanism of liquidity shocks among many stocks, the so-called "liquidity spillovers". Pasquariello and Vega (2015) described the relation between cross-price impact and informed multi-asset trading by assuming that dealers in one security can condition on prices of all other securities. Finally, common factors in the price discovery process have been investigated by Hasbrouck and Seppi (2001), Harford and Kaul (2005), Andrade et al. (2008) and Tookes (2008).

Econometric inference on the MLA can be conveniently carried out by casting the model into a state-space representation. The transition equation is a VAR(1) process for the returns of the "adjusted" price while the observation equation incorporates microstructure effects as an additive noise term. Estimation is performed through a Kalman-EM algorithm which easily handles missing observations. Thus, asynchronicity can be treated as a typical missing value problem, in a similar fashion to Corsi et al. (2015) and Shephard and Xiu (2016). This approach allows to estimate the parameters using all available observations and avoids the use of standard synchronization schemes. The latter may introduce spurious lead-lag correlations or destroy true short-term lead-lag effects. As shown in Section 4.2, the MLA estimator of lead-lag correlations is robust to asynchronous trading and to differences in the level of trading activity.

In the MLA, there is a one-to-one correspondence between the VAR matrix of lead-lag coefficients

and the speed of adjustment matrix in the lagged price adjustment process. As such, the presence of statistically significant lead-lag correlations can be interpreted as an evidence of the existence of a multi-asset price formation mechanism. Due to the VAR structure in the transition equation, the MLA can detect "latent" Granger causality, i.e. causality relationships among noisy and asynchronous observations.

In Section 4.3 the MLA is tested on a cross-section of NYSE tick data. We provide empirical evidence for the existence of a multi-asset price formation mechanism. This is done by recovering the speed of adjustment matrix from the estimated lead-lag correlations and showing that it contains statistical significant elements. A likelihood ratio test proves that the null hypothesis of a standard random walk plus noise model, along the lines of Corsi et al. (2015) and Shephard and Xiu (2016), is not able to capture relevant features of the dynamics of high-frequency prices. In particular, deviations from the null assumption of a random walk are more pronounced in periods of large volatility. In a similar fashion to Dobrev and Schaumburg (2017), this empirical finding can be interpreted in light of high-frequency trading: volatility can create short-living cross-autocorrelations that are exploited by short-term, high-frequency trading strategies. A positive relation between volatility and high-frequency trading is found by Zhang (2010), amongst others. We finally examine in detail the cross autocorrelation structure of the market and find that, in contrast to what is typically found with alternative, non-robust estimators, even assets characterized by lower trading activities can lead the dynamics of more liquid assets.

4.1 Theoretical framework

4.1.1 The MLA

A significant portion of the empirical research on market microstructure has been devoted to understanding the autocorrelation structure of univariate and multivariate high-frequency return series. There is well-established evidence of three key empirical properties: strong negative first-order autocorrelation in the return series, existence of positive autocorrelation at lags greater than one and, finally, existence of lead-lag correlations. Simple bid-ask models such as the model of Roll (1984) reproduce the negative first-order autocorrelation observed in the return series. Univariate bid-ask models were later generalized to capture correlations at orders greater than one through the introduction of lagged price adjustments (Hasbrouck, 1996). Here, we consider a multi-asset version of a model with lagged price adjustment that is also able to keep into account lead-lag correlations.

We assume that the efficient log-price P_t is a d-dimensional vector that evolves as Brownian semi-martingale defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$:

$$P_t = \int_0^t \mu_s ds + \int_0^t \sigma dW_s, \quad \Sigma = \sigma\sigma' \tag{4.1.1}$$

where $t \in [0, T]$, μ_s is a vector of predictable locally bounded drifts, σ is a volatility matrix and W_s is a vector of independent Brownian motions. The interval $[0, T]$ can be thought of as representing the trading day.

Let $0 \leq t_1, \ldots, t_n \leq T$ denote n equally-spaced observation times. Opposed to P_t, we consider the d-dimensional observed log-price process Y_{t_i}. The difference $\tau = t_{i+1} - t_i$ between consecutive observation times is assumed to be a very short time interval (e.g. $\tau = 1$ sec. in our empirical application). Note that, because of asynchronous trading, only the components of Y_{t_i} corresponding to traded assets are observed at time t_i, $i = 1, \ldots n$. Observations of other assets are missing. For simplicity, we assume that the drift term in eq. (4.1.1) is zero[2]. We can write:

$$P_{t_{i+1}} = P_{t_i} + u_{t_i}, \quad u_{t_i} \sim \text{NID}(0, \tau\Sigma) \tag{4.1.2}$$

These are the prices that, abstracting from microstructure effects, would be observed in a perfect market, i.e. one in which prices instantaneously react to new information. In real markets, dealers do not instantaneously adjust their quotes to new information. Instead, the adjustment process is gradual and reflects lagged dissemination of information and several market imperfections, such as trading costs, discreteness and price smoothing by market makers. In addition, due to cross-asset trading (Hasbrouck and Seppi 2001, Bernhardt and Taub 2008, Pasquariello and Vega 2015), dealers tend to look at more informative securities before setting their quotes. In order to capture lagged dissemination of information *across* stocks, we start from the simple univariate lagged adjustment mechanism proposed by Hasbrouck and Ho (1987) and adapt it to a multi-asset framework.

Let X_{t_i}, $i = 1, \ldots, n$ denote a d-dimensional vector of "adjusted" prices reflecting the imperfections of the trading process. We assume that X_{t_i} is related to the efficient log-price process P_{t_i} by:

$$X_{t_{i+1}} = X_{t_i} + \Psi(P_{t_{i+1}} - X_{t_i}) \tag{4.1.3}$$

where Ψ is a $d \times d$ matrix characterizing the speed of adjustment of X_{t_i} to the true efficient log-price P_{t_i}. If $\Psi = I_d$, then $X_{t_i} = P_{t_i}$ and the adjustment process is instantaneous. Instead, if $\Psi \neq I_d$ the adjustment process is gradual and, as a result, there is a delay between X_{t_i} and P_{t_i}. Note that the matrix Ψ may be non-diagonal. This implies that the adjustment process of one asset is affected by the adjustment process of other assets and the strength of this effect is quantified by the non-diagonal elements of Ψ.

Due to the presence of market microstructure effects (e.g. bid-ask bounces), the observed log-price process Y_{t_i} deviates from the lagged price X_{t_i}. Therefore, we assume that X_{t_i} is observed under additive noise:

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}, \quad \epsilon_{t_i} \sim \text{NID}(0, H) \tag{4.1.4}$$

[2]This assumption is not too restrictive since we are considering ultra-high-frequency returns for which drift effects are negligible.

where ϵ_{t_i} is a normal white noise term summarizing microstructure effects. In line with Corsi et al. (2015) and Shephard and Xiu (2016), the noise covariance matrix H is assumed to be diagonal. However, the model can be easily estimated even under a non-diagonal noise term. Denoting by $\Delta X_{t_{i+1}} = X_{t_{i+1}} - X_{t_i}$ the log-returns of the lagged price, eq. (4.1.2) and (4.1.3) imply:

$$\Delta X_{t_{i+1}} = (I_d - \Psi)\Delta X_{t_i} + \Psi u_{t_i} \tag{4.1.5}$$

that is, a first order vector autoregressive VAR(1) process. If Ψ is non-diagonal, the knowledge at time t_i of the return of one asset is useful for forecasting the return of another asset at time t_{i+1}. Therefore, in this multi-asset framework, lead-lag effects naturally arise as a consequence of the mutual influence between adjustment processes of different assets.

Let us assume, without loss of generality, $\tau = 1$ and re-write eq. (4.1.4) and (4.1.5) as:

$$Y_t = X_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, H) \tag{4.1.6}$$

$$\Delta X_{t+1} = F\Delta X_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q) \tag{4.1.7}$$

where $F = I_d - \Psi$ and $Q = \Psi\Sigma\Psi'$. eq. (4.1.6) is a measurement equation expressing the fact that observations of latent prices are affected by noise while eq. (4.1.7) is a transition equation describing the dynamics of latent returns. We name model (4.1.6), (4.1.7) Multi-Asset Lagged Adjustment (MLA) model. The MLA cannot be estimated as a standard VAR model since X_t is not observed. As discussed in Section 4.1.2, it can conveniently be estimated through a standard Kalman-EM algorithm with missing observations.

The assumption of a constant instantaneous matrix Σ in the efficient log-price process may be regarded as too restrictive, since there is well-established evidence that both volatilities and correlations exhibit strong intraday variation (see e.g Andersen and Bollerslev, 1997, Tsay, 2005, Bibinger et al., 2014, Buccheri et al., 2017b). However, by performing extensive Monte-Carlo simulations and using a misspecified DGP with a time-varying covariance matrix Σ_t, we will show (see Section 4.2.2) two relevant properties. First, the estimate $\hat{F}$ of the VAR(1) matrix of lead-lag coefficients remains unbiased even in presence of time-varying covariances. Second, denoting by $\hat{\Psi}$ and $\hat{Q}$ the two estimates of Ψ and Q, the matrix $\hat{\Sigma} = \hat{\Psi}^{-1}\hat{Q}\hat{\Psi}'^{-1}$ is an unbiased estimator of $\frac{1}{T}QV$, the quadratic covariation of the efficient log-price process:

$$QV = \int_0^T \Sigma_s ds \tag{4.1.8}$$

This result is similar to the one obtained by Shephard and Xiu (2016), who derived the asymptotic theory for the QML estimator of the integrated covariance of a Brownian semimartingale process observed under noise and asynchronicity but neglecting lead-lag effects. Thus, the MLA provides an estimator of the quadratic covariation of a Brownian semimartingale process that is robust to microstructure noise, asynchronicity and that takes into account the presence of lead-lag effects.

Note that making inference on model (4.1.7) can be regarded as testing for one-lag Granger causality in the latent process X_t. Testing for higher order lags would result in additional $d \times d$ lead-lag matrices to be estimated, thus increasing considerably the dimensionality of the parameter space. In practice, a less efficient but more feasible method is to consider observations sampled at a smaller frequency, which avoids estimating complex higher order VAR(p) models.

4.1.2 Estimation

The MLA can be conveniently estimated by writing the two equations in a linear Gaussian state-space representation. This is possible if one introduces the $2d$-dimensional state vector $\bar{X}_t = (X_t', X_{t-1}')'$ and re-writes the two equations as:

$$Y_t = M\bar{X}_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, H) \tag{4.1.9}$$

$$\bar{X}_t = \phi\bar{X}_{t-1} + \bar{\eta}_t, \quad \bar{\eta}_t \sim \text{NID}(0, \bar{Q}) \tag{4.1.10}$$

where:

$$\phi = \begin{pmatrix} I_d + F & -F \\ I_d & 0_{d\times d} \end{pmatrix}, \quad \bar{Q} = \begin{pmatrix} Q & 0_{d\times d} \\ 0_{d\times d} & 0_{d\times d} \end{pmatrix} \tag{4.1.11}$$

with $M = (I_d, 0_{d\times d})$ being a matrix that selects the first d components of $\bar{X}_t$ and $0_{d\times d}$ denoting a $d \times d$ matrix of zeros. We generically denote as Ω the set of parameters to be estimated, namely $\Omega = \{F, Q, H\}$.

Model (4.1.9), (4.1.10) is a linear Gaussian state-space representation for which the Kalman filter can be applied and the log-likelihood function can be written down in the form of the prediction error decomposition, as described in Section 2.2. Instead of numerically optimizing the log-likelihood function through a quasi-Newton method, we use the EM algorithm of A. P. Dempster (1977). The latter is particular advantageous for multivariate models, as it does not require inverting the Hessian matrix.

As a first step, we assume there are no missing observations. We will show how to handle missing observations in the next paragraph. We denote by $\mathcal{X}_n = \{\bar{X}_0, \ldots, \bar{X}_n\}$ the set of latent prices and by $\mathcal{Y}_n = \{Y_1, \ldots, Y_n\}$ the set of observed prices. Also, let us assume that $\bar{X}_0 \sim \text{N}(\mu, \Sigma)$. Note that, since the knowledge of $\bar{X}_{t-1}$ completely determines the last d components of $\bar{X}_t$, the density function $f(\bar{X}_t|\bar{X}_{t-1})$ can be written as:

$$f(\bar{X}_t|\bar{X}_{t-1}) = f(M\bar{X}_t|\bar{X}_{t-1}) \tag{4.1.12}$$

Therefore, denoting by $\log L = \log L(\mathcal{Y}_n, \mathcal{X}_n)$ the complete log-likelihood function, we have:

$$\begin{aligned} \log L = \text{const} &- \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\bar{X}_0 - \mu)'\Sigma^{-1}(\bar{X}_0 - \mu) \\ &- \frac{n}{2}\log|Q| - \frac{1}{2}\sum_{t=1}^{n}(\bar{X}_t - \phi\bar{X}_{t-1})'M'Q^{-1}M(\bar{X}_t - \phi\bar{X}_{t-1}) \\ &- \frac{n}{2}\log|H| - \frac{1}{2}\sum_{t=1}^{n}(Y_t - M\bar{X}_t)'H^{-1}(Y_t - M\bar{X}_t) \end{aligned} \tag{4.1.13}$$

One cannot maximize the complete log-likelihood to obtain the MLE of Ω since $\mathcal{X}_n$ is not observed. The EM algorithm provides an iterative method for finding the MLE by successively maximizing the conditional expectation of the complete log-likelihood function. The latter can be computed using the Kalman filter and smoothing recursions.

Let us introduce the following quantities which can be recovered as an output of the Kalman filter and smoothing recursions:

$$\bar{X}_t^s = \mathrm{E}[\bar{X}_t|\mathcal{Y}_s] \tag{4.1.14}$$

$$\bar{P}_t^s = \mathrm{Cov}[\bar{X}_t|\mathcal{Y}_s] \tag{4.1.15}$$

$$\bar{P}_{t,t-1}^s = \mathrm{Cov}[\bar{X}_t, \bar{X}_{t-1}|\mathcal{Y}_s] \tag{4.1.16}$$

With $s = t$, $s < t$ and $s > t$, the resulting conditional expectation is, respectively, an update filter, a predictive filter and a smoother. $\bar{X}_t^s$ and $\bar{P}_t^s$ can be computed through the Kalman filter and smoother recursions showed in Section 2.2 while $\bar{P}_{t,t-1}^s$ can be computed recursively as described in appendix (A.4). The Kalman filter is initialized with diffuse initial conditions, i.e. we set $\mathbb{E}[\bar{X}_1|Y_1] = 0$ and $\mathrm{Cov}[\bar{X}_1|Y_1] = \kappa I_d$ with $\kappa \to \infty$. At iteration r, the expectation step in the EM algorithm consists in taking the conditional expectation of the complete log-likelihood given the observations $\mathcal{Y}_n$ and using the estimate of Ω obtained at step $r-1$:

$$\begin{aligned}\mathrm{E}[\log L|\mathcal{Y}_n, \hat{\Omega}_{r-1}] = &-\frac{1}{2}\log|\Sigma| - \frac{1}{2}\mathrm{Tr}[\Sigma^{-1}[(\bar{X}_0^n - \mu)(\bar{X}_0^n - \mu)' + \bar{P}_0^n] \\ &- \frac{n}{2}\log|Q| - \frac{1}{2}\mathrm{Tr}[M'Q^{-1}M(C - B\phi' - \phi B' + \phi A\phi')] \\ &- \frac{n}{2}\log|H| - \frac{1}{2}\mathrm{Tr}[H^{-1}\sum_{t=1}^{n}[(Y_t - M\bar{X}_t^n)(Y_t - M\bar{X}_t^n)' + M\bar{P}_t^n M']\end{aligned} \tag{4.1.17}$$

where A, B and C are given by:

$$A = \sum_{t=1}^{n}(\bar{P}_{t-1}^n + \bar{X}_{t-1}^n \bar{X}_{t-1}^{n\prime}) \tag{4.1.18}$$

$$B = \sum_{t=1}^{n}(\bar{P}_{t,t-1}^n + \bar{X}_t^n \bar{X}_{t-1}^{n\prime}) \tag{4.1.19}$$

$$C = \sum_{t=1}^{n}(\bar{P}_t^n + \bar{X}_t^n \bar{X}_t^{n\prime}). \tag{4.1.20}$$

In the maximization step, the function $Q(\Omega|\hat{\Omega}_{r-1}) = \mathrm{E}[\log L|\mathcal{Y}_n, \hat{\Omega}_{r-1}]$ is maximized with respect to Ω. Let us consider the following terms depending on F, Q and H:

$$\begin{aligned}G_1(F,Q) &= -\frac{1}{2}\mathrm{Tr}[M'Q^{-1}M(C - B\phi' - \phi B' + \phi A\phi')] \\ G_2(F,Q) &= -\frac{n}{2}\log|Q| + G_1(F,Q) \\ G_3(H) &= -\frac{n}{2}\log|H| - \frac{1}{2}\mathrm{Tr}[H^{-1}[(Y_t - P\bar{X}_t)(Y_t - P\bar{X}_t)' + M\bar{P}_t^n M']\end{aligned}$$

We start by solving the first order condition $\nabla_F G_1(F,Q)=0$. Let us write the matrices A and B in the following form:

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},\quad B=\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix} \tag{4.1.21}$$

where A_{ij} and B_{ij}, $i=1,2$ are $d\times d$ submatrices of A and B. In appendix (A.5) we prove the following:

Proposition 2. *The solution of the matrix equation $\nabla_F G_1(F,Q)=0$ is:*

$$\hat{F}_r=\Gamma\Theta^{-1} \tag{4.1.22}$$

where $\Gamma=B_{11}-B_{12}-A_{11}+A_{12}$ and $\Theta=A_{11}+A_{22}-A_{12}-A_{21}$.

The estimated value of F is then used to solve $\nabla_Q G_2(\hat{F}_r,Q)=0$. In appendix (A.6), (A.4) we prove the following proposition:

Proposition 3. *The solution of the two matrix equations $\nabla_Q G_2(\hat{F}_r,Q)=0$, $\nabla_H G_3(H)=0$ is:*

$$\hat{Q}_r=\frac{\hat{\Upsilon}}{n},\quad \hat{H}_r=\frac{diag(\Lambda)}{n} \tag{4.1.23}$$

where $\hat{\Upsilon}=M(C-B\hat{\phi}_r'-\hat{\phi}_r B'+\hat{\phi}_r A\hat{\phi}_r')M'$, $\Lambda=\sum_{t=1}^n[(Y_t-M\bar{X}_t^n)(Y_t-M\bar{X}_t^n)'+M\bar{P}_t^n M']$ and

$$\hat{\phi}_r=\begin{pmatrix}I_d+\hat{F}_r & -\hat{F}_r\\ I_d & 0_{d\times d}\end{pmatrix} \tag{4.1.24}$$

Conditions under which the EM algorithm converges to a local maximum of the incomplete log-likelihood function are studied by Wu (1983). We check convergence by looking at the relative increase of the log-likelihood and stop the algorithm when it is lower than some small threshold ($\mu=10^{-6}$ in our simulation and empirical study). The log-likelihood can be computed in the prediction error decomposition form:

$$\log L=\text{const}-\frac{1}{2}\sum_{t=1}^n\log|F_t|-\frac{1}{2}\sum_{t=1}^n v_t'F_t^{-1}v_t \tag{4.1.25}$$

where $v_t=Y_t-M\bar{X}_t^{t-1}$ is the prediction error and $F_t=M\bar{P}_t^{t-1}M'+H$.

Once $\hat{F}$, $\hat{Q}$ and $\hat{H}$ have been estimated, the matrix of price adjustment Ψ and the covariance matrix of the efficient log-price process Σ can be computed as:

$$\hat{\Psi}=I_d-\hat{F},\quad \hat{\Sigma}=\hat{\Psi}^{-1}\hat{Q}\hat{\Psi}'^{-1} \tag{4.1.26}$$

The Kalman filter and smoothing recursions provide filtered and smoothed estimates of the lagged price X_t. From these, using eq. (4.1.3), one also obtains, as a byproduct, filtered and smoothed estimates of the martingale efficient log-price process.

4.1.3 Missing value modification

The update formulas in the maximization step can be modified to keep into account missing values. Let us assume that, at time t, d_1 components in the vector Y_t are observed while the remaining d_2 are not observed. We consider the d_1-dimensional vector $Y_t^{(1)}$ of observed components and the $d_1 \times d$ matrix $M_t^{(1)}$ whose lines are the lines of M corresponding to $Y_t^{(1)}$. Also, we consider the $d_1 \times d_1$ covariance matrix $H_t^{(11)}$ of observed components disturbances. Following Shumway and Stoffer (2015), the Kalman filter and smoothing recursions in appendix (A.4) and the prediction error decomposition form of the log-likelihood, eq. (4.1.25) are still valid, provided that one replaces Y_t, M and H with:

$$Y_{(t)} = \begin{pmatrix} Y_t^{(1)} \\ \underline{0} \end{pmatrix}, \quad M_{(t)} = \begin{pmatrix} M_t^{(1)} \\ \underline{0} \end{pmatrix}, \quad H_{(t)} = \begin{pmatrix} H_t^{(11)} & \underline{0} \\ \underline{0} & \mathcal{I}^{(22)} \end{pmatrix} \tag{4.1.27}$$

where $\mathcal{I}^{(22)}$ is the $d_2 \times d_2$ identity matrix and $\underline{0}$ generically denotes zero arrays of appropriate dimension. Note that the time dependence in $M_{(t)}$ and $H_{(t)}$ is only due to missing observations, while the matrices M and H are constant over time.

Taking the conditional expectation in eq. (4.1.13) requires some modifications in case of missing observations. The second and the fourth term remain as in eq. (4.1.17), provided that one runs Kalman filter and smoothing recursions as described in (4.1.27). The last term changes because one needs to evaluate expectations of Y_t conditioning to the incomplete data $\mathcal{Y}_n^{(1)} = \{Y_1^{(1)}, Y_2^{(1)}, \dots, Y_n^{(1)}\}$. If H is diagonal, as we are assuming here, Shumway and Stoffer (1982) showed that:

$$\begin{aligned} \mathbb{E}[(Y_t - M\bar{X}_t)(Y_t - M\bar{X}_t)'|\mathcal{Y}_n^{(1)}] &= (Y_{(t)} - M_{(t)}\bar{X}_t^n)(Y_{(t)} - M_{(t)}\bar{X}_t^n)' \\ &\quad + M_{(t)}\bar{P}_t^n M_{(t)}' + \begin{pmatrix} \underline{0} & \underline{0} \\ \underline{0} & \hat{H}_{22,t,r-1} \end{pmatrix} \end{aligned} \tag{4.1.28}$$

where $\hat{H}_{22,r-1}$ is the $d_2 \times d_2$ covariance matrix of unobserved components disturbances at time t obtained using the estimate at step $r-1$ of the matrix H. Therefore, the update equation for H becomes:

$$\hat{H} = \frac{\text{diag}(\Lambda^*)}{n} \tag{4.1.29}$$

where

$$\Lambda^* = \sum_{t=1}^{n} D_t \left[(Y_{(t)} - M\bar{X}_t^n)(Y_{(t)} - M\bar{X}_t^n)' + M_{(t)}\bar{P}_t^n M_{(t)}' + \begin{pmatrix} \underline{0} & \underline{0} \\ \underline{0} & \hat{H}_{22,t,r-1} \end{pmatrix} \right] D_t', \tag{4.1.30}$$

D_t being a permutation matrix that rearranges the components of Y_t in their original order.

4.2 Simulation study

4.2.1 Comparison with other estimators

As underlined in the introduction, non-synchronous trading can have deep consequences when one wants to make inference on multivariate high-frequency tick-by-tick data. For instance, the Epps effect is

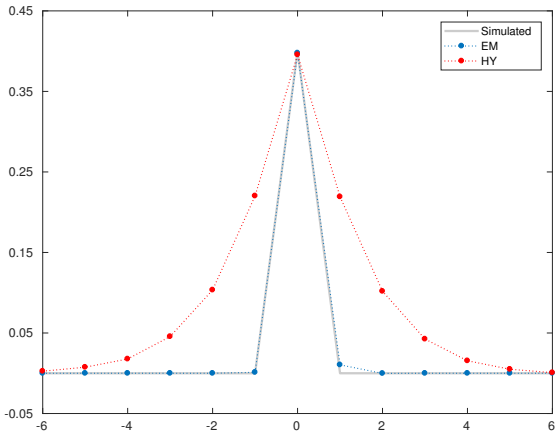


Figure 4.2.1: Cross-correlogram of two simulated Brownian motions with correlation $\rho = 0.4$ observed asynchronously over $T = 10000$ timestamps. The red line is obtained by averaging across HY estimates while the blue line is obtained by averaging EM estimates over $N = 250$ independent realizations.

certainly due to asynchronicity (see Epps, 1979 and Hayashi and Yoshida, 2005).

Non-synchronous trading is responsible for two main kinds of spurious lead-lag correlations. First, some assets seem to lead other assets simply because, being traded more frequently, they are more likely to show the effect of new information arriving on the market before other assets which are traded less frequently. This effect is only due to differences in the level of trading activity and is not related to true lead-lag dependencies. Second, as we will show below, even in presence of similar levels of trading activities, one can find spurious nonzero lead-lag correlations that are not related to true lead-lag dependencies. Another source of spurious lead-lag correlations can simply arise as a result of the combination between autocorrelation and contemporaneous correlations and is not due to nonzero non-diagonal elements in the VAR matrix.

A detailed analysis of the impact of asynchronicity on lead-lag correlations has been performed by Huth and Abergel (2014), who considered the standard previous-tick correlation estimator (Griffin and Oomen, 2011) and the estimator proposed by Hoffmann et al. (2013). The latter is computed on bivariate series by applying the Hayashi-Yoshida (HY) estimator (Hayashi and Yoshida 2005) after shifting the timestamps of one of the two series. In their simulation study, Huth and Abergel generated two (contemporaneously) correlated Brownian motions with different timestamps. Compared to the previous-tick correlation estimator, the HY estimator is not affected by differences in the levels of trading activity, meaning that the lead-lag correlogram remains symmetric even if the two processes are characterized by different average durations. However, as a consequence of asynchronicity, it has a bias at nonzero leads and lags that implies nonzero correlations even in absence of true lead-lag dependencies.

The MLA is robust to both effects. The reason is that asynchronicity is handled as resulting from

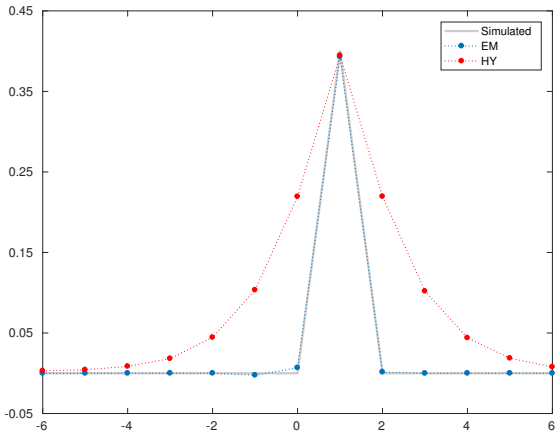


Figure 4.2.2: As in figure (4.2.1) but HY and EM are estimated over a time series obtained by shifting one of the two simulated Brownian motions. We show the average correlations provided by the HY (red line) and EM (blue line) estimators and the cross-correlogram of the simulated time-series (grey line).

missing observations which can easily be incorporated in the EM algorithm without jeopardizing the inference. In order to show this property, we sample two Brownian motions over a time grid of $T = 10000$ equally spaced points. The correlation between the two Brownian motions is $\rho = 0.4$. Asynchronicity is reproduced by censoring the simulated observations through Poisson sample. The probability of missing values is set equal to $\Lambda_1 = 0.3$ for the first series and $\Lambda_2 = 0.5$ for the second series. We repeat the experiment 250 times and for each realization compute lead-lag correlations by estimating model (4.1.6), (4.1.7) and using the HY estimator. Figure (4.2.1) shows the cross-correlogram obtained by averaging lead-lag correlations over all the simulations. We note that both correlograms are symmetric, meaning that both estimators are not affected by differences in the level of trading activity. However, the HY estimator provides nonzero correlations at nonzero lags. As shown by Huth and Abergel (2014) (appendix B) this is due to asynchronicity. In contrast, the EM estimator is not affected by asynchronicity and correctly reproduces the correlogram of simulated data.

Our definition of lead-lag effects is formally different from the one of Hoffmann et al. In the MLA framework, lead-lag effects arise as a consequence of nonzero non-diagonal coefficients in the lagged adjustment matrix Ψ. In their work, Hoffman et al. considered a continuous-time bivariate process (X_t, Y_t) and focused on the estimation of the time shift θ such that the shifted process $(X_t, Y_{t+\theta})$ is a semi-martingale with respect to some filtration. In order to understand how the MLA behaves in this different framework, we consider the bivariate time series of the previous experiment and shift by a lag $\theta = 1$ all the timestamps of one of the two series. Figure (4.2.2) shows the correlogram of the new bivariate time-series and those estimated by the HY and MLA estimators. The HY estimator correctly estimates the lagged cross correlation but provides nonzero correlations at other leads and lags. The

MLA estimator correctly captures the true cross-correlogram. Indeed, the shifted time series can be written as a VAR(1) process with a nonzero non-diagonal element in F and uncorrelated disturbances. In case $\theta > 1$, one can sample observations at a lower frequency and still use the MLA estimator.

As a final remark, it is interesting observing that, in contrast to the HY estimator, the MLA estimator is robust to microstructure noise in that it allows observations of the underlying process X_t to be contaminated by noise. Another relevant advantage is that the proposed estimator is not pairwise, i.e. it can be applied to a generic multivariate time-series of dimension $d \geq 2$.

4.2.2 Robustness to stochastic volatility

In eq. (4.1.1) the covariance matrix Σ of the efficient log-price is assumed to be constant over time. This assumption is too restrictive since real high-frequency data are characterized by significant changes in their covariance structure during the day. In order to assess the properties of $\hat{F}$ and $\hat{\Sigma}$ in a more realistic scenario, we simulate realizations from a misspecified DGP with a time-varying covariance matrix Σ_t. The latter is decomposed as:

$$\Sigma_t = D_t R D_t \tag{4.2.1}$$

where R is a constant correlation matrix and D_t is a diagonal matrix of time-varying standard deviations:

$$D_t = \begin{pmatrix} \sigma_{t,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{t,d} \end{pmatrix} \tag{4.2.2}$$

The dynamic terms $\sigma_{t,i}$ evolve through the following stochastic volatility model:

$$d\sigma_{t,i}^2 = k_i(v_i - \sigma_{t,i}^2)dt + w_i \sigma_{t,i} dB_{t,i} \quad i = 1, \ldots, d \tag{4.2.3}$$

where $B_{t,i}$ is Wiener process, $\mathbb{E}[dB_{t,i}dW_{t,j}] = \delta_{ij}\rho_{ij}dt$ and $\mathbb{E}[dB_{t,i}dB_{t,j}] = \delta_{ij}dt$.

We report the results obtained in the bivariate case ($d = 2$). The length of the trading day is assumed to be $T = 6.5$ hours. Thus, we simulate $n = 23400$ timestamps of 1-second prices. The stochastic volatility process is simulated by through Euler discretization. The parameters in eq. (4.2.3) are chosen in the following way: $v_1 = 0.01$, $v_2 = 0.02$, $w_1 = w_2 = 0.1$, $k_1 = 10$, $k_2 = 7$. We adopt the following choices for the matrix F of lead-lag coefficients and the leverage matrix ρ:

$$F = \begin{pmatrix} 0.1 & 0.5 \\ 0.3 & 0.1 \end{pmatrix}, \quad \rho = \begin{pmatrix} 0.05 & 0 \\ 0 & 0.1 \end{pmatrix} \tag{4.2.4}$$

The diagonal elements of the variance matrix H, that we denote as h_{ii}, $i = 1, 2$, are computed based on the average signal-to-noise ratio, that is defined as $\bar{\delta}_i = v_i / h_{ii}$.

	avg×100	stDev×100	p-value	avg×100	stDev×100	p-value	avg×100	stDev×100	p-value
	$\delta = 0.5$			$\delta = 1$			$\delta = 2$		
	$\Lambda = 0$								
F_{11}	0.1950	2.6117	**0.0184**	0.0725	3.0954	**0.4589**	0.4159	3.3105	**0.5314**
F_{12}	−0.2393	2.5798	0.0034	−0.0857	3.2803	**0.4090**	−0.0616	3.4449	**0.5717**
F_{21}	0.0812	2.2679	**0.2580**	0.0191	2.8767	**0.8341**	0.0002	3.2539	**0.9983**
F_{22}	−0.1742	2.9171	**0.0592**	−0.1233	3.6635	**0.2876**	0.0067	3.8570	**0.9563**
Σ_{11}	−0.0053	0.0879	**0.0588**	−0.0019	0.0861	**0.4937**	−0.0045	0.0873	**0.1043**
Σ_{12}	−0.0043	0.0416	0.0011	−0.0057	0.0500	0.0040	−0.0063	0.0563	**0.0521**
Σ_{22}	0.0083	0.1348	**0.0510**	−0.0038	0.1462	**0.4159**	−0.0043	0.1285	**0.2870**
	$\Lambda = 0.5$								
F_{11}	−0.1778	3.3155	**0.0903**	0.2465	3.7535	**0.0381**	0.1386	3.2338	**0.1756**
F_{12}	0.0212	3.4834	**0.8476**	−0.3053	3.8869	**0.0132**	−0.2190	3.3824	**0.0409**
F_{21}	−0.9099	5.2234	0.0002	−0.4047	3.9738	0.0075	−0.2804	3.3068	**0.0105**
F_{22}	0.8688	5.7858	0.0002	0.1713	4.7535	**0.2548**	0.1942	3.9592	**0.1211**
Σ_{11}	0.0068	0.0935	**0.0209**	−0.0014	0.0923	**0.6268**	0.0013	0.0949	**0.6696**
Σ_{12}	0.0028	0.0671	**0.1831**	0.0011	0.0650	**0.5876**	−0.0004	0.0550	**0.8036**
Σ_{22}	−0.0064	0.1473	**0.1723**	0.0030	0.1584	**0.5510**	9.81e-6	0.1471	**0.9983**

Table 4.2.1: We report the sample averages and standard deviations of all the elements of the pivotal matrices $\hat{\theta}_F$, $\hat{\theta}_\Sigma$ and the p-value of the one-sample t-test in all the simulated scenarios. The cases in which the null hypothesis is not rejected at the 1% c.l. are denoted by bold numbers.

For simplicity, we assume $\bar{\delta}_1 = \bar{\delta}_2$. In order to mimic realistic noise scenarios, we choose $\bar{\delta}_i = 0.5, 1, 2$. Indeed, as shown in Section 4.3, these are the values that are estimated on real markets. We consider both the case where observations are synchronized and the case where there are missing values. In the latter case, the simulated observations are censored using Poisson sampling. The probability of having a missing value is set equal to $\Lambda = 0.5$ for both series.

We estimate model (4.1.6), (4.1.7) for $N = 1000$ independent realizations and consider the pivotal statistics $\hat{\theta}^i_F = \hat{F}^i - F$ and $\hat{\theta}^i_\Sigma = \hat{\Sigma}^i - \frac{1}{T}QV$, for $i = 1, \ldots, N$. The term QV is the quadratic covariation defined in eq. (4.1.8). The three scenarios $\bar{\delta} = 0.5, 1, 2$ are combined to each of the two scenarios $\Lambda = 0, 0.5$, obtaining a total of 6 scenarios. In table (4.2.1) we show the sample mean and standard deviations of each entries of two matrices $\hat{\theta}^i_\Sigma$ and $\hat{\theta}^i_F$. A one sample t-test is performed in order to test the null assumption that the mean is zero. The p-value is reported in the table. The distribution of $\hat{\theta}_\Sigma$ and $\hat{\theta}_F$ is always centered in 0. This implies that, even in case of time-varying covariances, $\hat{F}$ is an unbiased estimate of the true matrix F of lead-lag coefficients while $\hat{\Sigma}$ is an unbiased estimate of the quadratic covariation of the efficient log-price process.

In figures (4.2.3), (4.2.4), (4.2.5), the two scenarios $\Lambda = 0, \bar{\delta} = 1$ and $\Lambda = 0.5, \bar{\delta} = 1$ are considered. We plot histograms of each element of $\hat{\theta}_\Sigma$ and $\hat{\theta}_F$ after normalizing by their sample standard deviations. In the first scenario there are no missing values and the histograms are perfectly compatible with a standard normal distribution. In the second scenario the distribution is still centered in zero but we observe slight deviations from the normal. This is due to the fact that censoring halves, on average, the number of observations and thus leads to less efficient estimates.

In order to assess the effect of lead-lag correlations, we plot in figure (4.2.6) the same histograms as in figure (4.2.5a), (4.2.5b), (4.2.5c) but now use the Hayashi-Yoshida estimator to compute the quadratic covariation. As can be seen, the Hayashi-Yoshida estimator is largely biased in case lead-lag effects are present. A similar bias is observed when using other estimators of the quadratic covariation, e.g. the pairwise estimator of Aït-Sahalia et al. (2010), the multivariate realized kernel of Barndorff-Nielsen et al. (2011), and the QML estimator of Shephard and Xiu (2016).

4.3 Empirical evidence

4.3.1 Dataset

Our dataset is provided by Thomson Reuters and contains intraday transaction data of 11 among the most frequently traded NYSE assets in the period between 03-01-2006 and 31-12-2014, a total of 2250 days. The time stamp precision is the second. We use the procedure described by Barndorff-Nielsen et al. (2009) to clean the data. In particular: (i) we consider trades in the time window from 9:30 to

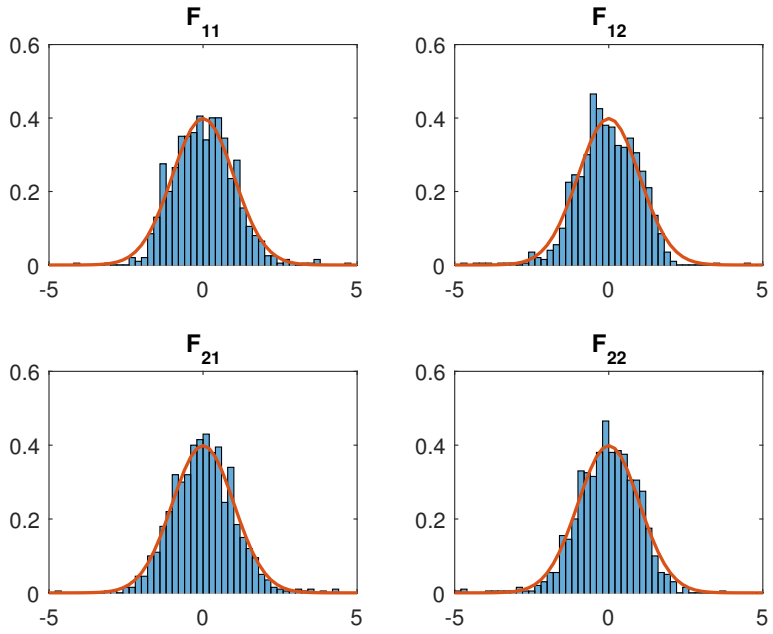


Figure 4.2.3: Histograms of the elements of the matrix $\hat{\theta}_F$ standardized by their sample standard deviations in the scenario $\bar{\delta} = 1, \Lambda = 0$ over $N = 1000$ independent realizations. The red line is the standard normal distribution.

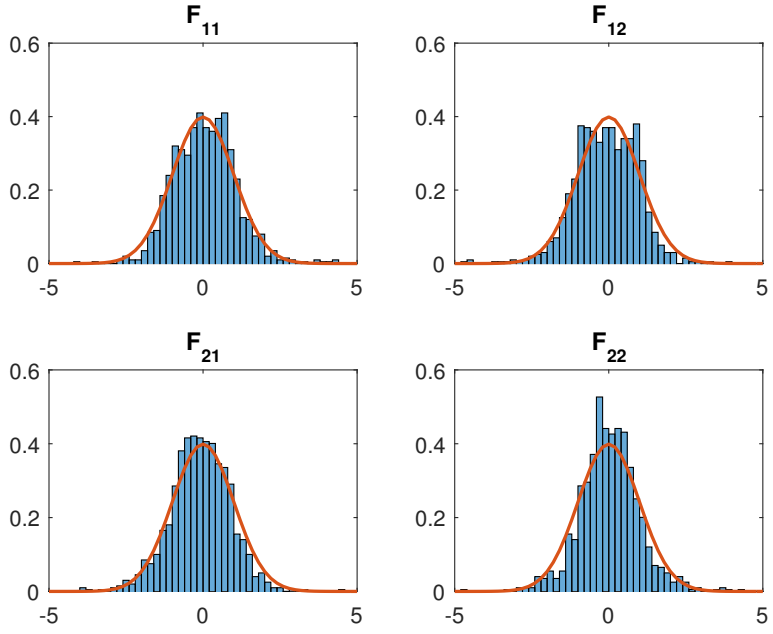


Figure 4.2.4: Histograms of the elements of the matrix $\hat{\theta}_F$ standardized by their sample standard deviations in the scenario $\bar{\delta} = 1, \Lambda = 0.5$ over $N = 1000$ independent realizations. The red line is the standard normal distribution.

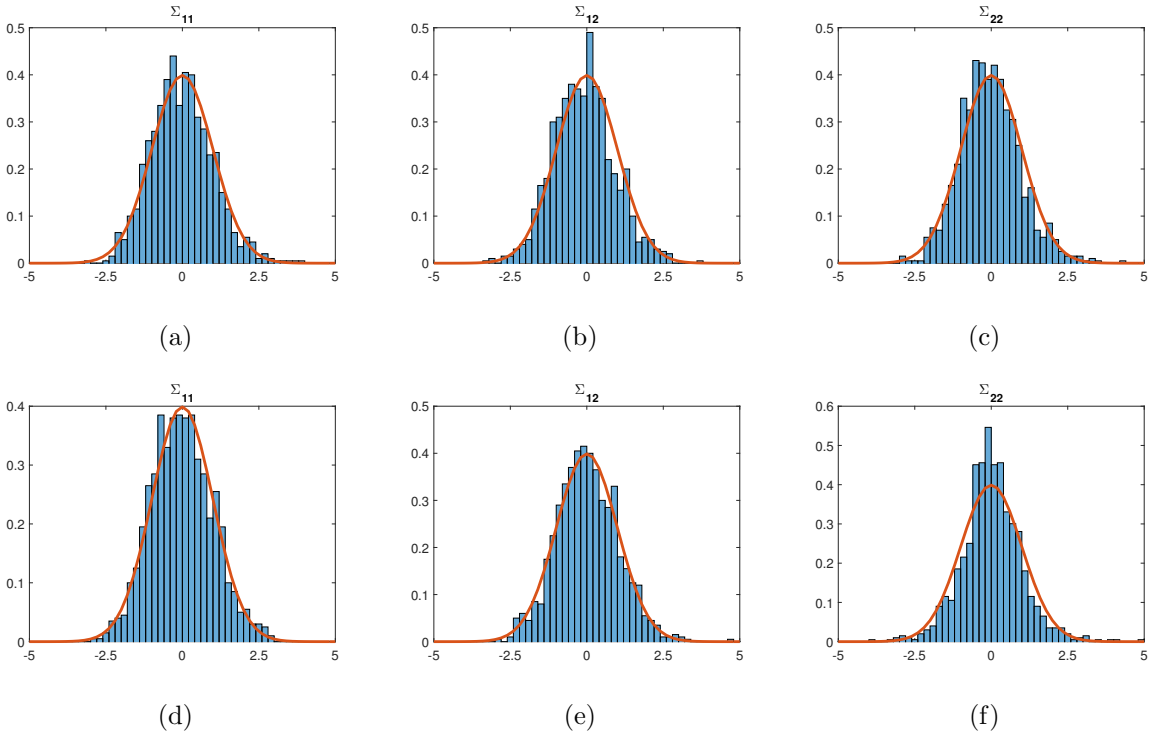


Figure 4.2.5: Histograms of the elements of the matrix $\hat{\theta}_\Sigma$ standardized by their sample standard deviations in the scenario $\bar{\delta} = 1, \Lambda = 0$ (a), (b), (c) and in the scenario $\bar{\delta} = 1, \Lambda = 0.5$ (d), (e), (f) over $N = 1000$ independent realizations. The red line is the standard normal distribution.

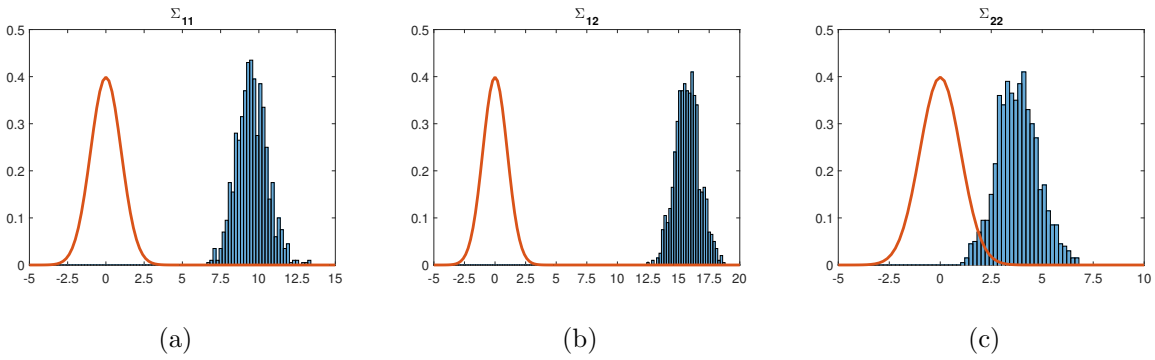


Figure 4.2.6: Histograms of the elements of the matrix $\hat{\theta}_\Sigma$ standardized by their sample standard deviations and computed using the Hayashi-Yoshida estimator in the scenario $\bar{\delta} = 1, \Lambda = 0$ over $N = 1000$ independent realizations. The red line is the standard normal distribution.

16:00; (*ii*) we aggregate high-frequency prices at the one-second frequency by taking the median price of trades within the same second; (*iii*) we delete entries with prices that are above the ask plus the bid-ask spread.

Table (4.3.1) shows the 11 assets used for the analysis together with the probability of missing values Λ, the average duration $\overline{\Delta t}$ in seconds between observations, the average number of observations per day $\bar{n}$ and the average signal-to-noise ratio $\bar{\delta}$ as estimated by the MLA in the last 250 days of the sample. Note that five of the selected assets, namely {C, JPM, BAC, MS, GS} belong to the banking sector while the remaining six, namely {XOM, CVX, SLB, GM, COP, GE} belong to the oil, energy and transport sectors. We name the two sets of assets as "Group I" and "Group II", respectively.

Stock	Symbol	Λ	$\overline{\Delta t}$	$\overline{n}$	$\overline{\delta}$
Exxon	XOM	0.816	5.434	4304	1.178
Citigroup	C	0.837	6.135	3832	1.246
JPMorgan	JPM	0.840	6.250	3743	0.999
Chevron	CVX	0.848	6.578	3553	0.850
Schlumberger	SLB	0.853	6.802	3454	0.613
General Motors	GM	0.866	7.462	3135	0.888
Bank of America	BAC	0.869	7.633	3079	0.328
ConocoPhillips	COP	0.880	8.333	2828	0.494
General Electric	GE	0.892	9.259	2543	0.641
Morgan Stanley	MS	0.897	9.708	2416	0.741
Goldman Sachs	GS	0.920	12.500	1873	0.630

Table 4.3.1: For each asset we show the probability of missing values Λ, the average duration $\overline{\Delta t}$ in seconds between consecutive observations, the average number of observations per day $\bar{n}$ and the average signal-to-noise ratio as estimated by the MLA. The averages are computed over the last 250 days of the sample.

4.3.2 Lead-lag effects and cross-asset trading

Our primary goal is to assess the statistical significance of the multi-asset lagged adjustment mechanism that we have introduced. To this end, we compare the MLA to a standard random walk plus noise model, also known as local level (LL) model (see e.g. Durbin and Koopman 2012). The latter is nested into the MLA, as it is obtained by setting to zero the VAR matrix F in eq. (4.1.7). The resulting model can be estimated through the EM algorithm with missing observations, as described by Corsi et al. (2015) and Shephard and Xiu (2016).

In order to compare the two models, for the 11 assets considered here, we estimate the LL and the MLA on each day of the sample and compute the likelihood ratio $\lambda = \frac{\mathcal{L}^{\mathrm{LL}}}{\mathcal{L}^{\mathrm{MLA}}}$. Figure (4.3.1) reports the

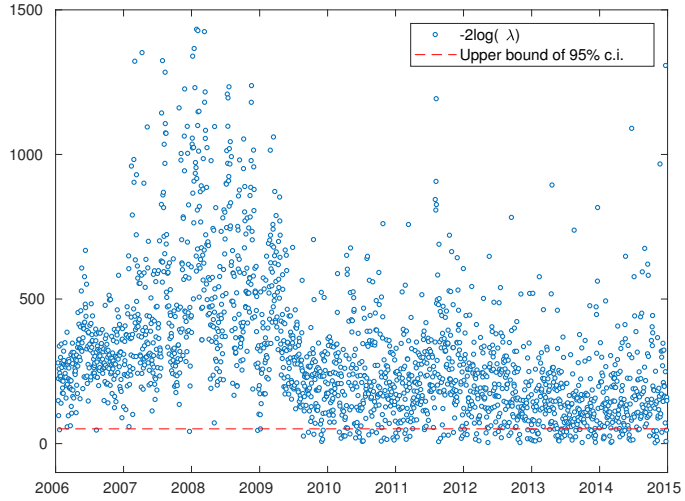


Figure 4.3.1: We show the test statistic $-2\log(\lambda)$ computed for each day of the sample and the upper bound of the 95% confidence interval evaluated using a χ^2 distribution with a number of degrees of freedom equal to d^2.

test statistic $-2\log(\lambda)$, which is distributed according to a χ^2 with a number of degrees of freedom equal to d^2, i.e. the difference between the number of parameters in the MLA and the one in the LL model. The null hypothesis that the matrix of lead-lag coefficients F is zero is strongly rejected in most of the cases. This indicates that the simple LL specification considered e.g. by Corsi et al. (2015) and Shephard and Xiu (2016) is not sufficient to capture important features of the dynamics of high-frequency prices. In particular, the fact that F is nonzero implies that the lagged adjustment matrix $\Psi = I_d - F$ is different from the identity and therefore the adjustment process is delayed, as a result of lagged dissemination of information across stocks.

The deviation of the MLA specification from the null assumption of a simple random walk process is larger in periods of high volatility. Figure (4.3.2) shows the logarithm of RV_{avg}. For each day of the sample, the latter is computed as the average realized variance of the 11 assets considered in table (4.3.1). Large values of $-2\log(\lambda)$ correspond to bursts in volatility. This is confirmed by figure (4.3.3), which shows a scatter plot of the two quantities and the line obtained from the OLS regression. The correlation is $\rho = 0.4296$ and is highly significant.

The relation between cross-asset effects and volatility can be ascribed to the impact of high-frequency trading (HFT). HFT is typically based on short-term statistical dependencies among assets. In periods of high uncertainty, volatility can create short-living cross autocorrelations that are exploited by high-frequency traders. A positive relationship between HFT and volatility, especially in periods of high market uncertainty, was found by Zhang (2010). A related result is the one of Dobrev and Schaumburg

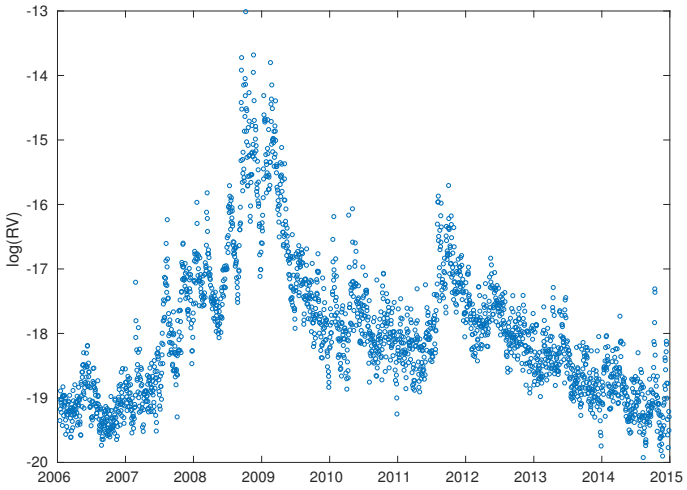


Figure 4.3.2: For each day of the sample, we show the logarithm of RV_{avg}. The latter is computed as the average realized variance of the 11 assets considered in table (4.3.1).

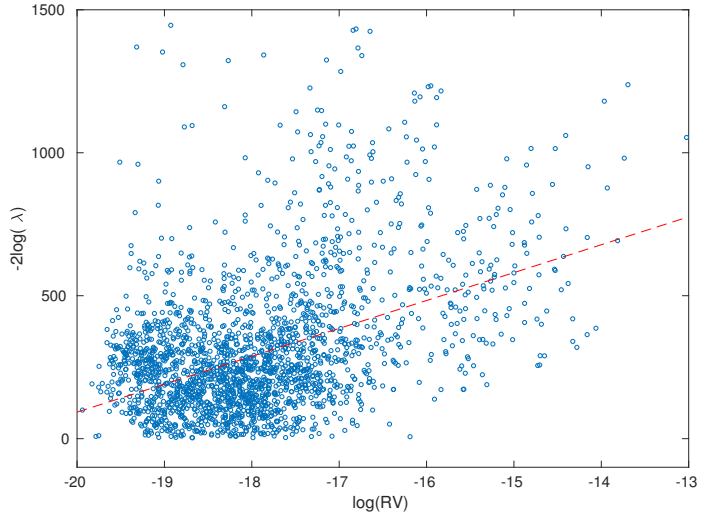


Figure 4.3.3: Scatter plot of $-2\log(\lambda)$ versus $\log(RV_{\text{avg}})$, where RV_{avg} is computed as the average realized variance of the assets. The correlation is $\rho = 0.4296$. We also show the result of the OLS regression of $-2\log(\lambda)$ on $\log(RV_{\text{avg}})$

(2017), who revealed a close relationship between volatility and lead-lag effects among different markets as a consequence of HFT surges in cross-market activity.

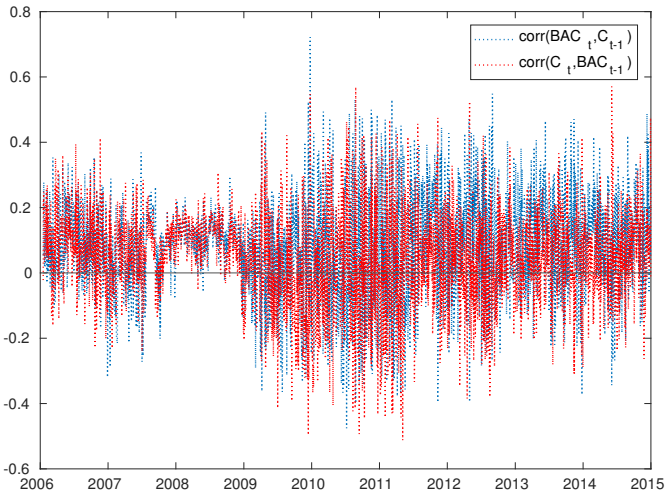


Figure 4.3.4: Dynamics of 1-sec lead-lag correlations between Bank of America and Citigroup.

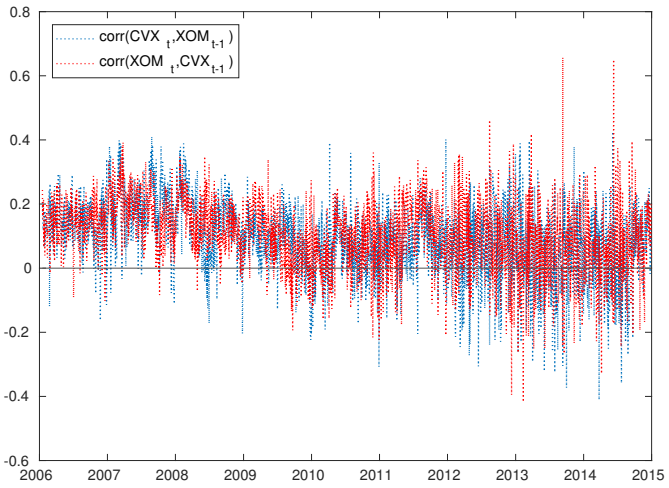


Figure 4.3.5: Dynamics of 1-sec lead-lag correlations between Chevron and Exxon.

A more detailed description of the relation between market volatility and cross-asset effects can be obtained by looking at the estimated lead-lag correlations. Figure (4.3.4) shows the daily dynamics of 1-sec lead-lag correlations between Bank of America and Citigroup. Lead-lag correlations are computed from the MLA estimated parameters as described in appendix (A.7). Figure (4.3.5) shows the same result for Chevron and Exxon. In periods of high uncertainty and large volatility, the cross-autocorrelation structure of the market is more stable, with leag-lag correlations being typically above zero. This is especially true during the 2007-2008 financial crisis, where correlations are generally positive for both

couples of assets. Note also that both lead and lag correlations are nonzero, meaning that sometimes asset A can lead the dynamics of asset B and other times B can lead the dynamics of A.

In periods of low volatility, the dynamic of leag-lag correlations is generally more erratic. Thus, it is necessary to focus on shorter periods of time to examine in more detail the multi-asset structure of the market. We concentrate here on the last 250 days of the sample, i.e. the subsample including the year 2014. In figures (4.3.6), (4.3.7) we plot the average of cross autocorrelations for every couple of assets in Group I and Group II, respectively. Bars denote 95% confidence intervals. Correlations at positive lags imply that the second asset is leading the first while correlations at negative lags imply the opposite. As suggested by the previous analysis, we find strong evidences of cross-asset effects in both groups. For instance, in Group I, Goldman Sachs appears to lead all other assets while Bank of America is lead by all the remaining assets. Note that Goldman Sachs is the less traded asset of Group I. The fact that it leads the dynamics of other assets implies that this effect is not merely due to differences in liquidity but to true lead-lag correlations that emerge as a consequence of nonzero lead-lag coefficients in the matrix F. In Group II we note that oil companies like Exxon, ConocoPhillips, Chevron, Schlumberger lead the dynamics of energy and transport companies like General Electric and General Motors. In contrast, we observe weaker lead-lag correlations among leaders (e.g. between XOM and CVX) and among laggers (e.g. between GE and GM).

There is statistical evidence of lead-lag correlations between assets belonging to different groups. These between-groups effects are in general weaker then within-group lead-lag effects. For instance, in figure (4.3.8) we show the estimated correlogram of GS-GM (4.3.8a) and the one of MS-CVX (4.3.8b). These are the two couples of assets belonging to different groups exhibiting the largest lead-lag correlations. In other cases we observe smaller or even non-significant correlations.

Note that lead-lag correlations can arise even if the non-diagonal elements of F are all zero, as a consequence of combined autocorrelation and contemporaneous correlation effects. This is not the case here, as the estimated matrix $\hat{F}$ has a lot of statistically significant non-diagonal elements. Tables (4.3.2), (4.3.3) show the average, over the subsample considered above, of the two sub-matrices of F corresponding to lead-lag coefficients of Group I and Group II, respectively. Non-diagonal elements are nonzero, with great significance. This implies that the recovered cross-asset structure arises as a direct consequence of the proposed multi-asset price formation mechanism. This result can be interpreted in light of cross-asset trading. Dealers tend to rely on the prices of more informative securities in order to set their quotes and this translates into a lagged dissemination of information across assets, as captured by the non-diagonal elements of the matrix Ψ.

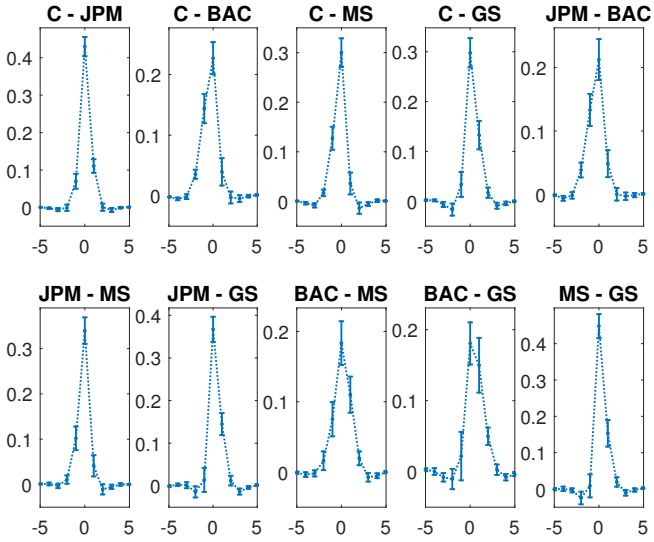


Figure 4.3.6: Cross autocorrelations for all the couples of assets in Group I. The correlograms are computed by averaging those obtained on all the business days of 2014. Error bars denote 95% confidence intervals. Correlations at positive lags imply that the second asset displayed in the title is leading the first and the other way around for negative lags.

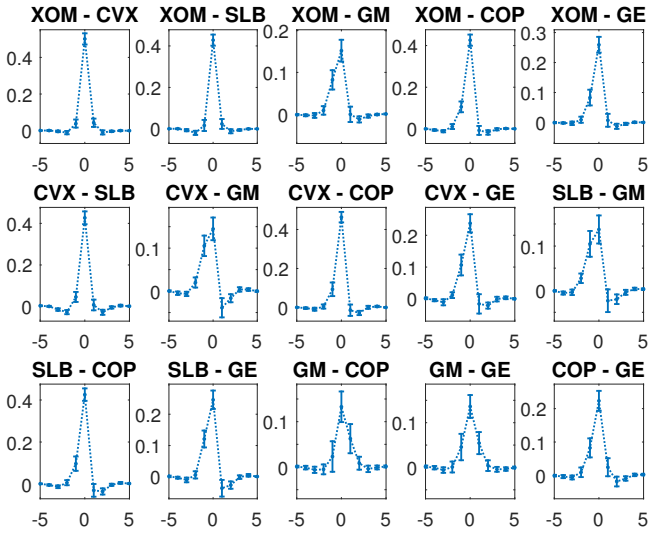


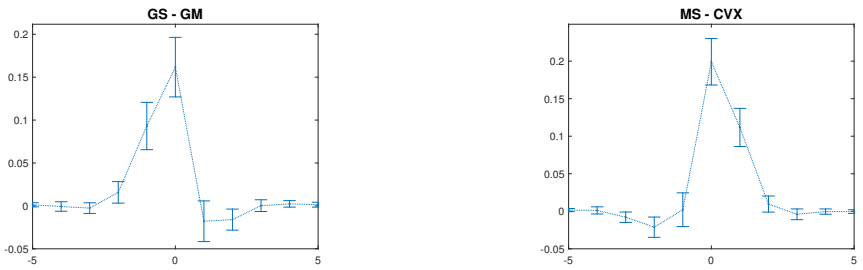
Figure 4.3.7: Cross autocorrelations for all the couples of assets in Group II. The correlograms are computed by averaging those obtained on all the business days of 2014. Error bars denote 95% confidence intervals. Correlations at positive lags imply that the second asset displayed in the title is leading the first and the other way around for negative lags.

	Group I				
	avg F_{ij}				
	C	JPM	BAC	MS	GS
C	0.0886****	0.0472****	0.0220*	-0.0635****	0.1276****
JPM	0.0318***	0.1023****	$0.0065^{(\mathrm{ns})}$	-0.0709****	0.1358****
BAC	0.0518****	0.0657****	0.0863****	$0.0201^{(\mathrm{ns})}$	0.1040****
MS	0.0752****	0.0973****	$0.0107^{(\mathrm{ns})}$	0.0193*	0.1542****
GS	0.0334**	$0.0011^{(\mathrm{ns})}$	$-0.0031^{(\mathrm{ns})}$	-0.0743****	0.1647****

Table 4.3.2: We report the sample average, over the sub-sample of $N = 252$ days, of the elements of the estimated matrices $\hat{F}$ corresponding to assets belonging to Group I, together with significance levels obtained based on the p-value of the one-sample t-test: $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$, $^{(\mathrm{ns})}p > 0.05$.

	Group II					
	avg F_{ij}					
	XOM	CVX	SLB	GM	COP	GE
XOM	0.0776****	0.0428***	0.0273****	-0.0143**	$-0.0263^{(\mathrm{ns})}$	-0.0408****
CVX	0.0232*	0.0981****	0.0155*	$-0.0137^{(\mathrm{ns})}$	-0.0327*	-0.0322***
SLB	$0.0135^{(\mathrm{ns})}$	0.0665***	0.0946****	-0.0251*	-0.0709****	-0.0355*
GM	0.0809****	0.0657**	0.0733****	0.0686****	$-0.0035^{(\mathrm{ns})}$	$0.0182^{(\mathrm{ns})}$
COP	0.0318**	0.0502***	0.0485****	$-0.0114^{(\mathrm{ns})}$	0.0531****	$-0.0045^{(\mathrm{ns})}$
GE	0.0321*	0.0542***	0.0424****	$0.0137^{(\mathrm{ns})}$	$-0.0015^{(\mathrm{ns})}$	0.0585****

Table 4.3.3: We report the sample average, over the sub-sample of $N = 252$ days, of the elements of the estimated matrices $\hat{F}$ corresponding to assets belonging to Group II, together with significance levels obtained based on the p-value of the one-sample t-test: $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$, $^{(\mathrm{ns})}p > 0.05$.



(a) (b)

Figure 4.3.8: Cross autocorrelations between stocks belonging to different groups.

Chapter 5

Intraday covariance dynamics

Almost all results in this chapter previously appeared in Buccheri et al. (2017b).

Multivariate conditional volatilities and correlation models have been largely applied to low-frequency (e.g. daily) financial data where prices are synchronized and microstructure effects are negligible. Popular multivariate dynamic time-series models include the class of multivariate extensions of the univariate GARCH model of Engle (1982) and Bollerslev (1986) and the dynamic conditional correlation (DCC) model of Engle (2002a). However, these models are difficult to apply to intraday high-frequency financial data, since assets are traded asynchronously and prices are contaminated by microstructure effects. At the same time, estimating and forecasting intraday volatilities and correlations of high-frequency asset prices is a problem of crucial interest from both a theoretical and an empirical perspective.

We propose a new multivariate conditional correlation model able to cope with observations of the underlying process that are asynchronous and affected by noise. This is done by considering a multivariate local level model with time-varying covariances and treating asynchronicity as a missing value problem. The dynamic of time-varying parameters is driven by the score of the predictive likelihood (Creal et al. 2013), thus allowing to write-down the likelihood in closed-form. The resulting conditionally Gaussian model is a Local Level with Score-Driven (LLSD) time-varying parameters. The LLSD estimates covariances using all available data, thus avoiding the use of standard synchronization techniques. It guarantees positive-definite correlation matrices and provides filtered estimates of both the time-varying parameters and the efficient martingale price process. Other than being estimated using standard maximum-likelihood techniques, the LLSD allows to deal with large covariance matrices through a two-step maximum likelihood procedure that couples simple univariate models to a parsimonious correlation model.

The modelling of asynchronicity as a missing value problem was first developed by Corsi et al. (2015) and Shephard and Xiu (2016), who considered a state-space representation with constant parameters for

the purpose of estimating the integrated covariance of high-frequency asset prices. Here, we are interested in a different problem, namely the *dynamic* modelling of the conditional correlation of high-frequency asset prices.

In the field of high-frequency financial econometrics, the most tangible effect of asynchronous trading and microstructure noise on covariance estimation is the so-called Epps effect (Epps 1979), i.e. a downward bias of sample correlations as the sampling interval shrinks. This problem motivated researchers to build robust estimators of the high-frequency integrated covariance. For instance, Hayashi and Yoshida (2005) proposed an unbiased and consistent estimator of the quadratic covariation of a bivariate semimartingale that is observed asynchronously over time. Examples of estimators that also take into account the presence of microstructure noise have been proposed by Aït-Sahalia et al. (2010), Barndorff-Nielsen et al. (2011), Bibinger (2012), Corsi et al. (2015), Shephard and Xiu (2016). Also, there was a parallel interest in the estimation of the intraday instantaneous covariance matrix. This stream of the high-frequency financial econometric literature has received less attention compared to the problem of quadratic covariation estimation and, consequently, is less developed. However, it is similar in spirit to our purpose since we are interested in modelling the intraday dynamics of conditional volatilities and correlations. For instance, Zu and Boswijk (2014) proposed a two-scale realized spot volatility estimator while Bibinger et al. (2014) proposed a spot covariance estimator based on a local method of moments.

As discussed in Section 2.4, the computation of the score of time-varying parameters in a linear Gaussian state-space representation requires deriving the Kalman filter recursions with respect to the time-varying parameters. This approach was described by Creal et al. (2008) for a univariate local level model with time-varying variances. Here, we propose a local level model with time-varying covariance matrices that can be regarded as the multivariate generalization of the latter. Recently, the estimation of linear Gaussian state-space models with score-driven parameters has been described in more detail by Delle Monache et al. (2015) and Delle Monache et al. (2016) in the context of inflation and GDP forecasting.

In a similar fashion to DCC models, we estimate covariances in two steps. In the first step, individual volatilities are estimated using independent univariate local level models with time-varying variances and in a further step correlations are estimated conditionally on the previous volatility estimates. The problem of estimating a large dimensional model is thus reduced to numerically simpler problems. In addition, restrictions can be imposed on the structure of the parameter space of the correlation model in such a way that the number of static parameters scales well with the cross-sectional dimension. The covariance matrix is parametrized using spherical coordinates (Jaeckel and Rebonato 1999). This choice guarantees positive-definite estimates and also avoids the typical problem of over-parametrization present

in the DCC.

In Section 5.4, by performing extensive Monte-Carlo experiments, we test the performance of the model in estimating several dynamic patterns of volatilities and correlations on a broad range of scenarios that closely mimic real market conditions. We find that the LLSD provides superior performances compared to standard techniques that are typically employed when dealing with noisy and asynchronous data.

Application to transaction data of 25 NYSE assets in Section 5.5 shows that volatilities exhibit the well-known U-shape. Correlations are found to be small at the opening and then they increase, especially during the first two hours, until the last 15 minutes of the trading day, after which they tend to decrease. We study variations of the intraday patterns over time and how the release of relevant information affects the day-specific dynamics of volatilities and correlations. Indeed, since intraday patterns can be recovered separately for each day, our estimator is able to capture in real-time abrupt changes of volatilities and correlations following particular events (e.g. macro-news announcements) occurring on specific days. Out-of-sample covariance forecasts provided by the model are empirically assessed through an economically meaningful application based on intraday portfolio optimization. Following the strategy suggested by Engle and Colacito (2006), we find significant improvements and large efficiency gains compared to standard techniques.

5.1 The model

5.1.1 General framework

Let $t \in [0, S]$ and denote by $X_t = (X_t^1, \ldots, X_t^n)'$ an n-dimensional vector of intraday efficient log-prices. We consider T equally-spaced observation times $0 \leq t_0 < t_1 < \cdots < t_{T-1} \leq S$ and propose to model X_{t_i}, $i = 0, \ldots, T-1$ as a random walk with heteroskedastic Gaussian innovations:

$$X_{t_{i+1}} = X_{t_i} + \eta_{t_i}, \quad \eta_{t_i} \sim \text{NID}(0, Q_{t_i}) \tag{5.1.1}$$

Due to microstructure effects, observed prices are contaminated by noise. This implies that we do not directly observe the efficient log-price X_t but a blurred version of it. Let $Y_{t_i}^*$ be the n-dimensional vector of log-prices. Due to asynchronous trading, the components of $Y_{t_i}^*$ corresponding to assets that are traded at time t_i are observed, while observations of the remaining components are missing.

Let us denote by Y_{t_i} the n_i, $0 < n_i \leq n$ components[1] of $Y_{t_i}^*$ that are observed at time t_i. For $i = 0, \ldots, T-1$ define an $n_i \times n$ matrix W_i such that $Y_{t_i} = W_i Y_{t_i}^*$. We write the observed log-price vector Y_{t_i} as:

$$Y_{t_i} = W_i X_{t_i} + W_i \epsilon_{t_i}, \quad \epsilon_{t_i} \sim \text{NID}(0, H_{t_i}) \tag{5.1.2}$$

[1]The case $n_i = 0$ is discussed in Session (5.1.2).

where ϵ_{t_i} is a Gaussian disturbance term which is assumed to be independent on the efficient log-price X_{t_i}. Hansen and Lunde (2006) argued that endogeneity (i.e. a nonzero correlation between the microstructure noise and the efficient price process) may be important, especially for mid-quote data. Our model uses the Kalman filter and the extension of the latter to deal with correlated measurement and model errors is standard in the literature (see e.g. Simon 2006). In line with Corsi et al. (2015) and Shephard and Xiu (2016), we assume that the noise covariance matrix H_t is diagonal.

5.1.2 State-space representation

Eq. (5.1.2) describes the measurement process of the latent efficient log-price, while eq. (5.1.1) describes the dynamics of the latter at discrete times. Thus, the disturbance term ϵ_t can be thought of as summarizing microstructure effects while η_t is the idiosyncratic innovation driving X_t. We can re-write both equations as a multivariate local level:

$$Y_t = W_t X_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \bar{H}_t) \tag{5.1.3}$$

$$X_{t+1} = X_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q_t) \tag{5.1.4}$$

where, without loss of generality, we have set $t_{i+1} - t_i = 1$ for $i = 0, \dots, T-1$ and $\bar{H}_t \equiv W_t H_t W_t'$. Note that, because of missing values, the dimensionality of the observation vector Y_t changes over time.

We model time-varying parameters in an observation-driven framework and assume that H_t and Q_t are known given the set of past observations $\mathcal{Y}_{t-1} = \{Y_1, \dots, Y_{t-1}\}$ and a vector Θ of static parameters:

$$H_t = H_t(\mathcal{Y}_{t-1}, \Theta) \tag{5.1.5}$$

$$Q_t = Q_t(\mathcal{Y}_{t-1}, \Theta) \tag{5.1.6}$$

Under this assumption, model (5.1.3)-(5.1.4) is conditionally Gaussian and is susceptible of treatment with the Kalman filter (see Section 2.4). By defining $x_{t|t} = \text{E}[X_t|\mathcal{Y}_t]$, $P_{t|t} = \text{Var}[X_t|\mathcal{Y}_t]$, $x_{t+1} = \text{E}[X_{t+1}|\mathcal{Y}_t]$, $P_{t+1} = \text{Var}[X_{t+1}|\mathcal{Y}_t]$, the latter can be computed recursively through equations (2.4.5)-(2.4.10). Note that, since the state vector X_t is non-stationary, the Kalman filter requires diffuse initialization, i.e. we set $x_1 = 0$ and $P_1 = \kappa \mathbb{I}_n$ with $\kappa \to \infty$. The log-likelihood function can be computed in the usual prediction error decomposition form:

$$\log L(Y_1, \dots, Y_T) = -\frac{nT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \left(\log |F_t| + v_t' F_t^{-1} v_t \right) \tag{5.1.7}$$

The advantages of having a linear Gaussian state-space representation are numerous. Beside the fact that the model can be estimated using standard maximum likelihood techniques, another relevant aspect is that Kalman filter recursions still hold in presence of missing observations. The only modification is the introduction of the matrix W_t which selects the components of Y_t that are actually observed at

time t. Therefore, the problem of modelling asynchronicity is equivalent to having missing values in the observation vector Y_t^* and these are easily tackled by the Kalman filter.

In case H_t and Q_t are constant over time, model (5.1.3)-(5.1.4) can be conveniently estimated using the expectation maximization (EM) method, as described by Shumway and Stoffer (1982) and Durbin and Koopman (2012). In order to apply the Kalman filter recursions and write down the log-likelihood function, we need to specify the law of motion of H_t and Q_t.

5.1.3 The LLSD

Let us collect all the time-varying parameters of the model in a k-dimensional column vector f_t. The score-driven update rule is given by:

$$f_{t+1} = \omega + As_t + Bf_t \tag{5.1.8}$$

where ω is a $k \times 1$ vector, A, B are $k \times k$ matrices and s_t is the scaled derivative of the log-density with respect to the time-varying parameter vector f_t. As described in Section 2.3, we write it as:

$$s_t = S_t \nabla_t \tag{5.1.9}$$

where

$$\nabla_t = \left[\frac{\partial \log p(y_t | f_t, \mathcal{Y}_{t-1}; \Theta)}{\partial f_t'} \right]' \tag{5.1.10}$$

The vector Θ of static parameters includes the elements of ω, A and B. In our application to high-frequency financial data, we will set $S_t = (\mathcal{I}_{t|t-1})^{-1/2}$, i.e. we take the principal square root matrix of the inverse of $\mathcal{I}_{t|t-1}$. Motivated by the generality and flexibility of the score driven approach[2], we model dynamically volatilities and correlations in the multivariate local level using the update equation (5.1.8). The resulting model is named Local Level with Score-Driven (LLSD) time-varying parameters.

The time-varying parameters in the LLSD are the elements of the diagonal noise covariance matrix H_t and those in the state vector covariance matrix Q_t. As in the DCC model and in the multivariate t-GAS model of Creal et al. (2011), we decompose Q_t as:

$$Q_t = D_t R_t D_t \tag{5.1.11}$$

where R_t is the correlation matrix and D_t is the diagonal matrix of standard deviations. The correlation matrix R_t is decomposed as:

$$R_t = Z_t' Z_t \tag{5.1.12}$$

where Z_t is an $n \times n$ matrix containing $k_Z = n(n-1)/2$ time-varying parameters. The parametrization of Z_t will be discussed in detail in Section 5.3. Note that the present decomposition of Q_t and R_t guarantees that both matrices are positive-definite by construction. The k-dimensional time-varying parameter

[2]See Section 2.3 for a general discussion on score-driven models.

vector f_t includes the $2n$ elements of H_t, D_t and the k_Z elements of Z_t and therefore $k = 2n + n(n-1)/2$ is the number of time-varying parameters in the LLSD model.

5.2 Estimation

A major challenge in modelling time-varying correlations is handling large dimensionality. The DCC model has a clear computational advantage over multivariate GARCH models in that the number of parameters to be estimated in the correlation process is independent on the cross-sectional dimension n. This is possible if one uses a two-step maximum likelihood procedure where one first separately estimates n GARCH models and then inputs the estimated variances into the DCC. If one assumes that correlations have same persistences, the number of static parameters to be estimated in the last step will be independent on the dimension n.

Thanks to the particular structure of the local level model, this relevant feature of the DCC can be incorporated into the LLSD model. In the language of Harvey (1991), the multivariate local level model is a particular case of a Seemingly Unrelated Time Series Equation (SUTSE) model in that each series is modeled as in the univariate case but disturbances are correlated across series. Therefore, one can employ n univariate LLSD models to separately estimate the time-varying parameters in H_t and D_t, i.e. the time-varying variances of the noise and the standard deviations of the innovations in the transition equation. In a further step, the time-varying correlations in R_t can be estimated as if H_t and D_t were known and given by the previous estimates.

In the following sections, we will describe in detail the two-step estimation procedure. First, we discuss the estimation of the time-varying parameters in Z_t. The k_Z-dimensional time-varying parameter vector will be denoted as f_t^Z. The results that will be recovered in this Section are independent on the parametrization of Z_t. We refer to the dynamics of f_t^Z as the *correlation model*. Second, we discuss the estimation of the noise covariance matrix H_t and the matrix of standard deviations D_t of the state vector. We will consider n independent univariate LLSD models where the time-varying parameter vector is a 2-dimensional vector that we will denote as $f_{t,i}$ for $i = 1, \ldots, n$. This univariate specification coincides with the one proposed by Creal et al. (2008).

In practice, one first estimates the n univariate models and then uses the estimated matrices H_t and D_t as an input to the correlation model. In the following paragraph, we describe in detail the correlation model. The univariate local-level with time-varying variances is discussed in appendix (A.8).

5.2.1 The correlation model

Let us consider the multivariate local level model (5.1.3), (5.1.4) and assume that H_t and D_t are known. As underlined above, f_t^Z is a k_Z-dimensional column vector that includes all the time-varying parameters

in Z_t. Let us define:

$$\nabla_t^Z = \frac{\partial \log p(Y_t|\mathcal{Y}_{t-1})}{\partial f_t^{Z\prime}}, \quad \mathcal{I}_{t|t-1}^Z = \mathrm{E}[\nabla_t^Z \nabla_t^{Z\prime}] \tag{5.2.1}$$

As discussed in Section 2.4, the score and the information matrix of the conditional log-likelihood are given by:

$$\nabla_t^Z = -\frac{1}{2}\left[\dot{F}_t'(I_{n_t} \otimes F_t^{-1})\mathrm{vec}(I_{n_t} - v_t v_t' F_t^{-1}) + 2\dot{v}_t' F_t^{-1} v_t\right] \tag{5.2.2}$$

$$\mathcal{I}_{t|t-1}^Z = \frac{1}{2}\left[\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})\dot{F}_t + 2\dot{v}_t' F_t^{-1}\dot{v}_t\right] \tag{5.2.3}$$

where $\dot{F}_t = \partial \mathrm{vec}(F_t)/\partial f_t^Z$ and $\dot{v}_t = \partial v_t/\partial f_t^Z$. The two terms $\dot{F}_t$ and $\dot{v}_t$ needed to evaluate ∇_t and $\mathcal{I}_{t|t-1}$ can be computed recursively as:

$$\dot{v}_t = -W_t \dot{x}_t \tag{5.2.4}$$

$$\dot{F}_t = (W_t \otimes W_t)\dot{P}_t \tag{5.2.5}$$

where:

$$\dot{x}_{t+1} = \dot{x}_t + (v_t' \otimes I_n)\dot{K}_t + K_t \dot{v}_t \tag{5.2.6}$$

$$\dot{P}_{t+1} = \dot{P}_t - (K_t W_t \otimes I_n)\dot{P}_t - (I_n \otimes P_t W_t')K_{nn}\dot{K}_t + \dot{Q}_t \tag{5.2.7}$$

$$\dot{K}_t = (F_t^{-1} W_t \otimes I_n)\dot{P}_t - (F_t^{-1} \otimes K_t)\dot{F}_t \tag{5.2.8}$$

$$\dot{Q}_t = (D_t \otimes D_t)\dot{R}_t \tag{5.2.9}$$

$$\dot{R}_t = (\dot{Z}_t \otimes I_n)K_{nn}\dot{Z}_t + (I_n \otimes Z_t')\dot{Z}_t \tag{5.2.10}$$

The above formulas are recovered by deriving the Kalman filter recursions (2.4.5)-(2.4.10) with respect to f_t^Z and are a particular case of the recursions appearing in Delle Monache et al. (2016). Note that D_t and H_t are not derived since they are freezed and equal to the output of the univariate estimation.

At time t, the one-step-ahead prediction of the time-varying vector f_t^Z is given by:

$$f_{t+1}^Z = \omega_Z + A_Z s_t + B_Z f_t^Z \tag{5.2.11}$$

where ω_Z, A_Z and B_Z are static parameters of appropriate dimensions. Let $\Theta_Z = \{\omega_Z, A_Z, B_Z\}$ collect all the static parameters in the correlation model. The scaled score s_t is chosen as $s_t = (\mathcal{I}_{t|t-1}^Z)^{-1/2}\nabla_t^Z$ and is computed using eq. (5.2.2), (5.2.3) and the derivative recursions (5.2.4)-(5.2.7), which run in parallel with the Kalman filter recursions. The conditional log-likelihood is given in the prediction error decomposition form:

$$\log L_Z(Y_1, \ldots, Y_T|\Theta_Z) = \mathrm{const} - \frac{1}{2}\sum_{t=1}^{T}\frac{1}{2}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \tag{5.2.12}$$

where v_t and F_t are evaluated using the Kalman filter recursions and the score-driven update equation with parameters given by Θ_Z. The maximum likelihood estimate of the static parameter vector Θ_Z is simply obtained by numerically optimizing the log-likelihood function through a quasi-Newton algorithm:

$$\hat{\Theta}_Z = \underset{\Theta}{\operatorname{argmax}} \log L_Z(Y_1, \ldots, Y_T|\Theta) \tag{5.2.13}$$

5.3 The TAP parametrization for Z_t

As a parametrization for Z_t, we use the triangular angles parametrization (TAP) introduced by Jaeckel and Rebonato (1999) and described in detail by Rapisarda et al. (2007). The TAP has been used in the financial econometric literature. For instance, Creal et al. (2011) employed it in their multivariate t-GAS model for time-varying covariances. In the TAP parametrization, the generic element $z_{ij,t}$ of the matrix Z_t in eq. (5.1.12) is given by:

$$z_{ij} = \begin{cases} 1 & i = j = 1 \\ \cos\theta_{1j} & i = 1, j > 1 \\ \prod_{k=1}^{i-1} \sin\theta_{kj} \cos\theta_{ij} \mathbf{1}_{j>i} + \prod_{k=1}^{i-1} \sin\theta_{kj} \mathbf{1}_{i=j} & i > 1 \end{cases} \tag{5.3.1}$$

where we have suppressed the subscript t for ease of notation. Thus, denoting by $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$, the matrix Z_t has the following form:

$$Z_t = \begin{pmatrix} 1 & c_{12} & c_{13} & \ldots & c_{1n} \\ 0 & s_{12} & c_{23}s_{13} & \ldots & c_{2n}s_{1n} \\ 0 & 0 & s_{23}s_{13} & \ldots & c_{3n}s_{2n}s_{1n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & \prod_{k=1}^{n-1} s_{kn} \end{pmatrix} \tag{5.3.2}$$

The i-th column of Z_t contains the spherical coordinates of a vector of unit norm in an ith-dimensional subspace of $\mathbb{R}^n$, which is parametrized by $i-1$ angles. Therefore, we have a total of $n(n-1)/2$ angles $[\theta_{12,t}, \theta_{13,t}, \ldots, \theta_{n-1n,t}]$. The two main advantages of the TAP are that it guarantees positive-definite estimates of R_t and that the number of free parameters is $n(n-1)/2$. In contrast to the DCC parametrization, where there are n extra parameters, the number of parameters is the same as the degrees of freedom of the correlation matrix. Thus, the Fisher information matrix is full-rank and the static parameters are identifiable. Note that, as pointed out by Rapisarda et al. (2007), every correlation matrix can be written in the TAP form.

At time t, the vector f_t^Z of time-varying parameters is a $k_Z = n(n-1)/2$ dimensional vector containing all the angles $\theta_{12,t}, \theta_{13,t}, \ldots, \theta_{n-1n,t}$. In order to perform recursion (5.2.10), we need to know the derivative

of Z_t with respect to f_t^Z. This is easily computed by observing that:

$$\frac{\partial z_{ij}}{\partial \theta_{lm}} = \begin{cases} 0 & i > j, j \neq m, l \geq m, l > i \\ -z_{ij} \tan \theta_{ij} & i < j, l = i \\ \frac{z_{ij}}{\tan \theta_{ij}} & i \leq j, l < i \end{cases} \tag{5.3.3}$$

5.4 Simulation study

5.4.1 Finite sample properties

The finite sample properties of the maximum likelihood estimator are tested in a Monte-Carlo experiment where bivariate time-series are generated using the LLSD as a DGP. We choose a simulation setting that resembles as closest as possible the main features of high-frequency data. In particular, we assume a trading day of 6.5 hours and simulate $N = 2000$ one-second time-series of $T = 23400$ observations. Asynchronicity is reproduced by censoring the simulated observations using Poisson sampling. The probability of missing values is set equal to $\Lambda = 0,\ 0.5,\ 0.8$. The last two scenarios are commonly observed in real transaction data, as it will be shown in Section 5.5.

Λ	ω_Z	A_Z	B_Z
	BIAS		
0	0.00056	0.00085	-0.00256
0.5	0.00261	-0.00660	-0.00562
0.8	0.00308	-0.00719	-0.00743
	Standard Dev.		
0	0.00778	0.01373	0.00725
0.5	0.01238	0.02258	0.01254
0.8	0.01435	0.03075	0.01569

Table 5.4.1: Bias and standard deviations of $N = 2000$ maximum likelihood estimates of static parameters ω_Z, A_Z, B_Z for different probabilities of missing values $\Lambda = 0,\ 0.5,\ 0.8$.

For simplicity, individual volatilities are assumed to be constant over time. This is obtained by setting $\omega_i = 0$, $A_i = 0_{2\times 2}$ and $B_i = \mathbb{I}_{2\times 2}$ for $i = 1, 2$. The initial values $f_{1,i}$ are chosen in order to have a signal-to-noise ratio $\delta = \xi_i^2/\sigma_i^2 = 1$, for $i = 1, 2$. The latter is close to the typical values that are found on empirical data. Different choices of the static parameters in the univariate model lead to similar results.

The static parameters of the correlation model have the following values:

$$\omega_Z = 0.021, \quad A_Z = 0.1, \quad B_Z = 0.98 \tag{5.4.1}$$

The LLSD is estimated using the two-step procedure described in Section 5.2. Initial values of the time-varying parameters are set equal to their unconditional means. Table (5.4.1) shows the bias and sample

standard deviations of the maximum likelihood estimates of ω_Z, A_Z and B_Z while figure (5.4.1) shows their kernel density estimates. In the no missing scenario, the bias is less than 1% for A_Z and B_Z while it is slightly larger for ω_Z. For $\Lambda = 0.5, 0.8$, as a result of missing observations, the distribution of ω_Z and B_Z becomes more skewed and fat tailed. However, the bias remains small in relative terms and even for $\Lambda = 0.8$, where the sample is highly asynchronous, the parameter B_Z is accurately recovered with a bias lower than 1%.

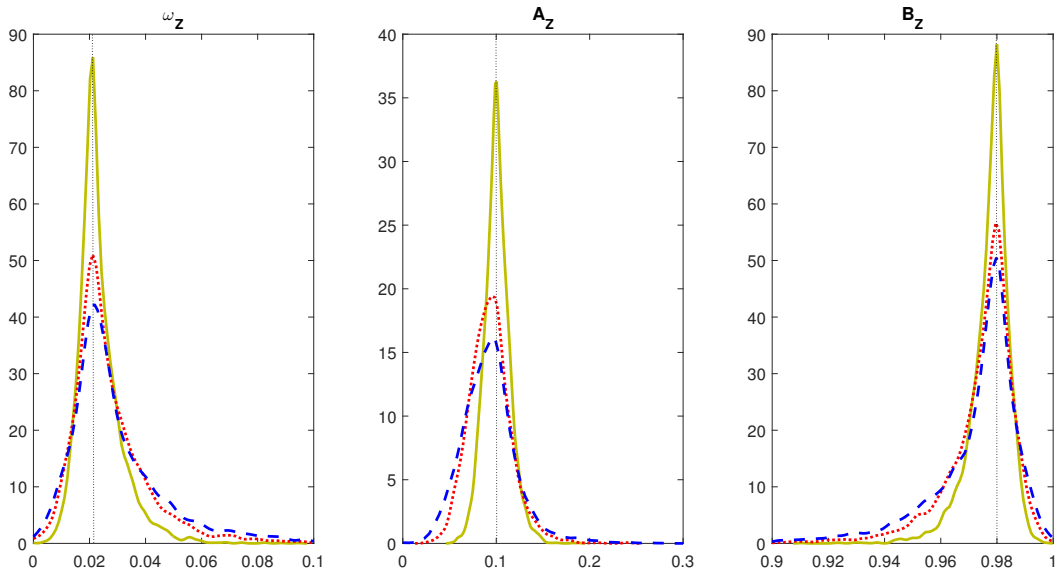


Figure 5.4.1: Kernel density estimates of maximum likelihood estimates of parameters ω_Z, A_Z and B_Z for $\Lambda = 0$ (yellow lines), $\Lambda = 0.5$ (dotted red lines) and $\Lambda = 0.8$ (dashed blue lines).

5.4.2 Monte-Carlo analysis based on misspecified DGP

We first deal with the problem of choosing challenging benchmark models with which to compare the forecasting performance of the LLSD. The DCC model and other existing multivariate conditional correlation models are misspecified in presence of additive noise. In order to assess the impact of the noise term, we simulate a bivariate local level model as in (5.1.3), (5.1.4) with $T = 5000$ timestamps. The two matrices H_t and D_t are kept constant while the correlation in R_t follows the deterministic path in figures (5.4.2a) - (5.4.2d). The signal-to-noise ratio δ is assumed to be the same for the two simulated series. If δ is very large (e.g. $\delta = 10$) the estimate provided by the DCC is close to that of the LLSD. However, as δ decreases, the noise leads to downward biased correlation estimates. Note that, except for the very large value $\delta = 10$, the other values are commonly observed in real markets, as we will see in the empirical application in Section 5.5.

In order to attenuate the effect of the noise on the estimates provided by the DCC, we proceed in

two different ways. The first is to compute returns by subsampling at larger time-scales, thus neglecting part of the available observations. This has the effect of increasing the signal-to-noise ratio. We name as DCC_{sub} the correlation estimates obtained by the DCC on the new sample with returns computed on a time-scale $K > 1$. The second method is to apply a pre-averaging procedure (Jacod et al. 2009). Indeed, since ϵ_t is uncorrelated, the variance of the noise term of $1/K \sum_{j=0}^{K-1} Y_{t_i-j}$, is $1/K$ times the variance of the noise of Y_{t_i}. Therefore, the effect of the pre-averaging is, again, to increase the signal-to-noise ratio. We name the correlation estimates obtained by the DCC on the pre-averaged sample as DCC_{pre}. In both cases the tuning parameter K is chosen as the one minimizing the mean square error in formula (5.4.2) over a sample of $N = 100$ independent realizations. In figures (5.4.2e), (5.4.2f) we show the estimates provided by the LLSD, DCC_{sub} and DCC_{pre} for $\delta = 1$ and $\delta = 0.5$, respectively. Differently from the previous case, DCC_{sub} and DCC_{pre} are now able to capture the dynamics of the simulated correlation pattern.

In order to test the forecasting performance of the LLSD under a misspecified DGP, we design two different simulation settings, one for the correlation model and a different one for the univariate variance model. In the former case, the Monte-Carlo experiments that will be performed are similar to those conducted by Engle (2002a) and Creal et al. (2011). While keeping variances constant, we estimate a bivariate LLSD over a set of deterministic and stochastic correlation patterns. Different levels of noise and asynchronicity are simulated in order to closely mimic real market conditions.

A relevant advantage of the LLSD is that the conditional correlation matrix is estimated using all the available data. This implies more efficient estimates compared to standard synchronization schemes where a significant fraction of the observations is typically neglected. We examine the impact of asynchronicity in a higher-dimensional experiment where LLSD estimates are compared to those provided by the DCC model estimated on a synchronized sample. The chosen synchronization scheme is the refresh-time described by Barndorff-Nielsen et al. (2011).

In the bivariate and univariate tests, our loss measures are the mean squared error (MSE) and the mean absolute error (MAE) which are computed as:

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(\theta_t - \hat{\theta}_t)^2, \quad \text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|\theta_t - \hat{\theta}_t| \tag{5.4.2}$$

where θ_t generically denote the simulated variable, i.e. the correlation in the first test and the variance in the second test, while $\hat{\theta}_t$ denotes the estimated variable. In the higher dimensional experiment, we use the Frobenius norm $\|\cdot\|_F$ and the $1-$norm $\|\cdot\|_1$ defined as:

$$\|R_t - \hat{R}_t\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(R_{t,ij} - \hat{R}_{t,ij})^2}, \quad \|R_t - \hat{R}_t\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|R_{t,ij} - \hat{R}_{t,ij}| \tag{5.4.3}$$

where R_t is the simulated correlation matrix and $\hat{R}_t$ is the estimated one.

5.4.3 Monte-Carlo design I: time-varying correlations

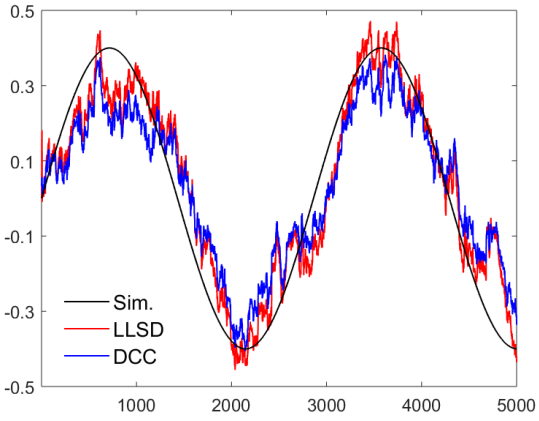
We assume again a trading day of 6.5 hours and simulate $N = 250$ bivariate time series of $T = 23400$ observations from the time-varying local level model (5.1.3), (5.1.4). Following Engle (2002a) and Creal et al. (2011), we keep the two variances $\sigma^2_{t,i}$, $\xi^2_{t,i}$, $i = 1, 2$ constant while the off-diagonal element of the correlation matrix R_t, that we will denote as ρ_t, evolves dynamically according to the following patterns:

1. Sine $\rho_t = a_s + b_s \sin(c_s \pi t)$
2. Fast Sine $\rho_t = a_f + b_f \sin(c_f \pi t)$
3. Step $\rho_t = \alpha_s - \beta_s(t > \gamma_s)$
4. Ramp $\rho_t = \frac{1}{b_r} \mod (t + a_r, b_r)$
5. Model $\rho_t = \exp(h_t)/(1 + \exp(h_t))$

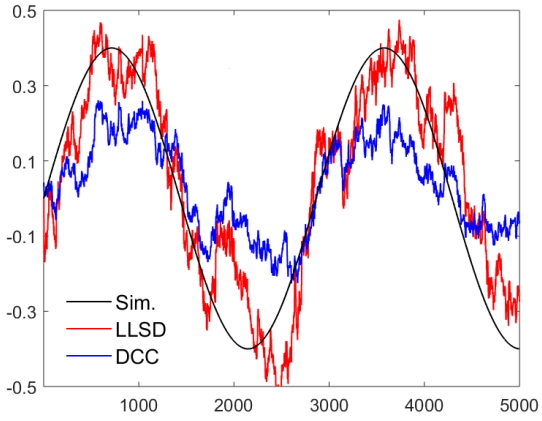
where h_t follows an AR(1) process:

$$h_{t+1} = c_m + b_m h_t + a_m \phi_t, \quad \phi_t \sim \mathrm{N}(0, 1)$$

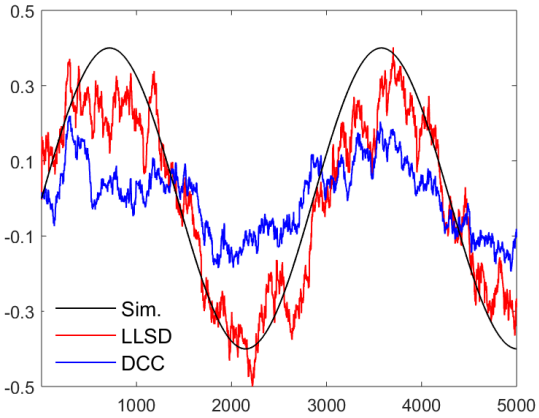
The parameters appearing in the dynamics of the correlation coefficient ρ_t are chosen in the following way: $a_s = a_f = 0$, $b_s = b_f = 0.4$, $c_s = 7/(2T)$, $c_f = 101/(2T)$, $\alpha_s = 0.25$, $\beta_s = 0.5$, $\gamma_s = T/2$, $b_r = T/20$, $a_r = b_r/2$, $b_m = 0.997$, $c_m = -0.4(1 - b_m)$, $a_m = 0.05$. The variance $\xi^2_{t,i}$ of the latent process is constant and equal to 0.1 for all the simulated patterns. Instead, the variance $\sigma^2_{t,i}$ of the noise is computed based on the chosen signal-to-noise ratio δ which is varied in the simulations in order to reproduce different scenarios: $\delta = 2$ (high signal), $\delta = 1$ (moderate signal) and $\delta = 0.5$ (low signal). As we will see in Section 5.5, these three scenarios are close to those observed in real data.



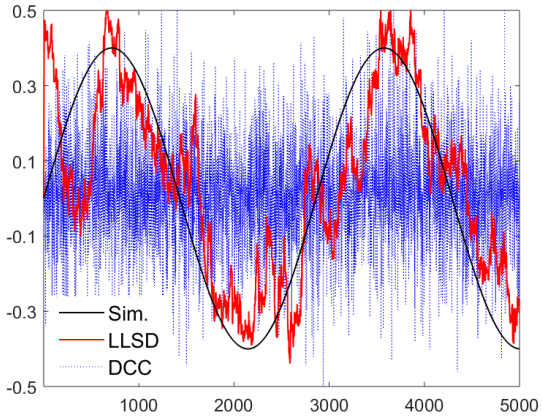
(a) $\delta = 10$



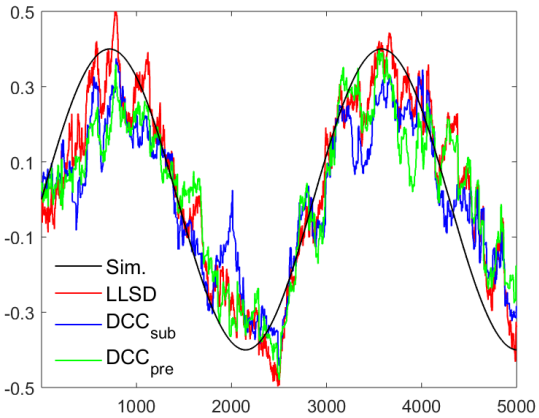
(b) $\delta = 2$



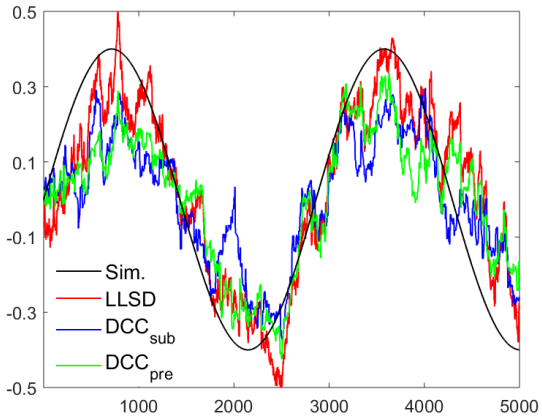
(c) $\delta = 1$



(d) $\delta = 0.5$



(e) $\delta = 1$



(f) $\delta = 0.5$

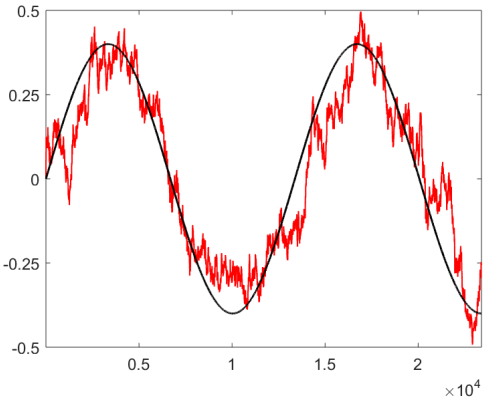
Figure 5.4.2: (a)-(d) Examples of correlation estimates on a bivariate time-series of $T = 5000$ timestamps provided by the LLSD and the DCC using different levels of noise. (e), (f) Estimates provided by the LLSD, DCC_{sub} and DCC_{pre} for large noise scenarios.

As in the previous experiment, the probability of missing values is set equal to $\Lambda = 0$, $\Lambda = 0.5$ and $\Lambda = 0.8$. In the simulations each signal-to-noise ratio scenario is combined to each missing values scenario in order to study the behavior of the filtered estimates under a multiplicity of realistic market conditions. Figures (5.4.3a) - (5.4.3e) show the simulated correlation patterns and the corresponding LLSD estimates in the scenario $\delta = 1$, $\Lambda = 0.5$.

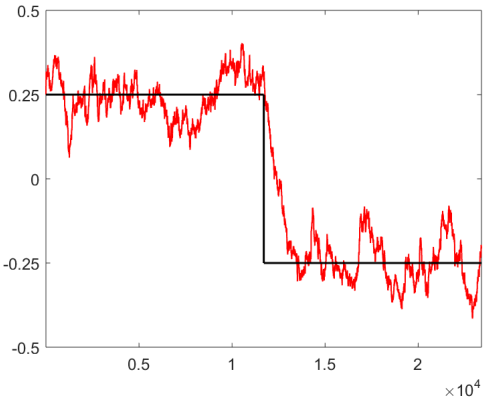
We divide the sample into two parts of 11700 timestamps each. The first sub-sample is used to estimate the LLSD, DCC_{sub}, and DCC_{pre} while the second is used for computing out-of-sample forecasts. In order to estimate the LLSD, we apply the two-step procedure described in Section 5.2. In both steps we first estimate a local level model with constant parameters and then set the initial values of time-varying parameters in the LLSD equal to the latter ML estimates.

The average MSE (MAE) of the LLSD is compared to that of DCC_{sub} and DCC_{pre} in table (5.4.3). In scenarios with missing values, the DCC_{sub} and DCC_{pre} are estimated by first synchronizing the sample using the refresh-time. The LLSD outperforms the DCC_{sub} and DCC_{pre} on all of the simulated scenarios while the DCC_{pre} outperforms the DCC_{sub} in most of the cases. In the "no missing" scenario, the relative difference between the LLSD and the other two estimators is large for all the simulated patterns. Indeed, the Kalman filter is able to optimally estimate the state vector, thus allowing to capture the dynamics of correlations even in presence of very noisy observations.

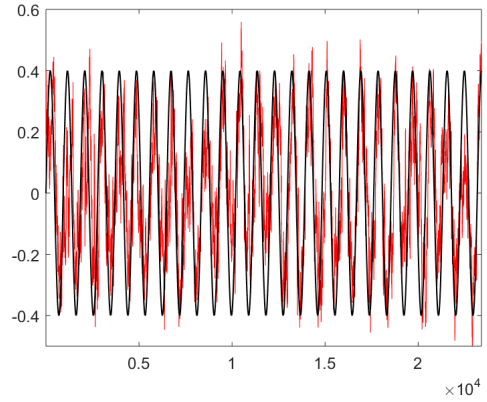
In the two scenarios ($\Lambda = 0.5, 0.8$), the effect of the noise diminishes because the average duration between observations increases. However, the use of the refresh-time implies neglecting a significant part of the available data and at the same time alters the instantaneous correlation structure of the observations. Indeed, we note that the relative difference between the LLSD and the other two estimators is larger in case of patterns with abrupt changes (e.g. step and ramp) and in stochastic patterns (model) while it is more moderate in case of smooth patterns (e.g. sine and fast sine).



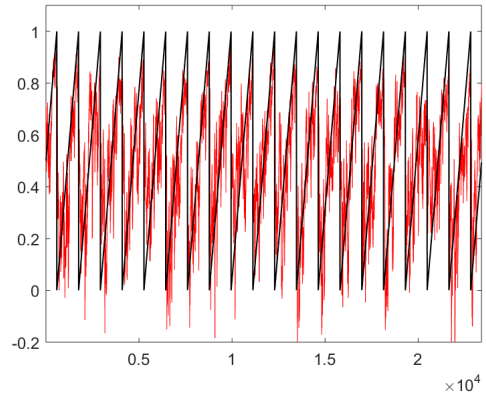
(a) Sine



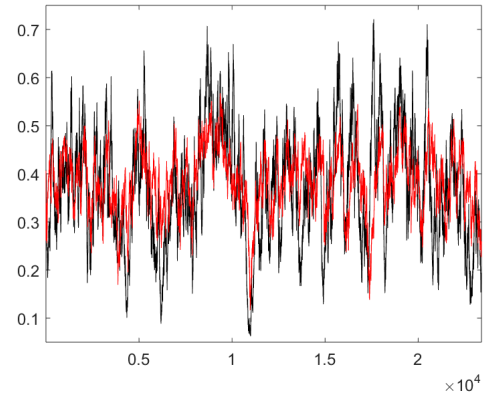
(b) Step



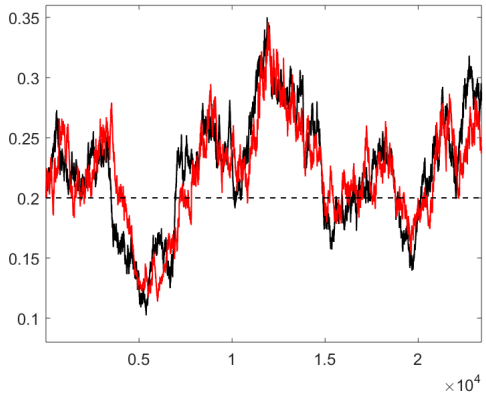
(c) Fast sine



(d) Ramp



(e) Model



(f) CIR

Figure 5.4.3: (a)-(e) Examples of correlation estimates provided by the LLSD for the five simulated patterns in Section 5.4.3 in the scenario $\Lambda = 0.5$, $\delta = 1$ for a bivariate time-series of $T = 23400$ timestamps. (f) LLSD estimate of the time-varying variance of the efficient log-price process. The horizontal dashed line denotes the value of the noise variance.

In order to provide a better understanding of the advantages of the LLSD in dealing with asynchronous data, we design a new Monte-Carlo experiment where 5-dimensional time-series of $T = 23400$ timestamps are simulated using the DCC as a DGP. The simulated observations are then randomly censored as in the previous experiment. The first 11700 timestamps are used to estimate the models, while the second part of the sample is used for assessing loss measures. We consider the five scenarios described in table (5.4.2).

As in the bivariate case, the DCC is estimated by first synchronizing the sample using the refresh-time. There is no noise added to observations and therefore the relative difference between the LLSD and the DCC is entirely due to the effect of asynchronicity. Table (5.4.4) shows the average, over $N = 250$ repetitions, of $\|\cdot\|_F$ and $\|\cdot\|_1$ norms between the simulated correlation matrix and the one estimated by the LLSD and the DCC. Note that, even though the DCC is correctly specified while the LLSD is not, the latter provides better estimates and the relative difference between the two increases as the level of asynchronicity increases.

Scenario	asset 1	asset 2	asset3	asset 4	asset 5
Λ_1	0.1	0.1	0.1	0.1	0.1
Λ_2	0.1	0.1	0.1	0.1	0.8
Λ_3	0.3	0.3	0.3	0.3	0.3
Λ_4	0.5	0.5	0.5	0.5	0.5
Λ_5	0.8	0.8	0.8	0.8	0.8

Table 5.4.2: Probability of missing values for each asset on each of the five scenarios

Estimator	Sine	Fast	Step	Ramp	Model	Sine	Fast	Step	Ramp	Model	Sine	Fast	Step	Ramp	Model
			$\delta = 0.5$					$\delta = 1$					$\delta = 2$		
								$\Lambda = 0$							
LLSD	0.0087	0.0395	0.0091	0.0434	0.0123	0.0067	0.0328	0.0078	0.0383	0.0111	0.0053	0.0272	0.0069	0.0329	0.0099
DCC_{sub}	0.0249	0.0718	0.0196	0.0898	0.0312	0.0183	0.0685	0.0148	0.0707	0.0210	0.0153	0.0666	0.0126	0.0626	0.0170
DCC_{pre}	0.0146	0.0668	0.0128	0.0595	0.0188	0.0141	0.0667	0.0124	0.0589	0.0184	0.0139	0.0665	0.0123	0.0588	0.0179
LLSD	0.0749	0.1637	0.0626	0.1577	0.0887	0.0657	0.1480	0.0570	0.1441	0.0842	0.0583	0.1341	0.0522	0.1309	0.0797
DCC_{sub}	0.1336	0.2354	0.1191	0.2506	0.1437	0.1121	0.2283	0.0958	0.2210	0.1168	0.1011	0.2235	0.0845	0.2033	0.1050
DCC_{pre}	0.0975	0.2188	0.0848	0.1954	0.1103	0.0954	0.2178	0.0831	0.1916	0.1092	0.0945	0.2176	0.0825	0.1900	0.1079
								$\Lambda = 0.5$							
LLSD	0.0139	0.0527	0.0116	0.0522	0.0143	0.0111	0.0491	0.0088	0.0470	0.0134	0.0099	0.0439	0.0074	0.0426	0.0125
DCC_{sub}	0.0277	0.0731	0.0215	0.0969	0.0347	0.0207	0.0702	0.0164	0.0764	0.0236	0.0174	0.0685	0.0139	0.0672	0.0189
DCC_{pre}	0.0180	0.0687	0.0154	0.0682	0.0204	0.0159	0.0682	0.0139	0.0627	0.0188	0.0151	0.0681	0.0133	0.0609	0.0187
LLSD	0.0945	0.1926	0.0733	0.1776	0.0960	0.0849	0.1844	0.0644	0.1658	0.0927	0.0791	0.1732	0.0601	0.1551	0.0895
DCC_{sub}	0.1421	0.2383	0.1274	0.2601	0.1523	0.1206	0.2321	0.1038	0.2311	0.1238	0.1089	0.2279	0.0915	0.2140	0.1108
DCC_{pre}	0.1103	0.2259	0.0976	0.2165	0.1150	0.1025	0.2225	0.0902	0.2039	0.1103	0.0991	0.2211	0.0871	0.1980	0.1101
								$\Lambda = 0.8$							
LLSD	0.0257	0.0737	0.0171	0.0641	0.0167	0.0233	0.0716	0.0160	0.0608	0.0161	0.0203	0.0695	0.0150	0.0582	0.0156
DCC_{sub}	0.0361	0.0768	0.0270	0.1133	0.0433	0.0293	0.0753	0.0222	0.0943	0.0321	0.0258	0.0744	0.0198	0.0848	0.0269
DCC_{pre}	0.0281	0.0811	0.0232	0.0956	0.0343	0.0239	0.0799	0.0201	0.0842	0.0293	0.0220	0.0787	0.0187	0.0793	0.0278
LLSD	0.1278	0.2358	0.0946	0.2053	0.1042	0.1218	0.2319	0.0905	0.1975	0.1023	0.1206	0.2288	0.0878	0.1912	0.1004
DCC_{sub}	0.1643	0.2455	0.1461	0.2800	0.1722	0.1463	0.2423	0.1269	0.2572	0.1459	0.1359	0.2401	0.1162	0.2442	0.1326
DCC_{pre}	0.1413	0.2423	0.1286	0.2579	0.1496	0.1287	0.2417	0.1152	0.2423	0.1370	0.1223	0.2420	0.1089	0.2344	0.1325

Table 5.4.3: Results of the Monte-Carlo experiment I for all the 9 scenarios described in Section 5.4.3. For each scenario and for each of the five simulated correlation patterns, we show the average MSE (MAE) of the LLSD, DCC_{sub} and DCC_{pre} estimated on bivariate time-series of $T = 23400$ timestamps and based on $N = 250$ independent repetitions.

Estimator	Λ_1	Λ_2	Λ_3	Λ_4	Λ_5
	$\|\cdot\|_F$ norm				
LLSD	0.4132	0.4496	0.4483	0.4771	0.5407
DCC	0.4558	0.5467	0.5937	0.6671	0.7653
relative	0.9065	0.8224	0.7551	0.7152	0.7065
	$\|\cdot\|_1$ norm				
LLSD	0.4254	0.4655	0.4608	0.4896	0.5530
DCC	0.4688	0.5598	0.6068	0.6797	0.7792
relative	0.9074	0.8315	0.7594	0.7203	0.7097

Table 5.4.4: Results of Monte-Carlo experiments with $n = 5$ and using the DCC as a DGP. The first two lines show the average of the $\|\cdot\|_F$ ($\|\cdot\|_1$) norm over $N = 250$ independent repetitions while the third line shows the average of the $\|\cdot\|_F$ ($\|\cdot\|_1$) norm of the LLSD relative to that of the DCC.

5.4.4 Monte-Carlo design II: time-varying volatilities

We simulate $N = 250$ univariate time series of $T = 23400$ observations from model (A.8.1), (A.8.2). The variance σ_t^2 of the noise is kept constant and equal to a value that we denote as σ^2. Thus, the only time-varying parameter is the volatility ξ_t of the state variable.

We consider a CIR stochastic volatility model as a DGP for the variance:

$$d\xi_t^2 = k(\theta - \xi_t^2)dt + w\xi_t dW_t \tag{5.4.4}$$

Simulations are performed by discretizing the process in the Euler scheme. The values assigned to the parameters are the following: $\theta = 0.2$, $k = 10$, $w = 0.5$. The first observation of the CIR process is drawn from a Gamma distribution $\Gamma(2k\theta/w^2, w^2/2k)$ centered in the mean variance. We define the average signal-to-noise ratio as $\bar{\delta} = \theta/\sigma^2$ and perform the simulation in three scenarios: low signal ($\bar{\delta} = 0.5$), moderate signal ($\bar{\delta} = 1$) and high signal ($\bar{\delta} = 2$). As in the previous experiment, the first 11700 observations are used for estimation purposes, while the second part of the sample is used for computing loss measures. An example of a simulated path, together with the estimates provided by the LLSD model in the scenario $\delta = 1$ is given in figure (5.4.3f).

In order to compare the estimates of the LLSD with a reliable benchmark, we use the two methods employed in the bivariate test, i.e. the computation of returns on larger time-scales and the pre-averaging approach. We name the variance estimates provided by the GARCH model estimated on these new samples as GARCH_{sub} and GARCH_{pre}, respectively.

Table (5.4.5) contains the results of the Monte-Carlo simulations. We show the average MSE (MAE) for the three scenarios $\bar{\delta} = 0.5, 1, 2$ obtained by estimating the LLSD, GARCH_{sub} and GARCH_{pre}. The LLSD provides better estimates than the other two estimators and the relative difference increases as the level of the noise increases. Again, thanks to the use of the Kalman filter which provides optimal

Estimator	$\bar{\delta} = 0.5$	$\bar{\delta} = 1$	$\bar{\delta} = 2$
	MSE×100		
1−LLSD	0.0866	0.0730	0.0623
GARCH$_{sub}$	0.2616	0.1946	0.1701
GARCH$_{pre}$	0.2139	0.1272	0.0862
	MAE		
1−LLSD	0.0233	0.0214	0.0198
GARCH$_{sub}$	0.0416	0.0354	0.0338
GARCH$_{pre}$	0.0386	0.0275	0.0232

Table 5.4.5: Results of Monte-Carlo experiment II. We show the average MSE (MAE) of the LLSD, GARCH$_{sub}$, GARCH$_{pre}$ estimators over $N = 250$ independent repetitions. The three simulated scenarios are: low signal ($\bar{\delta} = 0.5$), moderate signal ($\bar{\delta} = 1$) and high signal ($\bar{\delta} = 2$).

estimates of the state variable, the dynamics of the variance is well captured even in case of very noisy observations.

5.5 Empirical illustration

5.5.1 Dataset

The dataset contains transaction prices of 25 assets traded in the NYSE in 2014. The exchange opens at 9.30 and closes at 16.00 local time, so that the number of seconds per day is $T = 23400$. We follow the same procedure described in Section 4.3 to clean the data. Table (5.5.1) shows the assets together with the average number of observations per day $\overline{m}$, the probability of missing values $\Lambda = 1 - \overline{m}/T$, the average duration in seconds between observations $\overline{\Delta t}$ and the average signal-to-noise ratio as estimated by the LLSD. Note that the probability of missing values is $\Lambda \sim 80\%$ for Exxon, which is the most frequently traded asset. This implies that the high-missing scenario that we considered in the simulation study is quite realistic, even for the most liquid assets. In this highly asynchronous scenario, the LLSD provides reliable estimates of the time-varying correlations, even in presence of patterns characterized by abrupt changes and/or random behavior.

5.5.2 Model specification

We implement the univariate LLSD under the following restrictions:

- A_i is diagonal
- $B_i = I_2$
- $\omega_i = (0, 0)'$

Symbol	$\overline{\Delta t}$	Λ	$\overline{m}$	$\overline{\delta}$	Symbol	$\overline{\Delta t}$	Λ	$\overline{m}$	$\overline{\delta}$
XOM	5.434	0.816	4304	1.178	PG	7.633	0.869	3080	1.030
C	6.135	0.836	3832	1.246	BAC	7.633	0.869	3079	0.328
JPM	6.250	0.840	3743	0.999	PFE	7.874	0.873	2967	0.509
HAL	6.369	0.843	3690	0.872	WFC	7.936	0.874	2961	0.724
CVX	6.579	0.848	3553	0.850	MDT	8.196	0.878	2874	0.571
DIS	6.622	0.849	3543	0.846	CAT	8.196	0.878	2861	0.782
JNJ	6.666	0.850	3529	0.809	HD	8.264	0.879	2829	0.754
SLB	6.802	0.853	3454	0.613	COP	8.333	0.880	2828	0.494
DAL	6.993	0.857	3348	0.766	T	8.333	0.880	2809	0.492
WMT	7.042	0.858	3325	0.698	MCD	8.474	0.882	2756	0.673
VZ	7.407	0.865	3172	0.678	PM	8.547	0.883	2746	0.683
MRK	7.462	0.866	3146	0.685	CBS	8.620	0.884	2725	0.760
GM	7.462	0.866	3135	0.888	-	-	-	-	-

Table 5.5.1: Summary statistics for the top 25 most frequently traded assets in 2014 in the NYSE. We show the average duration (in seconds) between observations $\overline{\Delta t}$, the probability of missing values Λ, the average number of observations per day $\overline{m}$ and the average signal-to-noise ratio $\bar{\delta}$ as estimated by the LLSD.

Therefore, the score-driven update is:

$$f_{t+1,i} = f_{t,i} + A_i s_{t,i} \tag{5.5.1}$$

This choice allows to account for the intrinsic non-stationary behavior of variances during the trading day. For each i, we need to estimate the two diagonal elements of A_i. Thus, the number of parameters that are estimated in the first step is $2n$.

The parameters of the correlation model in eq. (5.2.11) are estimated under the following restrictions:

- $A_Z = aI_{k_Z}$
- $B_Z = I_{k_Z}$
- $\omega_Z = (0, \dots, 0)'$

where $k_Z = n(n-1)/2$. As it is common practice in the DCC model, we restrict the parameter space by requiring that the matrix A_Z is diagonal, with all the diagonal elements being equal to a scalar parameter a. This guarantees that the number of parameters to be estimated in the correlation model is independent on the number of time-series. As in the univariate model, the two restrictions on ω_Z and B_Z are due to the intrinsic non-stationary behavior of correlations during the trading day. Thus, there is a total of $2n + 1$ static parameters to be estimated. Initial values of time-varying parameters are set

equal to those found by estimating a local level model with constant parameters in the first 15 minutes of each trading day.

Before studying intraday patterns of volatilities and correlations, we show an interesting byproduct of the estimation procedure, i.e. the reconstruction of the efficient log-price using the Kalman filter. Once the LLSD is estimated, one can use the estimated parameters to obtain filtered and smoothed estimates of the efficient log-price. We show in figure (5.5.1) the observed log-price of Citigroup in a 5 minutes time window on 02/01/2014 together with the filtered estimate provided by the Kalman filter. Note that, even if observations are missing at a given point in time, the filter is able to use the multivariate information in order to optimally predict the state variable at that time.

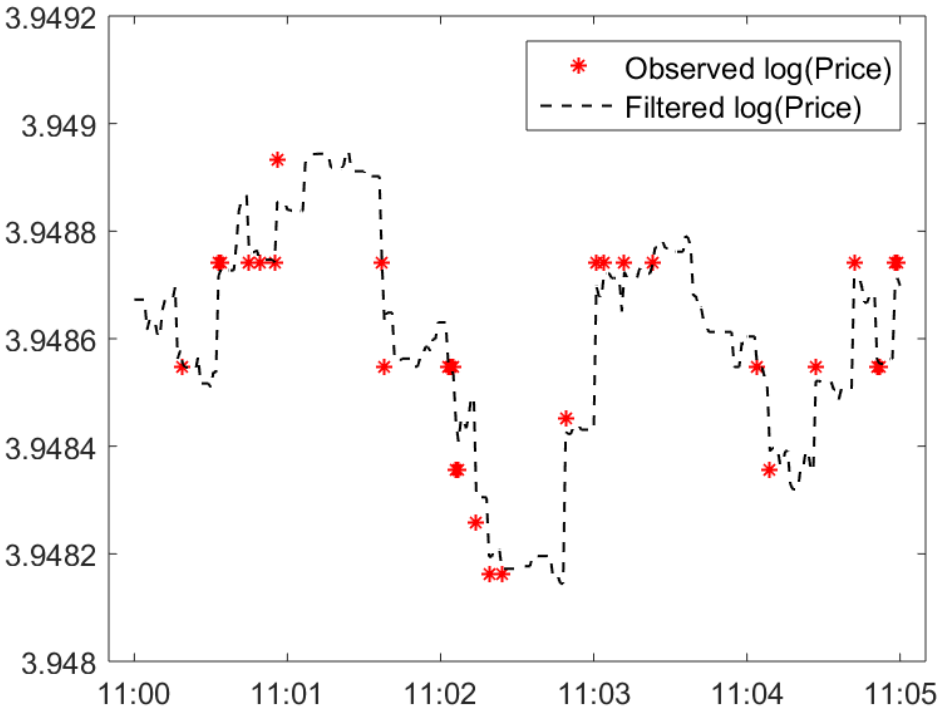


Figure 5.5.1: Reconstruction of the efficient log-price of Citigroup in a 5-minutes time-window on 02-01-2014 using the estimated LLSD.

5.5.3 Intraday patterns

The LLSD is estimated on the first $N = 120$ business days of 2014. As a result, estimates of the time-varying parameters D_t^j, H_t^j, Q_t^j and R_t^j are recovered, for $j = 1, \ldots, N$. In order to investigate the variation of intraday patterns among both different assets and different days, we compute the following

averages:

$$\bar{D}_t(i) = \frac{1}{N}\sum_{j=1}^{N} D_t^j(i,i), \quad \tilde{D}_t(j) = \frac{1}{n}\sum_{i=1}^{n} D_t^j(i,i)$$

$$\bar{H}_t(i) = \frac{1}{N}\sum_{j=1}^{N} H_t^j(i,i), \quad \tilde{H}_t(j) = \frac{1}{n}\sum_{i=1}^{n} H_t^j(i,i)$$

$$\bar{\delta}_t(i) = \frac{1}{N}\sum_{j=1}^{N} \delta_{i,t}^j, \quad \tilde{\delta}_t(j) = \frac{1}{n}\sum_{i=1}^{n} \delta_{i,t}^j$$

$$\bar{R}_t(p,q) = \frac{1}{N}\sum_{j=1}^{N} R_t^j(p,q), \quad \tilde{R}_t(j) = \frac{1}{n(n-1)}\sum_{p\neq q} R_t^j(p,q)$$

$$\bar{Q}_t(p,q) = \frac{1}{N}\sum_{j=1}^{N} Q_t^j(p,q), \quad \tilde{Q}_t(j) = \frac{1}{n(n-1)}\sum_{p\neq q} Q_t^j(p,q)$$

where $t = 1, \ldots, T$. For every asset $i = 1, \ldots, n$ or couple of assets (p, q), $p, q = 1, \ldots, n$ the quantities on the left are averages of the time-varying parameters over all the days of the sample while for every day j, $j = 1, \ldots, N$ the quantities on the right are averages of the time-varying parameters over all the assets (or couple of assets). For every $t = 1, \ldots, T$, we plot deciles of $\bar{D}_t$, $\sqrt{\bar{H}_t}$, $\bar{\delta}_t$, vech$(\bar{Q}_t)$, vech$(\bar{R}_t)$ in figure (5.5.2) and deciles of $\tilde{D}_t$, $\sqrt{\tilde{H}_t}$, $\tilde{\delta}_t$, $\tilde{Q}_t$, $\tilde{R}_t$ in figure (5.5.3). We also show in red the total average of the time-varying parameters over both assets and days.

The average volatility of the efficient price process exhibits the well known U-shape: volatility is huge at the beginning of the day and then it gradually declines until noon, except for a small raise at 10:00. After noon, we observe a steep raise at 14:00 and then a large increase at the end of the trading-day, especially in the last 30 seconds where volatility steeply increases. The average variance of the noise has two regimes: a steep decline from 9:30 to 10:00 and a slow decline from 10:00 until the end of the trading day. It resembles the typical intraday pattern of the bid-ask spread, suggesting that the LLSD is consistently setting apart the efficient log-price process from microstructure effects. The signal-to-noise ratio exhibits a more symmetric pattern: it is quite large at the beginning of the day ($\delta \sim 2$) and at the end, where it reaches similar values. Instead, it is small at noon, where $\delta \sim 0.5$. Note that these values are close to those used in the simulation study.

Average covariances follow a pattern which resembles the standard U-shape of volatility. However, their pattern is also driven by the non-trivial dynamics of the correlations which increase throughout the day. At the beginning of the day correlations are low, meaning that the dynamics of prices is largely affected by idiosyncratic risk. Then, we observe a steep increase of the correlations until 11:00 which is associated to the fast decline of the volatilities. After that, correlations still keep increasing, even at a slightly lower rate, until 15:45. At that time we observe a decline of all the correlations which

corresponds to the large increase of individual volatilities occurring during the last minutes of the trading day. Therefore, a market factor progressively gains explanatory power during the day, as shown by the intraday behavior of the first five eigenvalues of the correlation matrix $\bar{R}_t$ in figure (5.5.5). Note that the first eigenvalue increases during the day, following the same pattern of correlations, while the remaining eigenvalues decrease. At the end of the trading day the market factor accounts for $\sim 42\%$ of the total variance.

These results are in agreement with the empirical findings of Allez and Bouchaud (2011) who used standard sample correlations to study the intraday evolution of dependencies among asset prices. An increasing of correlations during the trading day was also found by Bibinger et al. (2014) using a spot covariance estimator.

Comparing figures (5.5.2) and (5.5.3), we note that the large increase of volatilities at the beginning and at the end of the trading day is common to all the assets and is observed every day. Also, the increase of correlations throughout the day and their sudden drop on the last 15 minutes is common to all the stocks and to all the days of our sample. There is, however, a remarkable variation of the level of correlations over different days, especially at the end of the trading day. For instance, the first decile of $\tilde{R}_t$ is ~ 0.3 while the last decile is ~ 0.5 at the closing time.

While these movements characterizing the intraday dynamics of volatilities and correlations are observed every day, there are other movements that are observed only in correspondence of some specific events. For instance, from figures (5.5.2), (5.5.3) we note that the increase of variances and covariances at 14:00 is common to all the stocks but is not observed every day. In order to understand the cause of this burst, we plot in figure (5.5.4) the estimate of the average D_t and $\text{vech}(R_t)$ obtained in correspondence of three different meetings of the Federal Open Market Committee (FOMC) in 2014, the first on 29-01-2014, the second on 19-03-2014 and the third on 30-04-2014.

Compared to the average intraday pattern, we observe significant deviations, especially at 14:00 and at 15:00. The time interval from 14:00 to 15:00 coincides with the press conference in which economic information is released by the central bank. We observe a huge increase of volatility at 14:00 followed by another peak at 15:00, especially in the second case. Moreover, in all the three days correlation decreases a bit immediately before the meeting and then it steeply increases until 15:45. It becomes significantly larger than the intraday pattern in the first two cases while in the third case correlation is significantly lower than the intraday pattern at the beginning of the day but it steeply increases after 14:00 and reaches the same level of the pattern in less than 30 minutes.

Thus, the LLSD is able to capture in real-time variations of volatilities and correlations occurring at a very small time-scale. This would not be possible using traditional methods since the presence of noise and asynchronous trading inevitably leads to neglect a lot of relevant information related to the

dynamics at smaller time-scales.

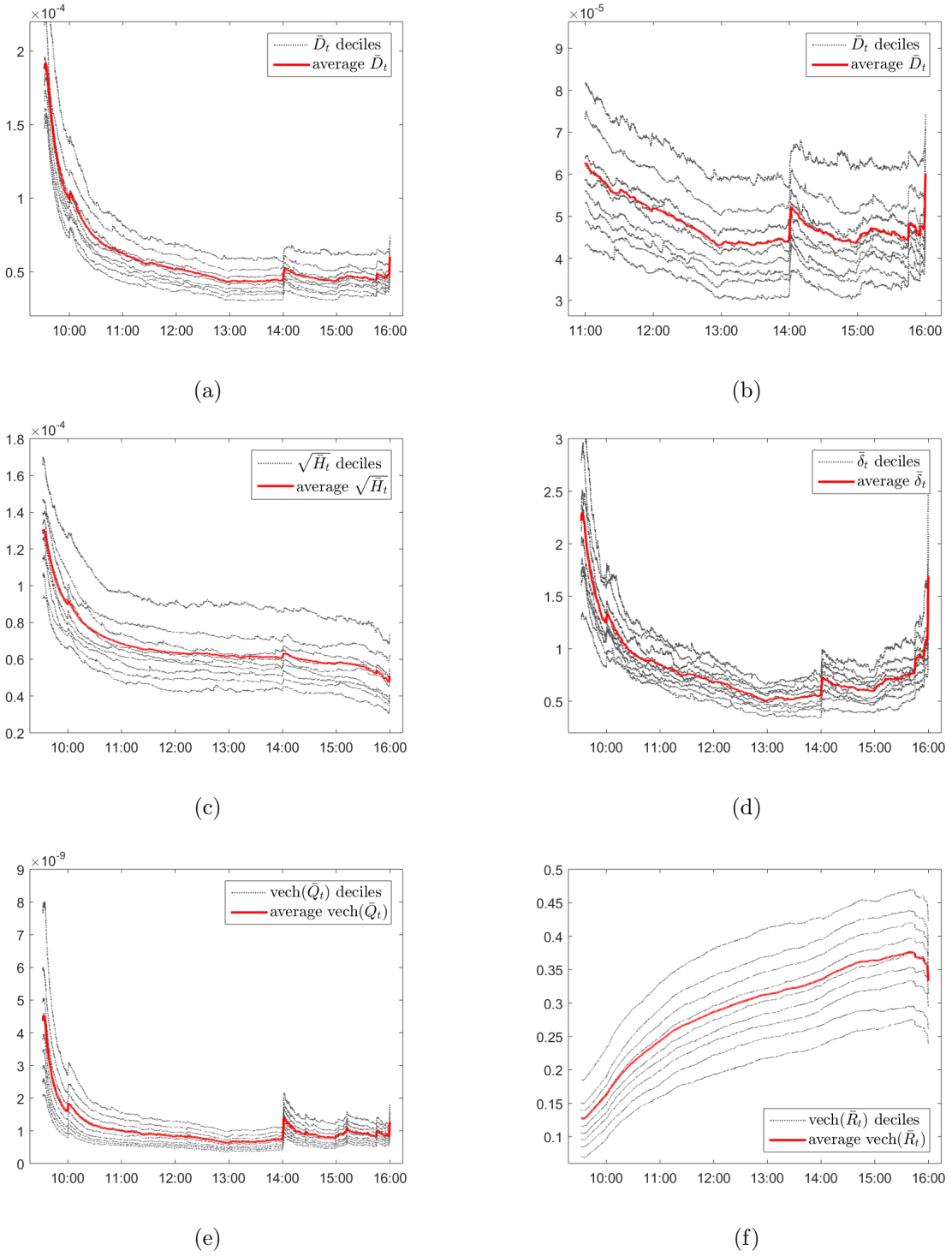


Figure 5.5.2: Intraday patterns of $\bar{D}_t$, $\sqrt{\bar{H}_t}$, $\bar{\delta}_t$, vech($\bar{Q}_t$) and vech($\bar{R}_t$) as defined in Section 5.5.3. Grey lines are deciles while the red line is the average over all the assets (or couple of assets). (b) is as (a) but zoomed from 11:00 to 16:00.

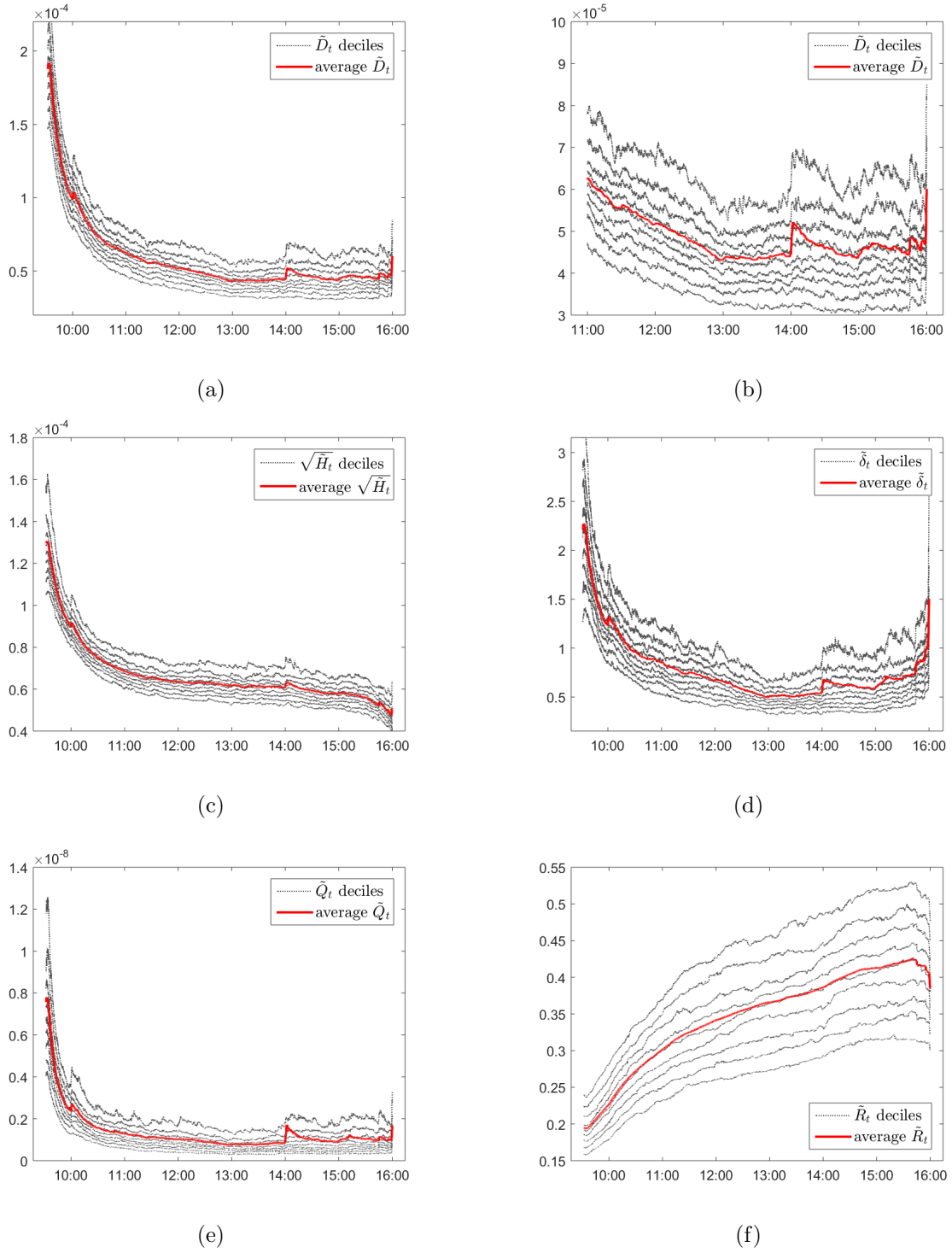
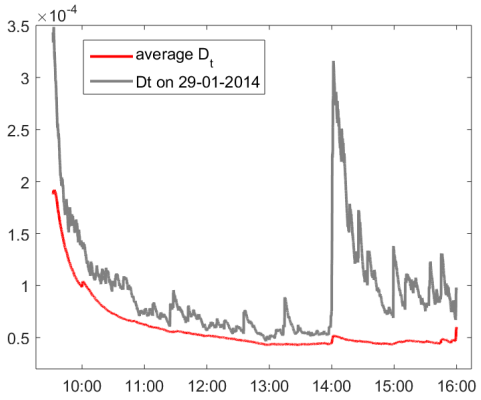
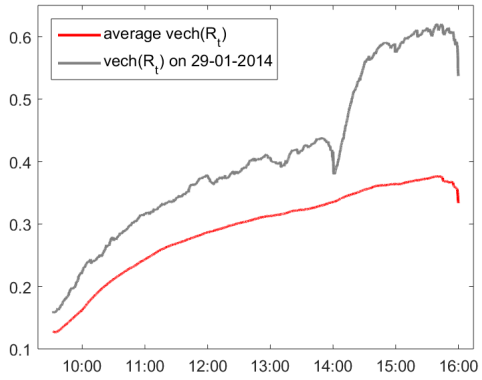


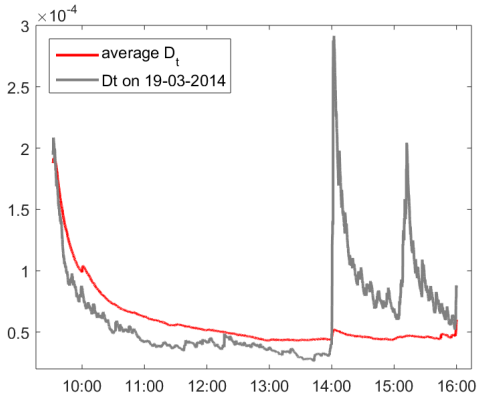
Figure 5.5.3: Intraday patterns of $\tilde{D}_t$, $\sqrt{\tilde{H}_t}$, $\tilde{\delta}_t$, $\tilde{Q}_t$ and $\tilde{R}_t$ as defined in Section 5.5.3. Grey lines are deciles while the red line is the average over all the days. (b) is as (a) but zoomed from 11:00 to 16:00.



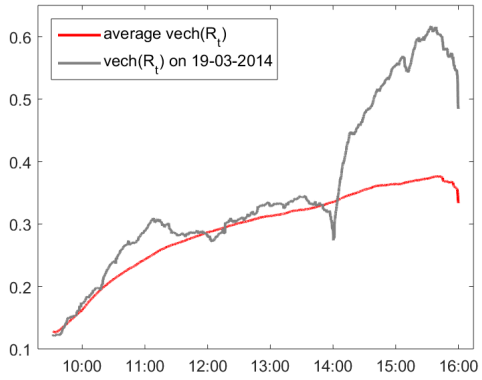
(a)



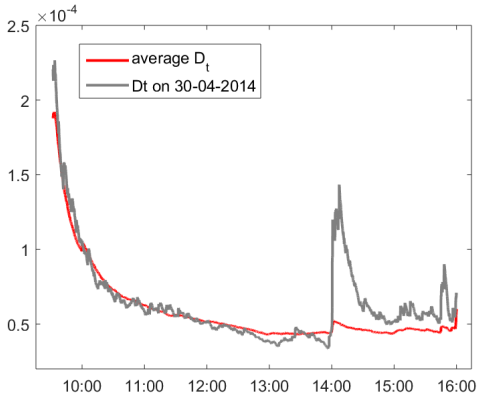
(b)



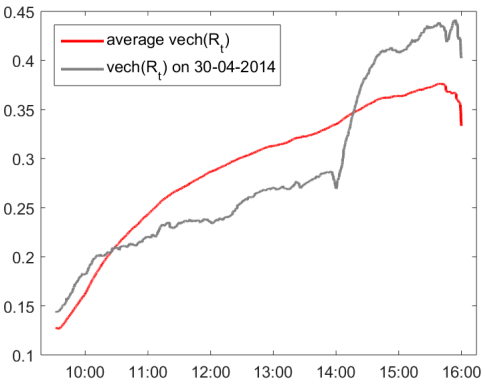
(c)



(d)

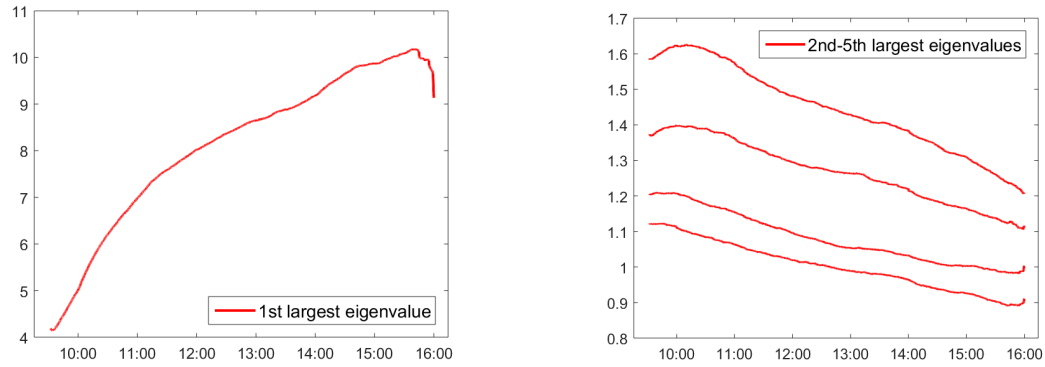


(e)



(f)

Figure 5.5.4: We show the intraday pattern of the average volatility and correlations (red line) and the estimate of D_t and vech(R_t) (averaged over all assets and couple of assets) obtained in correspondence of the first three meetings of the FOMC (black line) in 2014: on 29-01-2014 (a), (b), on 19-03-2014 (c), (d) and on 30-04-2014 (e), (f)



(a) (b)

Figure 5.5.5: Intraday patterns of the first eigenvalue (a) and of the following largest four eigenvalues (b) of the matrix $\bar{R}_t$.

5.5.4 Intraday portfolio selection

We assess covariance forecasts provided by the LLSD through an economically meaningful application based on portfolio optimization. As noted by Patton and Sheppard (2009), economic evaluation of volatility and correlation forecasts is an important tool for assessing the performance of the model. In our case, since the LLSD is designed to describe the dynamics of correlations at very small time-scales, we concentrate on short intraday investment horizons, of the order of few minutes.

Let us consider the problem of finding the Global Minimum Variance Portfolio (GMVP):

$$\min_{w_t} w_t' \Sigma_t w_t, \quad w_t' u = 1 \tag{5.5.2}$$

where u is a vector of ones of the same dimension as portfolio weights w_t. It is easy to show[3] that, if the GMVP solution w_t is constructed using the true conditional covariance matrix Σ_t, than the variance of a portfolio constructed using a different weight $\tilde{w}_t$ is larger. Engle and Colacito (2006) extended this result to a general mean-variance problem and showed that, if expected returns are constant, the portfolio constructed using the true covariance matrix has a variance which is lower than the one obtained through any covariance forecast, regardless of the assumed value of expected returns.

Based on these results, it is therefore possible to assess the forecasts of two different conditional correlation models by comparing the variances of optimal portfolios constructed using the two models. In this application we focus on the GMVP and denote by $\pi_t^j = w_t^{j\prime} r_t$, $j = 1, 2$, the portfolio realized return computed with model j. Then, we consider differences between the squared returns of the two portfolios:

$$u_t = (\pi_t^1)^2 - (\pi_t^2)^2 \tag{5.5.3}$$

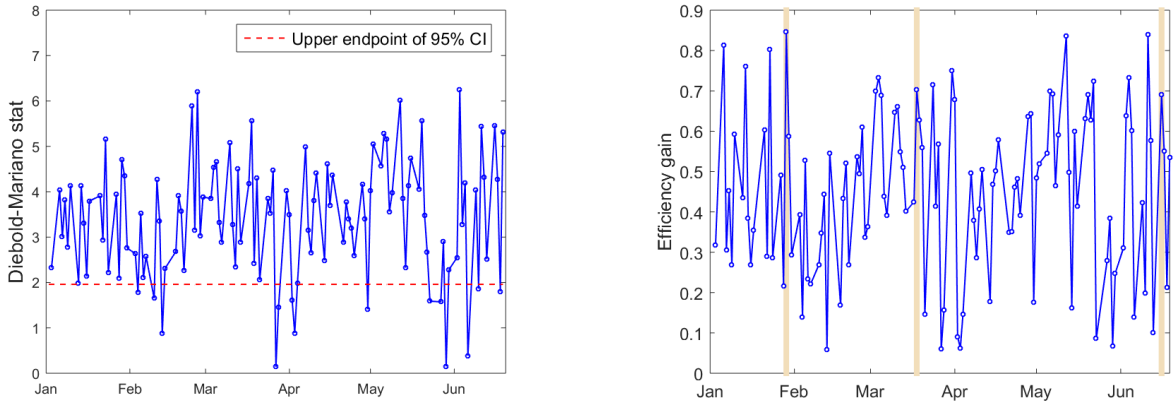
[3] see e.g. Patton and Sheppard 2009.

If the mean of u_t is significantly positive (negative), model 2 (1) provides better forecasts. Using squared returns as a proxy of the portfolio variance is based on the assumption that expected returns are zero. This is not restrictive since we are considering high-frequency returns computed on very small time-scales.

We consider the first $n = 2$ assets[4] in table (5.5.1) and find the GMVP using the forecasts provided by the LLSD and the DCC. The DCC model is estimated on a rolling window of 24 hours, using a regularly spaced grid of 4-min returns. This is the highest possible frequency at which it is possible to aggregate prices to have at least an observation per grid point. The LLSD is estimated on the same window but using all the observations. The choice of n is conservative since we verified that, as n increases, the relative performance of the LLSD over the DCC increases as well. The rolling window is shifted by 4 minutes each time and new GMVP's are constructed on a time-horizon $T = 4$ minutes. As such, the forecasts provided by the two models are non-overlapping.

Using the Diebold-Mariano (DM) test statistic, we test the null hypothesis that $u_t = (\pi_t^1)^{\mathrm{DCC}} - (\pi_t^2)^{\mathrm{LLSD}}$ has mean zero. Figure (5.5.6) shows the DM test statistic computed for each day of the sample. We also show the average of efficiency gains over all the portfolios constructed on a given day. The efficiency gain is defined here as the relative increase of the DCC portfolio standard deviation over the LLSD portfolio. Figure (5.5.6a) shows that the null hypothesis that the two portfolios have the same variance is rejected at the 95% c.l. for most of the days of the sample. In particular, we find that u_t is significantly positive in 104 days out of 120. The DM test statistic computed on the whole sample is $DM = 17.2729$, indicating strong evidence of rejection of the null hypothesis. Positive differences among portfolio variances translate into larger efficiency gains for LLSD portfolios, as shown in figure (5.5.6b). The average of efficiency gain over the whole sample is 45%, with peaks of more than 80% observed on certain days. For instance, the efficiency gain is $\sim 85\%$ on 29-01-2014, corresponding to a FOMC announcement.

[4] We obtained similar results with other couples of assets.



(a) (b)

Figure 5.5.6: (a) Daily DM test statistic computed on differences of intraday portfolios realized variance. (b) Daily average efficiency gains of LLSD portfolios relative to DCC portfolios. Vertical lines denote FOMC days.

Chapter 6

Realized volatility: errors and nonlinearity

Almost all results in this chapter previously appeared in Buccheri et al. (2018a).

Estimating and forecasting the volatility of financial markets is a prominent topic in theoretical and applied finance. Andersen and Bollerslev (1997) were the first to advocate the use of realized variance computed from high-frequency data as an accurate proxy of integrated variance. Linear reduced-form specifications for time series of realized variance are widely recognized today as being extremely powerful in predicting financial volatility (Andersen et al. 2003). However, the vast majority of these dynamic specifications ignore two key aspects: (i) realized variance is a noisy estimate of the true integrated variance and (ii) volatility dynamics are highly nonlinear. In particular, the HAR model of Corsi (2009), one of the most popular dynamic specification for realized volatility, suffers from three main forms of misspecification when estimated on realized variance series: (a) biased OLS estimates (Bollerslev et al. 2016a); (b) autocorrelated and highly heteroskedastic residuals (Corsi et al. 2008); (c) time-varying OLS coefficients (Chen et al. 2010). While (a) is related to measurement errors, (b) and (c) are due to nonlinear dependencies, i.e. deviations of the true underlying volatility dynamics from the linear HAR specification.

In the following, we aim to disentangle and quantify the impact of measurement errors and nonlinearity on volatility forecasts provided by the HAR. The strategy adopted here is a step by step approach. We first devise different HAR extensions aimed to account separately for each effect. Then, we combine the different approaches in a single model that takes into account both effects. In doing so, we can quantify the forecast gains resulting from removing each form of misspecification. The specific choice of the HAR model is not restrictive, as the same approach can be applied to any linear specification for realized variance.

As a first step, we write the HAR in a linear state-space representation where the time-varying variance of the measurement error is driven by realized quarticity, as prescribed by the asymptotic theory of realized variance (Barndorff-Nielsen and Shephard 2002). The Kalman filter allows to easily estimate bias-corrected HAR parameters and incorporates the effect of measurement errors through a time-varying Kalman gain. We name this model HAR-Kalman (HARK). The HARK corrects HAR forecasts based on the uncertainty with which volatility is measured. It provides more conservative forecasts when current volatility estimates are noisy and, in contrast, generates more responsive forecasts when volatility is measured with a good accuracy.

Compared to the HAR, the $\text{HAR}_{\log}$, i.e. the HAR estimated on $\log(RV_t)$ series, provides a better dynamic specification for realized variance (Corsi et al. 2008). However, even on $\log(RV_t)$ series, residuals feature significant heteroskedasticity. In addition, OLS coefficients tend to vary significantly when estimating the model on different time windows. This means that volatility forecasts can vary substantially when changing the estimation window. These empirical findings underline deviations of the volatility process from the linear HAR specification and can be viewed as an evidence of nonlinear dynamics. Blasques et al. (2014c) showed that general *nonlinear* autoregressive models[1] can be equivalently represented as *linear* autoregressive models with time-varying parameters. They proved that this formulation is optimal from an information theoretic perspective, provided that parameters are driven by the score of the conditional likelihood. Motivated by this result, we devise an $\text{HAR}_{\log}$ extension that features heteroskedastic errors and time-varying coefficients. In particular, we adopt an observation-driven approach where parameters evolve based on the score of the conditional likelihood. The resulting model, the Score-HAR (SHAR), is nonlinear in nature and features iid standardized residuals when estimated on real data.

As a final step, we combine together the HARK and the SHAR to obtain a more general model, the SHARK, that accounts for all forms of misspecification. In particular, we let the parameters of the HARK evolve through an observation-driven update scheme. The resulting model is conditionally Gaussian and can be handled by the Kalman filter, as underlined in Section 2.4. As the HARK, the SHARK can handle measurement errors but also accounts for heteroskedasticity and time-varying parameters. As the SHAR, the SHARK accounts for nonlinear dependencies but provides more responsive time-varying parameters, as static parameters are not bias attenuated by measurement errors.

Our empirical analysis, conducted on both index future and individual stock data, provides strong evidence that the effects captured by the HARK, SHAR and SHARK are relevant for volatility forecasting. As a matter of fact, the SHARK is always included in the model confidence set of Hansen et al. (2011). In some cases the test has not enough power to exclude the HARK and the SHAR while the

[1] Nonlinear autoregressive models have the general form $y_t = \phi(Y_{t-1}, \Theta) + u_t$, where $\phi(\cdot)$ is a nonlinear function of past observations $Y_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$, Θ is a set of parameters and u_t is a zero-mean sequence of independent innovations.

$\text{HAR}_{\log}$ and other competing models feature much lower p-values and are excluded in the vast majority of cases. Interestingly, the out-of-sample performance of SHAR and SHARK models does not change significantly if one estimates static parameters on a single window or on a rolling window. Indeed, while HAR parameters vary significantly when estimating the model on a moving window, SHAR and SHARK models have built-in time-varying coefficients which capture the variation of volatility persistence over time. Comparing the relative gains among our models, we obtain that measurement errors are important at small and intermediate sampling frequencies. The corresponding forecast gains slightly increase with the forecast horizon. The impact of heteroskedasticity is always relevant, especially at longer forecast horizons. Finally, time-varying parameters provide statistically significant improvements that are independent on the sampling frequency and slightly increase with the forecast horizon.

The problem of taking into account the effect of measurement errors on realized variance forecasts was recently tackled by Bollerslev et al. (2016a), who devised the HARQ model. The main idea is to augment the HAR with a term proportional to lagged RV_t's that also depends on realized quarticity. This new term adjusts HAR forecasts based on the current level of uncertainty on realized variance. Our approach differs in several aspects. First, assuming an HAR plus noise as a DGP for the realized variance, the Kalman filter provides MVLUE estimates (see e.g. Durbin and Koopman 2012) and corrections due to measurement errors are independent on the level of heteroskedasticity of the noise. Second, forecasting gains provided by the HARQ may in principle be imputable to potential nonlinear dependencies captured by the quarticity term, as recently pointed out by Cipollini et al. (2017), or result from a mixture of measurement errors and nonlinearity. Our approach, being based on the Kalman filter, allows to quantify the neat effect of estimation errors after disentangling it from nonlinearity.

Our work is related to that of Asai et al. (2012), who employed a state-space representation to account for measurement errors. However, the error variance in their work is homoskedastic and is not related to quarticity. Bekierman and Manner (2016) proposed a modification of the HARQ model by allowing the coefficients of the HAR to be driven by a latent process. Nevertheless, as acknowledged by the authors, the latter may also capture other sources of temporal variations. Applications of the Kalman filter to account for measurement errors are widespread in the literature. In the field of high-frequency financial econometrics, some examples are given by Barndorff-Nielsen and Shephard (2002), Shephard and Xiu (2016), Corsi et al. (2015). Examples of nonlinear models for realized variance are the class of MEM models developed by Engle (2002c), Engle and Gallo (2006), Cipollini et al. (2017).

6.1 Theoretical framework

6.1.1 Asymptotic theory of realized variance estimation

Let $\mathcal{P}_s$ denote the asset price at time s. We assume that the log-price $X_s = \log(\mathcal{P}_s)$ evolves according to a Brownian semimartingale process:

$$dX_s = \mu_s ds + \sigma_s dW_s \tag{6.1.1}$$

where μ_s and σ_s are drift and instantaneous volatility processes satisfying the usual assumptions and W_s is a Wiener process. The day t integrated variance of X_s is defined as:

$$IV_t = \int_{t-1}^{t} \sigma^2(s) ds \tag{6.1.2}$$

where the unit time interval corresponds to one trading day. Note that IV_t turns out to be equal to the quadratic variation of X_s computed on $[t-1, t]$ (see e.g. Protter 1992).

The day t realized variance is defined as:

$$RV_t = \sum_{i=1}^{M} r_{i,t}^2 \tag{6.1.3}$$

where the intraday returns $r_{i,t} = X_{t-1+i\Delta} - X_{t-1+(i-1)\Delta}$, $i = 1, \ldots, M$ are computed on M intraday time intervals of length $\Delta = 1/M$. The econometric theory of Barndorff-Nielsen and Shephard (2002) suggests that, as the intraday period $\Delta \to 0$, the estimation error is mixed normal distributed:

$$RV_t = IV_t + \epsilon_t, \quad \epsilon_t \sim \text{MN}(0, 2\Delta IQ_t) \tag{6.1.4}$$

where $IQ_t = \int_{t-1}^{t} \sigma^4(s) ds$ is the integrated quarticity of the underlying semimartingale process. IQ_t can consistently be estimated using the realized quarticity estimator:

$$RQ_t = \frac{M}{3} \sum_{i=1}^{M} r_{i,t}^4 \tag{6.1.5}$$

For later convenience, we also report the asymptotic distribution of $\log(RV_t)$. Using the delta method, it is immediate to see that, as $\Delta \to 0$:

$$\log(RV_t) = \log(IV_t) + \xi_t, \quad \xi_t \sim \text{MN}\left(0, 2\Delta \frac{IQ_t}{IV_t^2}\right) \tag{6.1.6}$$

A consistent estimator V_t of the variance of ξ_t can be obtained by replacing IV_t and IQ_t with their consistent estimators (6.1.3), (6.1.5), namely:

$$V_t = \frac{2}{3} \frac{\sum_{i=1}^{M} r_{i,t}^4}{(\sum_{i=1}^{M} r_{i,t}^2)^2} \tag{6.1.7}$$

As pointed out by Barndorff-Nielsen and Shephard (2002), this approximation is quite accurate even at moderate values of M, in which case approximation (6.1.4) is less reliable. Note that jumps and microstructure effects have not been considered. To account for them, one needs to replace RV_t and RQ_t with robust estimators. However, our general approach remains unchanged.

6.1.2 Reduced-form models for volatility estimation and forecasting

If one neglects the measurement error term in eq. (6.1.4), then $RV_t = IV_t$ and the integrated variance is observable. This assumption is at the basis of reduced-form specifications that have been employed in last years to model and forecast volatility using time series of realized variance.

One of the most popular linear specification is the approximate long-memory HAR model of Corsi (2009). The HAR reads:

$$RV_{t+1} = \beta_0 + \beta_1 RV_t + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-6|t-22} + \eta_{t+1} \tag{6.1.8}$$

where $\eta_{t+1} \sim \text{NID}(0, q)$ and $RV_{t_1|t_2}$ denotes the average of daily RV's from day t_1 to day t_2. Even though the HAR is not a long-memory process, the aggregation of volatilities at short and long time-scales leads to a slowly decaying autocorrelation function that closely resembles the one observed on real financial data. The model can be simply estimated by OLS and provides out-of-sample forecasts which have been proved to be comparable to those of long-memory ARFIMA-RV models (Corsi 2009). Corsi and Renò (2009) extended the HAR to include the effect of leverage and jumps. Given the simplicity and effectiveness of the HAR, we will use it as a basis for our modeling framework and empirical work. However, any other linear specification for RV_t is susceptible of treatment within our approach.

6.1.3 Measurement errors

The main effect of the measurement error term in eq. (6.1.4) is the well-known attenuation bias of OLS coefficients in presence of latent regressors (see e.g. Wansbeek and Meijer 2000). As a consequence, the estimated model features less persistence and volatility forecasts are less accurate.

Recently, there was an increased interest in examining the effect of measurement errors on HAR volatility forecasts. Bollerslev et al. (2016a) suggested to augment the HAR with a term depending on quarticity. The latter underrates RV forecasts in case estimation errors are large and in turn generates more responsive forecasts when errors are small. Their HARQ model reads:

$$RV_{t+1} = \beta_0 + (\beta_1 + \beta_{1Q} RQ_t^{1/2}) RV_t + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-6|t-22} + \eta_{t+1} \tag{6.1.9}$$

If $\beta_{1Q} < 0$, the term $\beta_{1Q} RQ_t^{1/2}$ corrects RV_{t+1} based on the degree of uncertainty with which RV_t is measured. Similar correction terms can also be included in the remaining terms.

In order to show the effect of measurement errors on OLS coefficients and the correction provided by the HARQ, we simulate IV_t using an HAR specification and contaminate the simulated observations with a white noise term:

$$RV_t = IV_t + \epsilon_t \tag{6.1.10}$$

$$IV_{t+1} = \beta_0 + \beta_1 IV_t + \beta_2 IV_{t-1|t-5} + \beta_3 IV_{t-6|t-22} + \eta_{t+1} \tag{6.1.11}$$

where $\epsilon_t \sim \text{NID}(0, H_t)$ and $\eta_{t+1} \sim \text{NID}(0, q)$ are uncorrelated measurement and model disturbances. Observations are simulated by first sampling IV_t, $t = 1, \ldots, 22$ from a lognormal distribution with mean and variance given by the unconditional mean and variance of process (6.1.11) and then iteratively applying eq. (6.1.10) and (6.1.11). Since the HARQ relies on having an heteroskedastic measurement error, we model dynamically H_t through a GARCH process:

$$z_t = \sqrt{H_t}\zeta_t \tag{6.1.12}$$

$$H_{t+1} = \omega + \gamma z_t^2 + \rho H_t \tag{6.1.13}$$

where $\zeta_t \sim \text{N}(0, 1)$. We define the signal-to-noise ratio as $\delta = q(1 - \gamma - \rho)/\omega$ and in this example we choose $q = 0.1$, $\gamma = 0.01$, $\rho = 0.95$ and $\omega = q(1 - \gamma - \rho)$, such that $\delta = 1$. The remaining parameters are set as: $\beta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 0.2$, $\beta_3 = 0.1$.

After generating $T = 1000$ observations for $N = 1000$ Monte-Carlo repetitions, the simulated RV_t and H_t are used to estimate[2] the HAR and HARQ models. In figure (6.1.1) we plot kernel density estimates of $z_i = \hat{\theta}_i - \theta_i$, where θ_i, $\hat{\theta}_i$, $i = 1, \ldots, 4$, denote the parameters of the HAR specification in eq. (6.1.11) and their estimates obtained using the HAR and HARQ models. In both cases the coefficient β_1 of RV_t exhibits a strong negative bias. However, the bias is lower in the HARQ which in turn is more persistent than the standard HAR model. As expected, the other two coefficients are less affected since measurement errors decrease when averaging out lagged RV_t 's.

Note that the estimated HARQ coefficient β_{1Q} depends on the level of heteroskedasticity of ϵ_t. If H_t is constant, then β_{1Q} is not identifiable, even in presence of large measurement error variance H. If H_t is time-varying, larger variations are related to larger estimates of β_1 and thus better forecasts. Finally, note also that specifications (6.1.8) and (6.1.9) do not guarantee RV_t to be positive.

6.1.4 The HARK model

The Kalman filter provides a natural way of recovering consistent and unbiased estimates of the HAR coefficients when observations of IV_t are contaminated by noise. Let $\mathcal{F}_t$ be the σ-field generated by RV_t and RQ_t. Let n_w be the number of daily RV_t's used in the computation of the second term in eq. (6.1.11) and n_m the number of daily RV_t's used in the computation of the third term. For instance, in the model in eq. (6.1.11), $n_w = 5$ and $n_m = 17$. Let also define $n = n_w + n_m + 1$. We consider the following linear state-space representation:

$$RV_t = Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, H_t) \tag{6.1.14}$$

$$\alpha_{t+1} = c + T\alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q) \tag{6.1.15}$$

[2] On real data H_t is not available and is replaced by the RQ_t estimator in eq. (6.1.5).

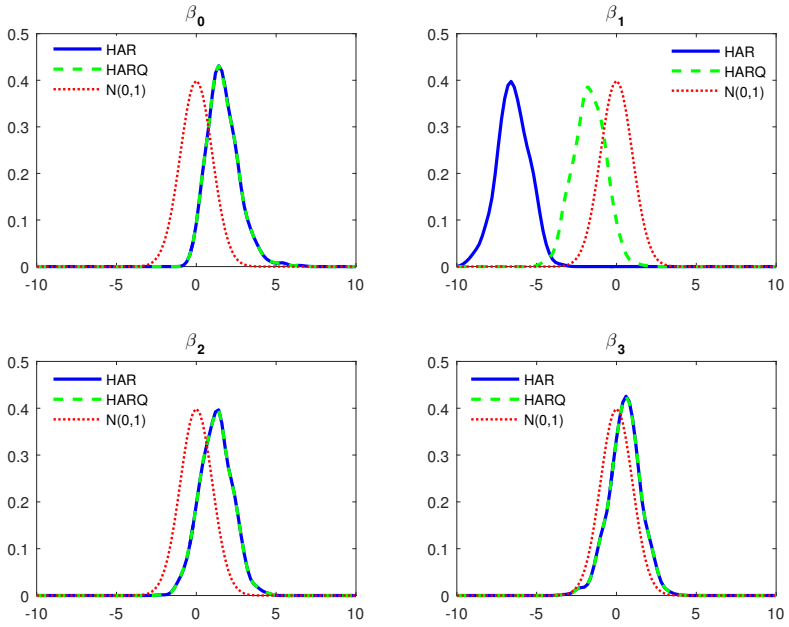


Figure 6.1.1: Kernel density estimates of standardized pivotal HAR and HARQ statistics estimated on $N = 1000$ Monte-Carlo simulations of $T = 1000$ observations of model (6.1.10), (6.1.11)

where we have introduced the $n \times 1$ state vector and the $n \times n$ transition matrix:

$$\alpha_{t+1} = \begin{pmatrix} IV_{t+1} \\ IV_t \\ \vdots \\ IV_{t-n+2} \end{pmatrix}, \quad T = \begin{pmatrix} \beta_1, & \frac{1}{n_w}\beta_2 & \overbrace{\cdots}^{n_w \ terms} & \frac{1}{n_w}\beta_2, & \frac{1}{n_m}\beta_3 & \overbrace{\cdots}^{n_m \ terms} & \frac{1}{n_m}\beta_3 \\ 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & \ddots & & & & \vdots & \vdots \\ \vdots & & \ddots & & & \vdots & \vdots \\ \vdots & & & \ddots & & \vdots & \vdots \\ \vdots & & & & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 1 & 0 \end{pmatrix} \tag{6.1.16}$$

together with the $n \times 1$ vector of constants and the $n \times n$ covariance matrix:

$$c = \begin{pmatrix} \beta_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad Q = \begin{pmatrix} q & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \tag{6.1.17}$$

The n-dimensional vector $Z = (1, 0 \dots, 0)$ selects the first element of α_{t+1}. The variance of the measurement error H_t is assumed to be measurable with respect to the information set $\mathcal{F}_t$.

Model (6.1.14), (6.1.15) is a linear Gaussian state-space representation. The Kalman filter recursions (2.2.6)-(2.2.11) allow to recursively compute conditional forecasts of the mean $a_{t+1} = \mathrm{E}[\alpha_{t+1}|\mathcal{F}_t]$ and the

variance $P_{t+1} = \text{Var}[\alpha_{t+1}|\mathcal{F}_t]$ of the state vector α_t. The log-likelihood can be computed as:

$$\log L(RV_1, \dots, RV_T|\Phi) = -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \tag{6.1.18}$$

where $v_t = RV_t - Za_t$ is the prediction error, $F_t = ZP_tZ' + H_t$ is the corresponding covariance matrix and $\Phi = \{\beta_0, \beta_1, \beta_2, \beta_3, q\}$ is the set of parameters of the model. Note that, compared to the Kalman filter recursions (2.2.6)-(2.2.11), the measurement error variance H_t is time-varying and is assumed to be known at time t. Parameters are estimated by maximizing the log-likelihood through a quasi-Newton algorithm:

$$\hat{\Phi} = \underset{\Phi}{\text{argmax}} \log L(RV_1, \dots, RV_T|\Phi) \tag{6.1.19}$$

In order to select a proxy for the variance H_t of the measurement error term in eq. (6.1.14), we rely on the asymptotic theory of Barndorff-Nielsen and Shephard (2002) in eq. (6.1.4) and use the realized quarticity estimator in eq. (6.1.5). In Section 6.4 we will examine more flexible specifications for H_t to test deviations from this assumption.

Measurement errors are taken into account in the Kalman filter recursions through the Kalman gain K_t. If the error variance H_t is large, meaning that RV_t is estimated with large uncertainty, the predicted variance is penalized by a small gain. In contrast, if H_t is small, meaning that RV_t is estimated with great accuracy, the predicted variance will be more responsive to RV_t. We name this model HAR-Kalman (HARK).

Figure (6.1.2) shows kernel density estimates of the pivotal HARK statistics obtained in the same simulation settings of Section 6.1.3. As done with the HARQ in previous experiment, we use the true measurement error variance H_t to estimate the HARK. The estimates provided by the HARK are unbiased and distributed according to a normal, as predicted by the econometric theory on linear state-space models. Estimation of the model is very fast since system matrices are sparse.

It is important to examine whether improvements in parameter estimates translate into better out-of-sample forecasts. For different values of the signal-to-noise ratio δ, we generate $N = 250$ Monte-Carlo realizations of model (6.1.10), (6.1.11) with $T = 2000$ observations. The latter 1000 observations are predicted using a moving window of 1000 observations. In figure (6.1.3) we show out-of-sample MSE of HARQ and HARK relative to the MSE of the HAR, for each signal-to-noise ratio δ. As expected, there are large gains when δ is small, as observations of the underlying volatility process are mainly dominated by noise. However, the MSE provided by the HARK is always lower than the one of the HARQ. This is true for all the choices of the parameters in model (6.1.10), (6.1.11) and in the GARCH model that we used to generate H_t. This is not surprising, as the HARQ model is misspecified on the DGP (6.1.10), (6.1.11) while the HARK provides the right specification.

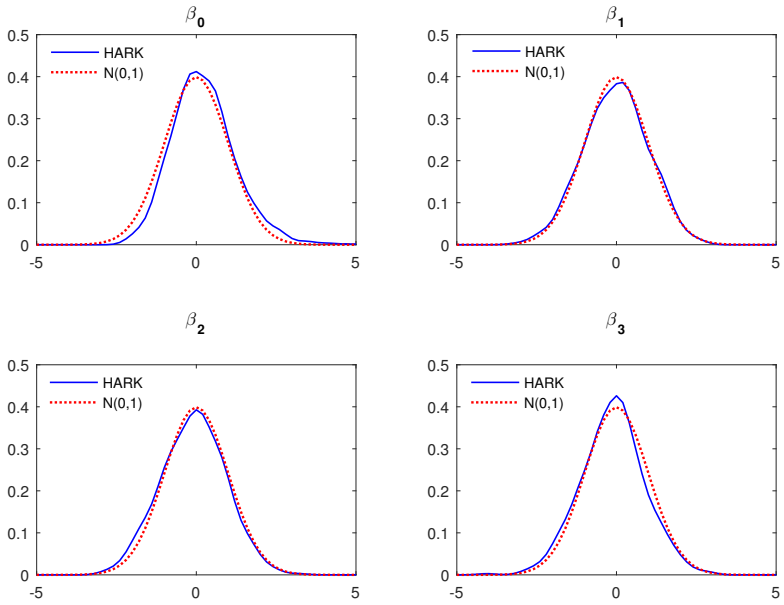


Figure 6.1.2: Kernel density estimates of standardized pivotal HARK statistics estimated on $N = 1000$ Monte-Carlo simulations of $T = 1000$ observations of model (6.1.10), (6.1.11)

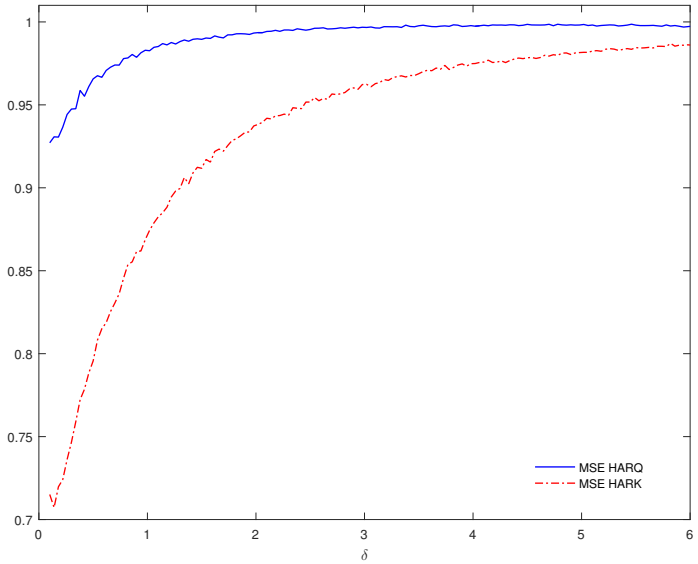


Figure 6.1.3: Out-of-sample MSE of HARQ and HARK models for different values of signal-to-noise ratio δ. Each MSE is divided by the MSE provided by the HAR model.

6.1.5 Nonlinear dynamics

The DGP (6.1.10), (6.1.11) used in the previous analysis takes into account the effect of measurement errors when forecasting IV_t but is characterized by a linear transition equation. In order to understand whether a linear specification fits well realized variance series, we consider high-frequency futures prices of the S&P 500 market index and compute daily series of realized variance by summing squares of 5-min returns. The resulting sample is shown in figure (6.1.4) and includes 4259 days, from 03-01-1995 to 21-06-2013.

It is known that the $\text{HAR}_{\log}$, i.e. the HAR estimated on $log(RV_t)$ series, features less heteroskedastic residuals (see e.g. Corsi et al. 2008). This is evident from figure (6.1.5), which shows the OLS residuals of HAR and $\text{HAR}_{\log}$, and from figure (6.1.6), which shows the sample autocorrelation function of residuals and squared residuals of the two models. Compared to the HAR, the residuals of the $\text{HAR}_{\log}$ are uncorrelated, while squared residuals show weaker serial correlation.

As a consequence, being the HARK based on a linear state-space representation, it is convenient to formulate it on $\log RV_t$ series rather then on RV_t. The only modification to the formulation in Section 6.1.4 is that we now rely on the asymptotic distribution of $\log RV_t$ and, consequently, the measurement error variance H_t is computed using eq. (6.1.6) and (6.1.7), i.e. $H_t = V_t$. It is useful to investigate whether the weaker misspecification of the $\text{HAR}_{\log}$ model translates into superior out-of-sample forecasts. Indeed, this is the case, as it will be verified in Section 6.2 using simulated data and in Section 6.3 using real data.

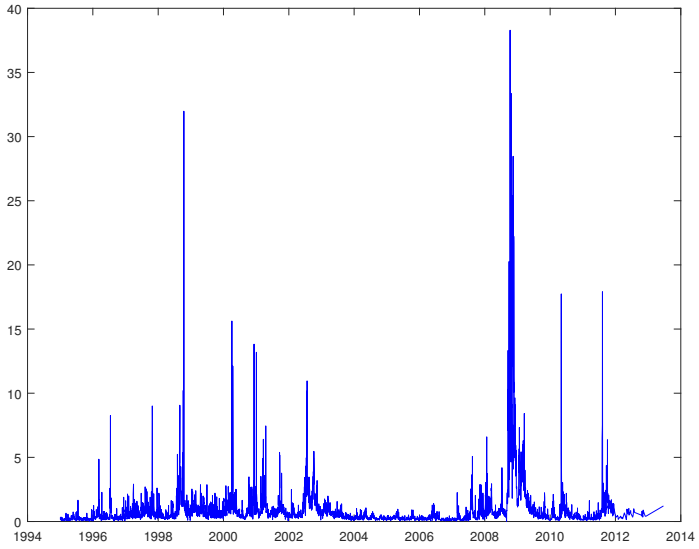


Figure 6.1.4: Time series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013.

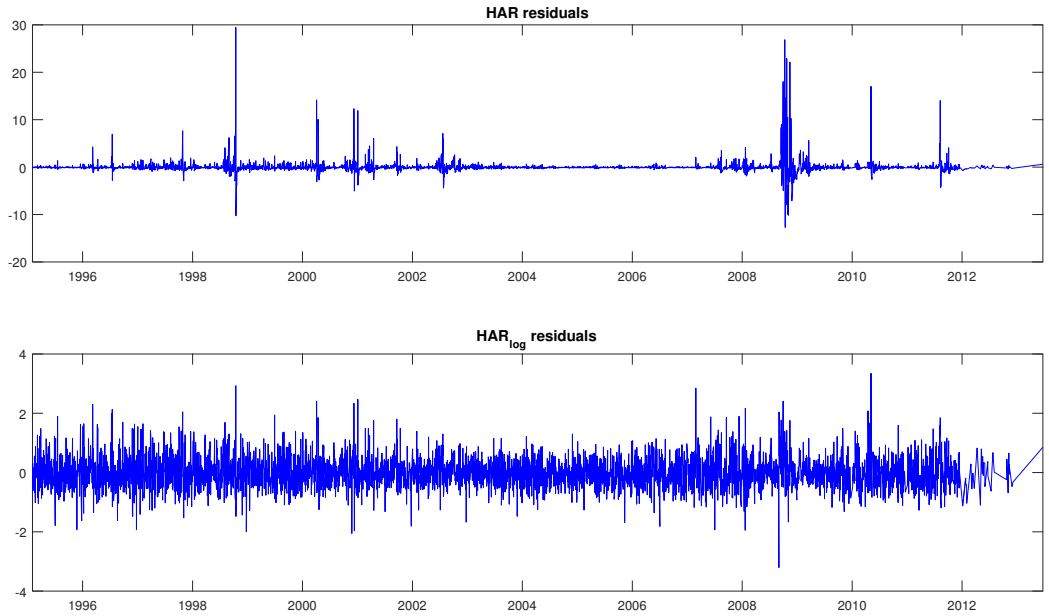


Figure 6.1.5: Residuals of HAR and HAR_{log} models estimated on the series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013.

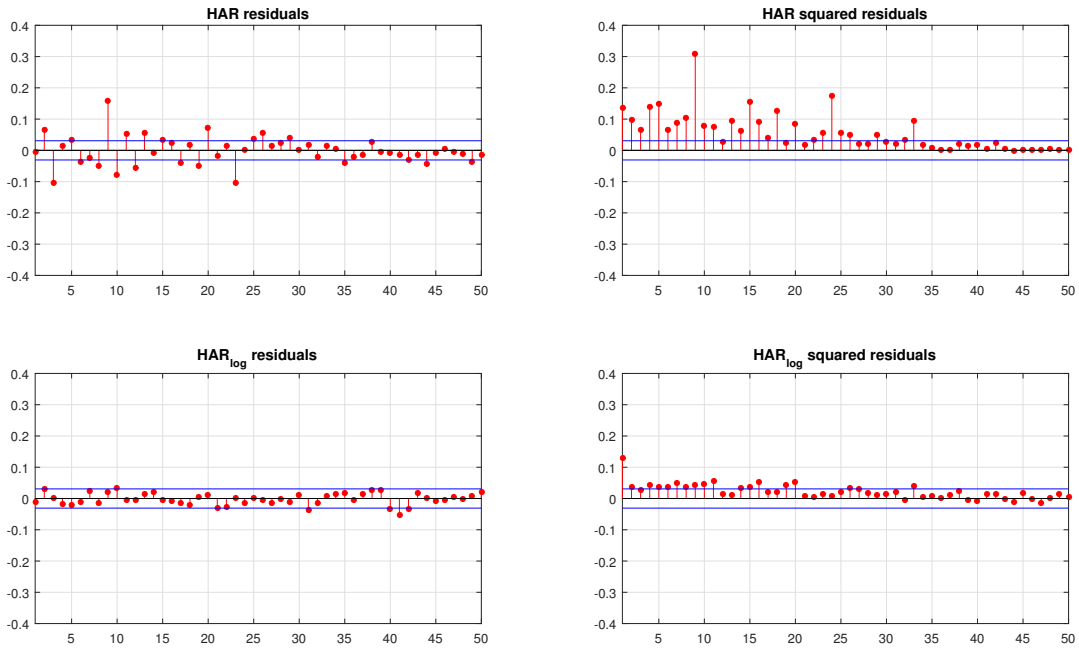


Figure 6.1.6: Sample autocorrelation function of HAR and HAR_{log} residuals and squared residuals estimated on the series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013.

6.1.6 The SHAR model

We can write the $\text{HAR}_{\log}$ model as:

$$RV^l_{t+1} = \beta_0 + \beta_1 RV^l_t + \beta_2 RV^l_{t-1|t-n_w} + \beta_3 RV^l_{t-n_w-1|t-n_w-n_m} + \eta_{t+1}, \quad \eta_{t+1} \sim \text{NID}(0, q) \tag{6.1.20}$$

where $RV^l_t = \log(RV_t)$ and $RV^l_{t1|t2}$ is built as in the HAR but is computed on $\log(RV_t)$ series. The heteroskedasticity featured by the $\text{HAR}_{\log}$ suggests to remove this residual misspecification by modeling dynamically the variance q of η_t. The $\text{HAR}_{\log}$ coefficients are also subject to time variations. Figure (6.1.7) shows the dynamics of $\beta_0, \beta_1, \beta_2, \beta_3$ obtained by estimating the $\text{HAR}_{\log}$ on daily realized variances of S&P500 future prices using a moving window of 2000 observations. Variations over time of the estimated coefficients are therefore relevant and suggest that volatility forecasts might be improved by modeling dynamically the $\text{HAR}_{\log}$ coefficients.

As shown by Blasques et al. (2014c), autoregressive models with time-varying coefficients can be employed as alternative representations for general nonlinear autoregressive models. We consider the following $\text{HAR}_{\log}$ model:

$$RV^l_{t+1} = \beta_{0,t+1} + \beta_{1,t+1} RV^l_t + \beta_{2,t+1} RV^l_{t-1|t-n_w} + \beta_{3,t+1} RV^l_{t-n_w-1|t-n_w-n_m} + \eta_{t+1}, \quad \eta_{t+1} \sim \text{NID}(0, q_{t+1}) \tag{6.1.21}$$

where now coefficients change over time.

Score-driven models, as described in Section 2.3, provide a useful methodological framework to model time-varying parameters. By defining the vector of time-varying parameters at time t as

$$f_t = \Big(\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \log q_t\Big)'$$

the next value f_{t+1} is determined by:

$$f_{t+1} = \omega + A s_t + B f_t \tag{6.1.22}$$

where s_t is the scaled score vector:

$$s_t = (\mathcal{I}_{t|t-1})^{-1} \nabla_t$$

Let $\Theta = \{\omega, \text{vec}(A), \text{vec}(B)\}$ denotes the set of all the static parameters of the model. The conditional log-likelihood is given by:

$$\log p(RV^l_t | f_t, \mathcal{B}_{t-1}, \Theta) = -\frac{1}{2}\left(\log q_t + \frac{(RV^l_t - \mu_{t|t-1})^2}{q_t}\right) \tag{6.1.23}$$

where $\mu_{t|t-1}$ is the conditional mean:

$$\mu_{t|t-1} = \text{E}[RV^l_t | f_t, \mathcal{B}_{t-1}] = \beta_{0,t} + \beta_{1,t} RV^l_{t-1} + \beta_{2,t} RV^l_{t-2|t-n_w-1} + \beta_{3,t} RV^l_{t-n_w-2|t-n} \tag{6.1.24}$$

The static parameters Θ are estimated by numerically optimizing the log-likelihood function:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \log L(RV_1^l, \ldots, RV_T^l|\Theta) \tag{6.1.25}$$

where $\log L(RV_1^l, \ldots, RV_T^l|\Theta)$ is obtained by summing all the conditional log-likelihood functions (6.1.23) and updating the time-varying parameters using eq. (6.1.22):

$$\log L(RV_1^l, \ldots, RV_T^l|\Theta) = \sum_{t=1}^{T} \log p(RV_t^l|f_t, \mathcal{B}_{t-1}, \Theta) \tag{6.1.26}$$

We name model (6.1.21), equipped with the update rule (6.1.22), as Score-HAR (SHAR) model. In appendix (A.9) we will recover expressions for ∇_t and $\mathcal{I}_{t|t-1}$.

In order to examine whether the SHAR is able to remove the misspecification of the $\text{HAR}_{\log}$, we look at the standardized residuals $\hat{\eta}_t = (RV_t^l - \mu_{t|t-1})/\sqrt{q_t}$ obtained by estimating the model on the same sample used in Section 6.1.5. Figure (6.1.8) shows sample autocorrelations of $\hat{\eta}_t$ and $\hat{\eta}_t^2$ while figure (6.1.9) shows kernel density estimates for $\hat{\eta}_t$. The weak heteroskedasticity of $\text{HAR}_{\log}$ residuals in figure (6.1.6) has now disappeared and $\hat{\eta}_t$ is iid. Slight deviations from normality in the tails are imputable to the huge spikes observed during the 2008-2009. Similar results are obtained by estimating the SHAR on individual stock data.

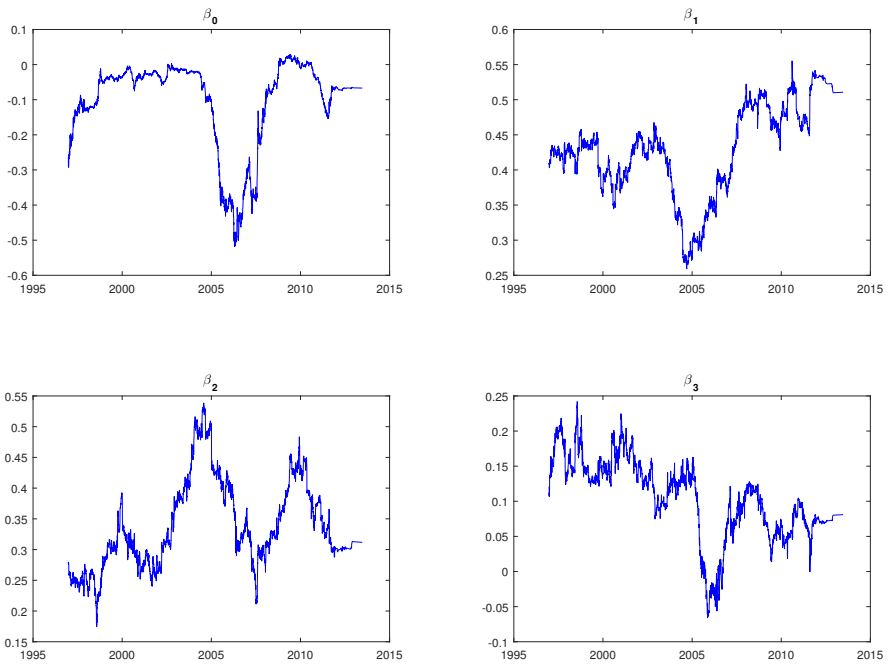


Figure 6.1.7: Dynamics of $\text{HAR}_{\log}$ coefficients obtained by estimating the model on a moving window of 500 observations in the period 23-12-1996 − 21-06-2013.

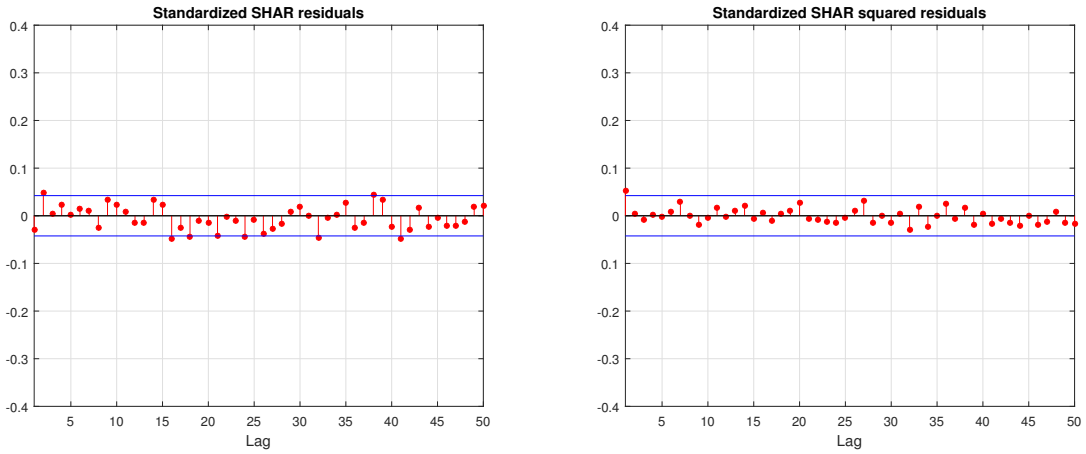


Figure 6.1.8: Sample autocorrelation function of SHAR standardized residuals and squared standardized residuals estimated the series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013.

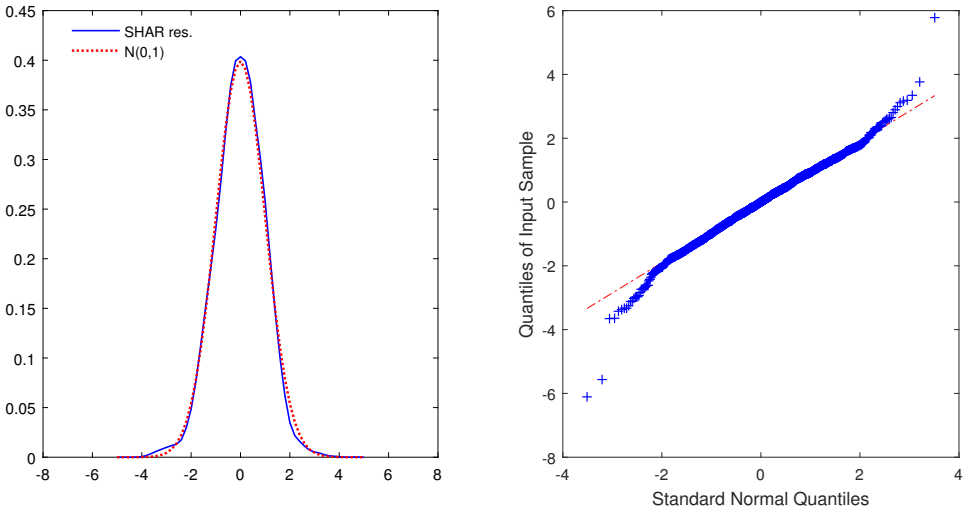


Figure 6.1.9: Kernel density estimate of SHAR standardized residuals computed on the series of 5-min daily realized variance of S&P500 future prices from 03-01-1995 to 21-06-2013. Q-Q plot of standardized residuals.

6.1.7 The SHARK model

The HARK allows to account for measurement errors when estimating and forecasting volatility with an $\text{HAR}_{\log}$ specification. However, the $\text{HAR}_{\log}$ exhibits heteroskedastic residuals and time-varying coefficients. As done with the SHAR, we can introduce time-varying parameters in the HARK in order to correct for its residual misspecification. The new model reads:

$$RV_t^l = Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, H_t) \tag{6.1.27}$$

$$\alpha_{t+1} = c_t + T_t\alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q_t) \tag{6.1.28}$$

where the system matrices have the same structure as in eq. (6.1.16) and (6.1.17) but parameters $\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, q_t$ are now time-varying.

As shown in Section 2.4, score-driven models are convenient when one wants to model dynamically the parameters of a linear Gaussian state-space representation. Indeed, the resulting model is conditionally Gaussian and can be estimated by standard maximum likelihood. As in the SHAR, the vector of time-varying parameters is given by:

$$f_t = \Big(\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, \log q_t\Big)' \tag{6.1.29}$$

and follows the usual update rule:

$$f_{t+1} = \zeta + Cs_t + Df_t \tag{6.1.30}$$

where $s_t = (\mathcal{I}_{t|t-1})^{-1}\nabla_t$. The score ∇_t and the information matrix $\mathcal{I}_{t|t-1}$ can be computed as:

$$\nabla_t = -\frac{1}{2}\left[\dot{F}_t'(I_{n_t} \otimes F_t^{-1})\text{vec}(I_{n_t} - v_t v_t' F_t^{-1}) + 2\dot{v}_t' F_t^{-1} v_t\right] \tag{6.1.31}$$

$$\mathcal{I}_{t|t-1} = \frac{1}{2}\left[\dot{F}_t'(F_t^{-1} \otimes F_t^{-1})\dot{F}_t + 2\dot{v}_t' F_t^{-1}\dot{v}_t\right] \tag{6.1.32}$$

where v_t and F_t are the Kalman filter prediction error and its covariance matrix. The two quantities v_t and F_t are computed as an output of the Kalman filter recursions (2.4.5)-(2.4.10) for conditionally Gaussian models. Instead, $\dot{v}_t$ and $\dot{F}_t$ denote derivatives of v_t and F_t with respect to f_t and can be computed through a parallel set of recursions that is reported in appendix (A.10).

The log-likelihood can be computed in the prediction-error decomposition:

$$\log L(RV_1^l, \ldots, RV_T^l|\Omega) = -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|F_t| + v_t' F_t^{-1} v_t\right) \tag{6.1.33}$$

where $\Omega = \{\zeta, \text{vec}(C), \text{vec}(D)\}$ denotes the set of static parameters of the model. The time-varying parameters are updated at each time step using eq. (6.1.30) and computing ∇_t and $\mathcal{I}_{t|t-1}$ as described above. The static parameters are estimated by optimizing numerically the log-likelihood with a quasi-Newton method:

$$\hat{\Omega} = \underset{\Omega}{\text{argmax}} \log L(RV_1^l, \ldots, RV_T^l|\Omega) \tag{6.1.34}$$

We name model (6.1.27), (6.1.28), equipped with the update rule (6.1.30), as Score-HAR-Kalman (SHARK). As we will see in our empirical study in Section 6.3, compared to the SHAR, the SHARK provides more responsive time-varying parameters, as they are not affected by the attenuation bias due to measurement errors.

6.1.8 Forecast

In this Section we will provide guidelines for computing one-step and multi-step-ahead forecasts from our models. Except for the HAR, all other models are estimated on $\log(RV_t)$ series. Thus, when computing forecasts of RV_t, it is essential to take into account the bias generated by the logarithm transformation.

The one-step-ahead forecast of the $\text{HAR}_{\log}$ model is simply computed using the moment generating function of the normal distribution:

$$\text{E}_t[\exp(RV^l_{t+1})] = \exp\left(\beta_0 + \beta_1 RV^l_t + \beta_2 RV^l_{t-1|t-n_w} + \beta_3 RV^l_{t-n_w-1|t-n_w-n_m} + \frac{q^2}{2}\right) \tag{6.1.35}$$

In the HARK, from the theory of linear Gaussian state-space models, we know that $\alpha_{t+1}|\mathcal{F}_t \sim \text{N}(a_{t+1}, P_{t+1})$. Thus, using the moment generating function of the multivariate normal distribution, we have:

$$\text{E}_t[\exp(Z\alpha_{t+1})] = \exp\left(Za_{t+1} + \frac{1}{2}ZP_{t+1}Z'\right) \tag{6.1.36}$$

Since the time-varying parameters in the SHAR and SHARK are $\mathcal{F}_t$-measurable, we obtain analogous formulas to eq. (6.1.35) and (6.1.36), respectively. In particular, in the case of the SHAR, we have:

$$\text{E}_t[\exp(RV^l_{t+1})] = \exp\left(\beta_{0,t+1} + \beta_{1,t+1} RV^l_t + \beta_{2,t+1} RV^l_{t-1|t-n_w} + \beta_{3,t+1} RV^l_{t-n_w-1|t-n_w-n_m} + \frac{q^2_{t+1}}{2}\right) \tag{6.1.37}$$

while in the case of the SHARK we have the same one-step-ahead forecast of the HARK in eq. (6.1.36) but a_{t+1} and P_{t+1} are computed with the Kalman filter recursions (2.4.5)-(2.4.10).

We will provide closed form bias-corrected expressions for multi-step ahead forecasts of $\text{HAR}_{\log}$ and HARK models. In the SHAR and SHARK models the presence of time-varying parameters does not allow for similar closed form expressions. In this case, as suggested by Creal et al. (2014), we evaluate the conditional mean of the predictive density through simulations. This is easily done by simulating recursively eq. (6.1.21), (6.1.22) for the SHAR and eq. (6.1.27), (6.1.28), (6.1.30) for the SHARK.

As done with the HARK, we write the $\text{HAR}_{\log}$ model in a vectorial representation by introducing the n-dimensional column vector $\overline{RV}^l_t = (RV^l_{t+1}, RV^l_t, \cdots, RV^l_{t-n+2})'$. Eq. (6.1.20) can thus be written as:

$$\overline{RV}^l_{t+1} = c_l + T_l \overline{RV}^l_t + \overline{\eta}_t, \quad \overline{\eta}_t \sim \text{NID}(0, Q_l) \tag{6.1.38}$$

where c_l, T_l and Q_l have the same structure as c, T and Q in Section 6.1.4 but contain the parameters of the $\text{HAR}_{\log}$ model. In appendix (A.11), (A.12), we prove the following two propositions:

Proposition 4. *The j-th step-ahead forecast of the HAR_{log} model is given by:*

$$E_t[\exp(Z\overline{RV}_{t+j})] = \exp\left\{ Z\left[(\mathbb{I}_n + \cdots + T_l^{j-1})c_l + T_l^j \overline{RV}_t^l\right] + \frac{1}{2} Z\left[Q + \cdots + T^{j-1} Q (T^{j-1})'\right] Z' \right\}$$

Proposition 5. *The j-th step-ahead forecast of the HARK model is given by:*

$$E_t[\exp(Z\alpha_{t+j})] = \exp\left\{ Z\left[(\mathbb{I}_n + \cdots + T^{j-2})c + T^{j-1} a_{t+1}\right] + \frac{1}{2} Z\left[T^{j-1} P_{t+1} (T^{j-1})' + Q + \cdots + T^{j-2} Q (T^{j-2})'\right] Z' \right\}$$

6.2 Simulation study

The linear DGP used in Section 6.1.3 does note provide a good description of empirical realized variance series, as we have seen that they exhibit evidence of nonlinear dependencies, such as heteroskedasticity and time-varying parameters. In order to test our models on realistic time-series, we perform an extensive simulation study using the two-factor stochastic volatility model of Huang and Tauchen (2005). The same DGP was also used by Bollerslev et al. (2016a) to test the HARQ model (see their appendix A). We simulate 1-sec data on a regularly spaced grid of 23400 timestamps per day and then compute daily realized variances by summing up $M = 39, 78, 390$ intraday squared returns, corresponding to 10, 5, 1-min sampling frequencies. The simulated RV series have $T = 3000$ observations. The in-sample analysis is performed on the last 1000 observations. The out-of-sample analysis is performed on the same sample of 1000 observations but the models are estimated on a rolling window of 2000 observations, starting from the first 2000 simulated data. We generate $N = 1000$ Monte-Carlo realizations and estimate the following models for both the in-sample and out-of-sample analysis: (i) HAR, (ii) HARQ, (iii) HAR_{log}, (iv) HARK, (v) SHAR and (vi) SHARK.

The SHAR and the SHARK, featuring dynamic HAR coefficients, tend to provide very close forecast performances when static parameters are estimated on the initial window and when they are re-estimated on a daily basis. As such, we generate out-of-sample forecasts from these models using parameter estimates recovered on the initial window. All other models are re-estimated on a daily basis. We found particularly effective and parsimonious constraining the dynamic equations (6.1.22), (6.1.30) of the time-varying parameters in the SHAR and SHARK as:

$$f_{t+1} = f_t + A s_t \tag{6.2.1}$$

$$f_{t+1} = f_t + C s_t \tag{6.2.2}$$

where A and C are 5×5 diagonal matrices. More flexible specifications are possible, but they do not lead to superior out-of-sample forecasts. Thus, the static parameters to be estimated in the SHAR are the five diagonal elements of A, that we denote as a_1, a_2, a_3, a_4, a_5 and, similarly, the static parameters to be estimated in the SHARK are the five diagonal elements of C, that we denote as c_1, c_2, c_3, c_4, c_5.

	HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK
	In-sample					
	$M = 39$					
MSE	1.0000	**0.97620**	0.98331	0.99210	0.98376	0.99994
MAE	1.0000	0.98676	0.97516	**0.96248**	0.97643	0.96478
QLIKE	1.0000	0.97062	**0.9499**	0.96401	0.95050	0.96581
Signal-to-Noise				7.54		8.00
	$M = 78$					
MSE	1.0000	**0.97981**	0.98710	0.99148	0.98916	0.99474
MAE	1.0000	0.98754	0.97748	**0.97048**	0.97937	0.97310
QLIKE	1.0000	0.96751	**0.9487**	0.95535	0.94931	0.95572
Signal-to-Noise				13.62		14.30
	$M = 390$					
MSE	1.0000	**0.98222**	0.98884	0.99017	0.99069	0.99024
MAE	1.0000	0.98870	0.97817	**0.97639**	0.97982	0.97905
QLIKE	1.0000	0.96532	**0.9446**	0.94634	0.94502	0.9457
Signal-to-Noise				61.17		62.70
	Out-of-sample					
	$M = 39$					
MSE	1.0000	0.98924	0.96767	0.97639	**0.96051**	0.98691
MAE	1.0000	0.98696	0.96911	0.95757	0.97252	**0.95645**
QLIKE	1.0000	0.97056	0.92626	0.94155	**0.92271**	0.94842
	$M = 78$					
MSE	1.0000	0.99382	0.97347	0.97772	**0.96987**	0.98167
MAE	1.0000	0.98805	0.97159	0.96532	0.97537	**0.96447**
QLIKE	1.0000	0.96044	0.92487	0.93186	**0.92185**	0.93414
	$M = 390$					
MSE	1.0000	0.99856	0.97658	0.99175	**0.97406**	0.97688
MAE	1.0000	0.98878	**0.97313**	0.97600	0.97803	0.97371
QLIKE	1.0000	0.95602	0.92355	0.94976	**0.92073**	0.92364

Table 6.2.1: Average relative in-sample and out-of-sample losses of HAR, HARQ, HAR_{log}, HARK, SHAR, SHARK models on 1000 simulated daily RV data. Signal-to-Noise denotes the average signal-to-noise ratio estimated by HARK and SHARK models.

The first part of table (6.2.1) compares the average in-sample mean square error (MSE), mean absolute error (MAE) and QLIKE, defined as:

$$\text{QLIKE} = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{IV_t}{\widehat{IV}_t} - \log\frac{IV_t}{\widehat{IV}_t} - 1\right) \tag{6.2.3}$$

where $\widehat{IV}_t$ denotes the one-step-ahead forecast obtained by the model. The true IV_t used in the com-

putation of the loss measures is computed by summing up 1-sec squared returns[3]. We also show the average signal-to-noise ratio estimated by the HARK and SHARK models.

The $\text{HAR}_{\log}$ outperforms the HAR. Indeed, contrary to the linear DGP in Section 6.1.3, the simulated series feature nonlinear behavior and, as it has been shown on S&P500 data, the logarithmic transformation removes most of the heteroskedasticity of the HAR. This translates into better in-sample estimates.

As expected, the HARQ provides better estimates than the HAR in terms of all the three loss measures. Note that the relative MSE of the HARQ approaches 1 as M becomes larger. This is due to the effect of measurement errors. However, the relative QLIKE is significantly lower than 1 even at $M = 390$, a setting in which measurement errors are very small. This can be interpreted as an indication that the HARQ is also capturing other effects than measurement errors.

Except for the fact that the HARK outperforms all other models in terms of MAE, in this in-sample analysis there are no significant advantages in using the HARK, SHAR and SHARK in place of the $\text{HAR}_{\log}$. The reason is twofold. On the one hand, estimation errors are small, as indicated by the large values of the signal-to-noise ratio δ. Indeed, the amount of noise on real data is larger, as indicated by the lower values of δ that we have found in our empirical application in Section 6.3. Second, the filtered time-varying parameters resulting from estimating the SHAR and the SHARK are less erratic than what found on empirical data.

It is more interesting to look at out-of-sample results, summarized in the second part of table (6.2.1). Both the $\text{HAR}_{\log}$ and the HARQ perform better than the HAR but the $\text{HAR}_{\log}$ outperforms the HARQ in terms of all the three loss measures. This implies that, if one is interested in forecasting, modeling logarithmic time series through the $\text{HAR}_{\log}$ is more effective than correcting for measurement errors through the HARQ. Note that the relative performance of the HARQ that we have found in the in-sample and out-of-sample analysis is in agreement with that of Bollerslev et al. (2016a) obtained in the same simulation setting.

The HARK still provides better estimates than the $\text{HAR}_{\log}$ in terms of MAE. In contrast to the in-sample analysis, introducing time-varying parameters allows improving over the forecasts provided by the $\text{HAR}_{\log}$ and the HARK. Indeed, the SHAR has lowest MSE and QLIKE while the SHARK has lowest MAE. Therefore, even in presence of weak nonlinear dependencies, modeling the residual misspecification of $\text{HAR}_{\log}$ and HARK leads to significantly better out-of-sample forecasts.

[3]The simulated prices are not contaminated by microstructure noise, so the realized variance estimator is a consistent and unbiased estimator of the true IV. In presence of microstructure noise, one only needs to replace the realized variance estimator with a robust estimator (see e.g. Zhang et al. 2005) and using a consistent estimator for the variance of the error to employ as a proxy of H_t.

6.3 Empirical evidence

In order to test the proposed models on real time series, we use both high-frequency data of S&P500 market index future and individual stock prices. The S&P500 sample is the same as the one used in Section 6.1.5. It includes 4259 business days, from 03-01-1995 to 21-06-2013. Equity data are provided by Thomson Reuters and comprise 18 frequently traded NYSE stocks. The sample goes from 03-01-2006 to 31-12-2014, spanning a total of 2250 business days. Table (6.3.1) provides summary statistics of the series of realized variances computed with 5-min returns.

Stock	Symbol	Min	Mean	Median	Max
SP500		0.0177	0.9243	0.4808	38.2914
Citigroup	C	0.1473	9.6360	2.5460	972.4663
Morgan Stanley	MS	0.1847	11.1067	2.8120	1.64e+03
Goldman Sachs	GS	0.2056	4.9457	1.6992	394.4546
JPMorgan Chase	JPM	0.1035	5.2330	1.6949	254.1726
Bank of America	BAC	0.1088	7.8898	2.1562	377.5072
ConocoPhillips	COP	0.1303	2.8366	1.4524	191.3988
Exxon Mobil	XOM	0.1130	1.9655	0.9875	135.4296
Chevron	CVX	0.1052	2.2589	1.1585	142.7586
Schlumberger	SLB	0.3042	4.6566	2.7218	165.4489
General Electric	GE	0.1088	3.2878	1.1553	172.7190
CBS Corporation	CBS	0.2013	5.5533	2.2906	165.6725
Walt Disney	DIS	0.1380	2.2760	1.1164	112.5801
Halliburton Company	HAL	0.1964	5.5976	3.1641	205.0786
Johnson & Johnson	JNJ	0.0668	0.9579	0.5167	49.6791
McDonald's	MCD	0.0867	1.3879	0.7031	124.4459
Pfizer	PFE	0.1620	1.9173	1.1373	64.0720
Verizon Communications	VZ	0.1229	1.8900	0.9336	108.7225
Wal-Mart	WMT	0.1134	1.4193	0.7576	73.2718

Table 6.3.1: Summary statistics of realized variance series computed with 5-min returns for S&P500 and NYSE stocks data.

Since one of the main purposes of this analysis is to assess the effect of measurement errors, we compute realized variance at different sampling frequencies. As done in the simulation study, we choose $M = 39, 78, 390$, corresponding to sampling 10, 5, 1-min returns[4].

For the out-of-sample analysis, a moving window of 2000 observations is employed to estimate the models and recover forecasts of the last $T - 2000$ RV's, where $T = 4259$ for S&P500 data and $T = 2250$

[4]The average realized variance for $M = 39, 78, 390$ is $0.90, 0.92, 0.89$, respectively, suggesting that microstructure effects are small and that RV_t provides unbiased estimates of IV_t for all the three sampling frequencies. Similar considerations hold for individual stock data.

for stock data. In the case of S&P500 data, the same subsample of 2259 observations is used to perform the in-sample analysis. For individual stock data, we perform the in-sample analysis on the last 1000 observations.

We constraint the static parameters of the SHAR and SHARK models as in eq. (6.2.1), (6.2.2). As found in the simulation study, re-estimating the SHAR and SHARK on a moving window or employing parameter estimates recovered on the first window leads to similar loss measures and same models ranking. The HARQ may provide negative RV_t estimates. In that case, as suggested by Bollerslev et al. (2016b), we replace the negative estimate with the average of past realized variances.

6.3.1 In-sample analysis

Tables (6.3.2), (6.3.3), (6.3.4) show OLS estimates of HAR, HARQ, $\text{HAR}_{\log}$ coefficients and maximum likelihood estimates of HARK, SHAR, SHARK parameters obtained by estimating the models on S&P500 data with $M = 39, 78, 390$ in the period 23-12-1996 − 21-06-2013. The first 500 observations, from 03-01-1995 to 22-12-1996, are used for initializing the time-varying parameters. We estimate the $\text{HAR}_{\log}$ and HARK on this pre-sample and then use the estimated coefficients as starting values of the time-varying parameters of the SHAR and SHARK, respectively. In the same table we also report the MSE, QLIKE and the R^2 of the Mincer-Zarnowitz regression computed on the period used for the in-sample analysis. On the last three lines we show the average MSE, QLIKE and R^2 obtained on individual stock data.

The Kalman filter consistently takes into account measurement errors. For $M = 39$, the coefficient β_1 estimated by the HARK is 0.54 while the one of the $\text{HAR}_{\log}$ is 0.41, meaning that the former is correcting for the bias induced by estimation errors. As M increases, the bias becomes smaller and thus the estimate provided by the HARK becomes closer to that of the $\text{HAR}_{\log}$. The same effect is observed on loss measures, with the relative difference between the two models being larger for $M = 39$ and then reducing as M increases. Note that the signal-to-noise ratio increases as M increases and is smaller than the one estimated on simulated data.

Figure (6.3.1), (6.3.2), (6.3.3), (6.3.4) show, for $M = 39, 78, 390$, filtered estimates of time-varying parameters $\beta_{0,t}$, $\beta_{1,t}$, $\beta_{3,t}$, q_t. The static parameter in matrices A and C corresponding to $\beta_{2,t}$ is found to be zero for both the SHAR and the SHARK. The latter provides more responsive estimates, especially for $M = 39$. As M increases, the parameters filtered by the two models tend to show similar dynamics. Indeed, while the SHARK is able to filter out observational noise through the Kalman filter, SHAR estimates are bias-attenuated due to measurement errors. In figure (6.3.4), we see that both models capture the residual heteroskedasticity that we observed in figure (6.1.6) and that is left after taking the logarithm of RV_t. However, the time-varying variance estimated by the SHAR is larger than the one of the SHARK. This is another consequence of measurement errors. Indeed, the SHAR is not able to

disentangle the dynamics of the quarticity from those of q_t and provides filtered variance estimates that include the contribution of both terms.

The coefficient β_{1Q} of the HARQ is negative, in agreement with the empirical results of Bollerslev et al. (2016a). Since the correction term in the HARQ accounts for measurement errors, as M increases β_{1Q} should become smaller in absolute value. In contrast, for $M = 390$, it is still significantly different form zero and it is close to the one computed for $M = 39$. This empirical finding is in agreement with what we have found in the simulation study and indicates that the HARQ term may also capture nonlinearities.

Further insights related to the effect of measurement errors are given by log-likelihood estimates. As expected, the SHARK has the largest estimated log-likelihood for all M's, implying better in-sample fit resulting from modeling both measurement errors and time-varying parameters. The estimated log-likelihood of the HARK is larger than the one of the SHAR for $M = 39$ and $M = 78$, while it is smaller for $M = 390$, as the effect of measurement errors becomes less important.

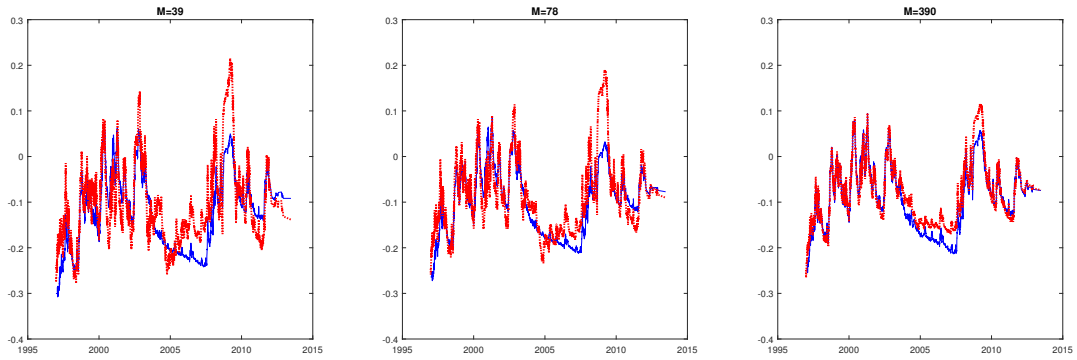


Figure 6.3.1: Dynamics of filtered SHAR (blue lines) and SHARK (red lines) $\beta_{0,t}$ parameter obtained by estimating the two models on S&P500 realized variance computed with $10, 5, 1$-min returns in the period 23-12-1996 − 21-06-2013.

6.3.2 Out-of-sample analysis

Table (6.3.5) shows the results of the out-of-sample analysis. As suggested by Patton (2011), we considered the MSE and QLIKE which are known to provide robust loss measures when comparing out-of-sample forecasts of noisy volatility estimates[5].

In our framework, the relative average loss between two models can be interpreted as the forecast

[5]Since the MSE is highly sensitive to the huge volatility spikes observed during the financial crisis, we considered averages of squared errors until 01-09-2008, a total of 1437 days. The QLIKE is more robust to huge spikes and it provides the same ranking when computed in the whole sample. However, to make it comparable with the MSE, we reported its value computed on the same sample.

	$M = 39$					
	HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK
β_0	0.0879	-0.0402	-0.0372	-0.0309		
	(0.0264)	(0.0270)	(0.0119)	(0.0153)		
β_1	0.3460	0.7068	0.4167	0.5420		
	(0.0160)	(0.0284)	(0.0158)	(0.0338)		
β_2	0.3962	0.3110	0.3655	0.2944		
	(0.0229)	(0.0229)	(0.0219)	(0.0445)		
β_3	0.1656	0.0947	0.1602	0.1165		
	(0.0229)	(0.0227)	(0.0203)	(0.0308)		
β_{1Q}		-0.0119				
		(0.0007)				
a_1, c_1					2.2996	3.6931
					(0.0794)	(0.2418)
a_2, c_2					0.2300	0.5442
					(0.0804)	(0.0525)
a_3, c_3					0.0000	0.0001
					(0.0000)	(0.0002)
a_4, c_4					0.0001	0.1944
					(0.0003)	(0.0847)
a_5, c_5					1.5477	1.9583
					(0.1651)	(0.0415)
Signal-to-Noise				3.0487		3.7064
MSE	2.2490	2.0464	2.2108	2.2037	2.1689	2.1617
QLIKE	0.1844	0.6328	0.1847	0.1837	0.1822	0.1783
R^2	0.5676	0.6065	0.5798	0.5810	0.5832	0.5859
$-\text{LogL} \cdot 10^{-3}$				1.8585	1.8716	1.8455
$\overline{\text{MSE}}$	5.7066	5.5769	6.3001	6.2670	6.1582	6.2732
$\overline{\text{QLIKE}}$	0.1772	0.1774	0.1732	0.1756	0.1739	0.1758
$\overline{R}^2$	0.3833	0.3954	0.3397	0.3490	0.3442	0.3318

Table 6.3.2: OLS estimates of HAR, HARQ, HAR_{log} coefficients and maximum likelihood estimates of HARK, SHAR, SHARK parameters obtained on RV series of S&P500 future prices computed with 10-min returns. Robust standard errors are indicated inside parenthesis. We also show the in-sample MSE, QLIKE, R^2, estimated log-likelihood and the average in-sample MSE, QLIKE, R^2 obtained on RV series of individual stock prices. The SHAR and SHARK parameters are multiplied by 100.

	$M = 78$					
	HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK
β_0	0.0921	0.0127	-0.0307	-0.0270		
	(0.0280)	(0.0279)	(0.0106)	(0.0147)		
β_1	0.3249	0.5780	0.4500	0.5480		
	(0.0159)	(0.0237)	(0.0157)	(0.0299)		
β_2	0.4166	0.3211	0.3531	0.2914		
	(0.0232)	(0.0236)	(0.0214)	(0.0367)		
β_3	0.1661	0.1309	0.1442	0.1147		
	(0.0236)	(0.0232)	(0.0193)	(0.0276)		
β_{1Q}		-0.0059				
		(0.0004)				
a_1, c_1					2.0385	3.3315
					(0.0629)	(0.1300)
a_2, c_2					0.3252	0.5295
					(0.0358)	(0.0304)
a_3, c_3					0.0000	0.0000
					(0.0002)	(0.0013)
a_4, c_4					0.0020	0.2346
					(0.0021)	(0.0370)
a_5, c_5					1.9973	1.7019
					(0.0632)	(0.0368)
Signal-to-Noise				5.0582		6.0415
MSE	2.1577	1.9766	2.1549	2.1587	2.1431	2.1542
QLIKE	0.1536	0.5941	0.1532	0.1507	0.1501	0.1465
R^2	0.5973	0.6311	0.6025	0.6046	0.6054	0.6075
$-\text{LogL} \cdot 10^{-3}$				1.6737	1.6851	1.6561
$\overline{\text{MSE}}$	5.5758	5.3013	5.9656	5.8987	5.8485	5.9093
$\overline{\text{QLIKE}}$	0.1452	0.1464	0.1400	0.1407	0.1401	0.1408
$\overline{R}^2$	0.3808	0.4073	0.3586	0.3687	0.3617	0.3564

Table 6.3.3: OLS estimates of HAR, HARQ, HAR_{log} coefficients and maximum likelihood estimates of HARK, SHAR, SHARK parameters obtained on RV series of S&P500 future prices computed with 5-min returns. Robust standard errors are indicated inside parenthesis. We also show the in-sample MSE, QLIKE, R^2, estimated log-likelihood and the average in-sample MSE, QLIKE, R^2 obtained on RV series of individual stock prices. The SHAR and SHARK parameters are multiplied by 100.

	$M = 390$					
	HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK
β_0	0.0853	-0.0687	-0.0272	-0.0264		
	(0.0257)	(0.0246)	(0.0091)	(0.0138)		
β_1	0.3539	0.8679	0.4996	0.5428		
	(0.0160)	(0.0257)	(0.0155)	(0.0273)		
β_2	0.4166	0.2652	0.3442	0.3161		
	(0.0227)	(0.0219)	(0.0206)	(0.0343)		
β_3	0.1429	0.0131	0.1056	0.0931		
	(0.0220)	(0.0211)	(0.0176)	(0.0269)		
β_{1Q}		-0.0104				
		(0.0004)				
a_1, c_1					2.7082	2.8887
					(0.2107)	(0.0212)
a_2, c_2					1.0119	0.9008
					(0.1249)	(0.0139)
a_3, c_3					0.0000	0.0000
					(0.0028)	(0.0064)
a_4, c_4					0.1531	0.3086
					(0.0298)	(0.0093)
a_5, c_5					2.1238	1.7671
					(0.1465)	(0.0719)
Signal-to-Noise				16.5013		19.0204
MSE	2.3724	1.9556	2.3027	2.3058	2.3188	2.2867
QLIKE	0.1246	0.3402	0.1210	0.1194	0.1191	0.1174
R^2	0.5890	0.6612	0.6048	0.6052	0.6024	0.6060
$-\text{LogL} \cdot 10^{-3}$				1.3861	1.3674	1.3540
$\overline{\text{MSE}}$	4.0100	3.8106	4.1341	4.1382	4.1397	4.0882
$\overline{\text{QLIKE}}$	0.1011	0.0997	0.0962	0.0966	0.0963	0.0960
$\overline{R}^2$	0.4358	0.4627	0.4322	0.4410	0.4330	0.4356

Table 6.3.4: OLS estimates of HAR, HARQ, HAR_{log} coefficients and maximum likelihood estimates of HARK, SHAR, SHARK parameters obtained on RV series of S&P500 future prices computed with 1-min returns. Robust standard errors are indicated inside parenthesis. We also show the in-sample MSE, QLIKE, R^2, estimated log-likelihood and the average in-sample MSE, QLIKE, R^2 obtained on RV series of individual stock prices. The SHAR and SHARK parameters are multiplied by 100.

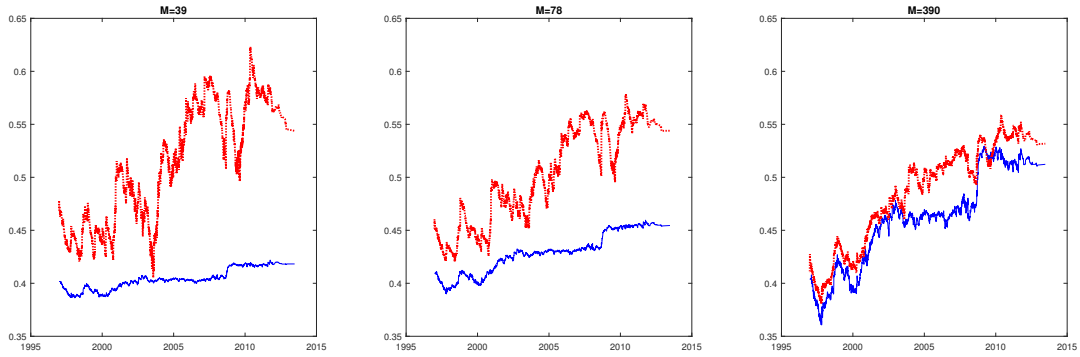


Figure 6.3.2: Dynamics of filtered SHAR (blue lines) and SHARK (red lines) $\beta_{1,t}$ parameters obtained by estimating the two models on S&P500 realized variance computed with $10, 5, 1$-min returns in the period 23-12-1996 – 21-06-2013.

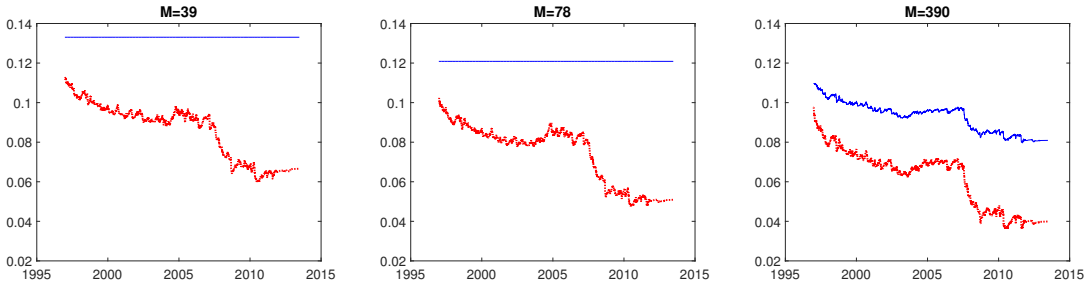


Figure 6.3.3: Dynamics of filtered SHAR (blue lines) and SHARK (red lines) $\beta_{3,t}$ parameters obtained by estimating the two models on S&P500 realized variance computed with $10, 5, 1$-min returns in the period 23-12-1996 – 21-06-2013.

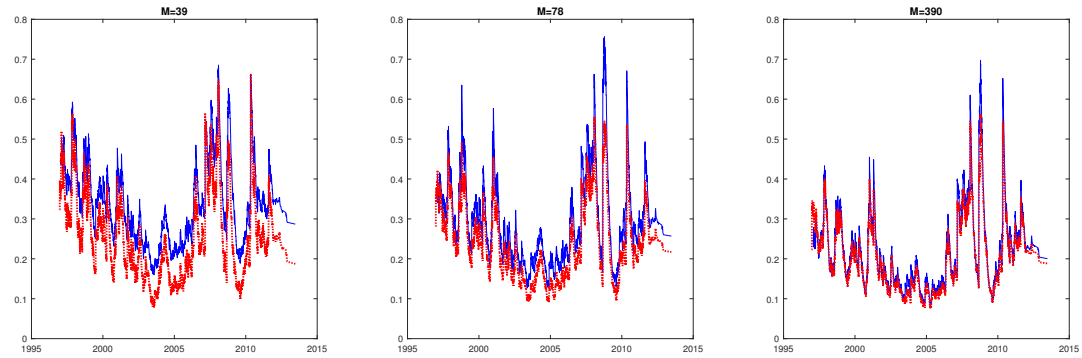


Figure 6.3.4: Dynamics of filtered SHAR (blue lines) and SHARK (red lines) q_t parameter obtained by estimating the two models on S&P500 realized variance computed with $10, 5, 1$-min returns in the period 23-12-1996 – 21-06-2013.

gain resulting from taking into account different forms of misspecification of the HAR. Given a generic loss measure $L(\cdot)$, we define:

$$\varphi_{\text{het}} = \frac{\text{E}[L(\text{HAR}_{\text{log}})]}{\text{E}[L(\text{HAR})]} \tag{6.3.1}$$

$$\varphi^{(1)}_{\text{tvp}} = \frac{\text{E}[L(\text{SHAR})]}{\text{E}[L(\text{HAR}_{\text{log}})]}, \quad \varphi^{(2)}_{\text{tvp}} = \frac{\text{E}[L(\text{SHARK})]}{\text{E}[L(\text{HARK})]} \tag{6.3.2}$$

$$\varphi^{(1)}_{\text{err}} = \frac{\text{E}[L(\text{HARK})]}{\text{E}[L(\text{HAR}_{\text{log}})]}, \quad \varphi^{(2)}_{\text{err}} = \frac{\text{E}[L(\text{SHARK})]}{\text{E}[L(\text{SHAR})]} \tag{6.3.3}$$

The first measure, φ_{het}, quantifies the forecast gain resulting from removing part of the heteroskedasticity of HAR residuals. The other two measures, $\varphi^{(1)}_{\text{tvp}}$ and $\varphi^{(2)}_{\text{tvp}}$, quantify the effect of modeling the residual nonlinearities by introducing time-varying parameters. Finally, $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(2)}_{\text{err}}$ measure the contribution of measurement errors. Table (6.3.6) shows the values of the forecast gains defined in eq. (6.3.1)-(6.3.3) computed on S&P500 and on individual stocks using the MSE and the QLIKE as loss measures.

On S&P500 data, the SHARK provides the lowest MSE and QLIKE. The only exception is for $M = 390$. In this case, the SHAR provides a slightly lower MSE, as measurement errors are less important at large sampling frequencies. Except for $M = 390$, where the HARQ has slightly lower MSE loss, the HARK always outperforms the latter in terms of both loss measures. Compared to the basic HAR_{log} model, all the three extensions considered here provide lower MSE and QLIKE.

These results are corroborated by the analysis based on the model confidence set (MCS) of Hansen et al. (2011). For S&P500 data, table (6.3.5) shows in parenthesis the p-values resulting from a 90% MCS, that we denote as $\widehat{\mathcal{M}}_{90\%}$, constructed using the MSE and QLIKE. The three HAR extensions are the only models belonging to $\widehat{\mathcal{M}}_{90\%}$ constructed with the QLIKE. If one uses the MSE, the HARK and the SHARK are the only models included in $\widehat{\mathcal{M}}_{90\%}$ for $M = 39, 78$. For $M = 390$, the test has not enough power to exclude any of the models from the confidence set. Such results indicate that the effects captured by the HARK, SHAR and SHARK models are relevant for volatility forecasting.

The HAR_{log} has always lower MSE and QLIKE compared to the HARQ, except for $M = 39, 390$. However, in the first case the HARQ is not included in $\widehat{\mathcal{M}}_{90\%}$, while in the latter both models are included. Note that, as M increases, $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(2)}_{\text{err}}$ approach 1, as one would expect from a model which accounts exclusively for measurement errors. On average, $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(2)}_{\text{err}}$ are smaller for $M = 78$ than $M = 39$, suggesting that the estimator of the noise variance in eq. (6.1.7) provides less precise estimates of the true variance H_t when computed using 10-min returns.

The forecast gain coming from taking into account heteroskedasticity turns out to be substantial if one uses the QLIKE as a loss measure. Indeed, from table (6.3.6), we see that φ_{het} is around 0.9 for $M = 390$ and even smaller in the remaining cases. More moderate gains are obtained using the MSE.

	HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK
			S&P500			
			$M = 39$			
MSE	1.0000	0.9897	0.9927	$0.9699^{\star}$	0.9770	$\mathbf{0.9508}^{\star}$
	(0.033)	(0.033)	(0.033)	(0.233)	(0.033)	(1.000)
QLIKE	1.0000	0.9037	0.8863	0.8765	$0.8643^{\star}$	$\mathbf{0.8536}^{\star}$
	(0.015)	(0.022)	(0.022)	(0.057)	(0.200)	(1.000)
			$M = 78$			
MSE	1.0000	0.9714	0.9555	$0.9284^{\star}$	0.9452	$\mathbf{0.9163}^{\star}$
	(0.003)	(0.057)	(0.089)	(0.587)	(0.089)	(1.000)
QLIKE	1.0000	0.8689	0.7844	$0.7668^{\star}$	$0.7733^{\star}$	$\mathbf{0.7455}^{\star}$
	(0.000)	(0.000)	(0.013)	(0.180)	(0.180)	(1.000)
			$M = 390$			
MSE	$1.0000^{\star}$	$0.9595^{\star}$	$0.9781^{\star}$	$0.9636^{\star}$	$\mathbf{0.9524}^{\star}$	$0.9538^{\star}$
	(0.141)	(0.932)	(0.141)	(0.736)	(1.000)	(0.932)
QLIKE	1.0000	0.9244	0.9022	0.8911	$0.8830^{\star}$	$\mathbf{0.8759}^{\star}$
	(0.008)	(0.019)	(0.019)	(0.096)	(0.295)	(1.000)
			Stocks			
			$M = 39$			
MSE	$1.0000^{(3)}$	$0.7651^{(5)}$	$0.5639^{(5)}$	$0.5433^{(14)}$	$0.5504^{(14)}$	$\mathbf{0.5377}^{(18)}$
QLIKE	$1.0000^{(3)}$	$0.9579^{(3)}$	$0.7366^{(10)}$	$0.7295^{(14)}$	$0.7242^{(16)}$	$\mathbf{0.7188}^{(17)}$
			$M = 78$			
MSE	$1.0000^{(2)}$	$0.9034^{(3)}$	$0.6612^{(7)}$	$0.6485^{(12)}$	$0.6525^{(15)}$	$\mathbf{0.6428}^{(18)}$
QLIKE	$1.0000^{(3)}$	$1.2628^{(3)}$	$0.7071^{(11)}$	$0.7010^{(13)}$	$0.6952^{(17)}$	$\mathbf{0.6890}^{(18)}$
			$M = 390$			
MSE	$1.0000^{(1)}$	$0.8492^{(1)}$	$0.5437^{(3)}$	$0.5691^{(9)}$	$0.5315^{(9)}$	$\mathbf{0.5216}^{(18)}$
QLIKE	$1.0000^{(2)}$	$2.3093^{(3)}$	$0.6002^{(10)}$	$0.5972^{(12)}$	$0.5944^{(17)}$	$\mathbf{0.5866}^{(17)}$

Table 6.3.5: Relative out-of-sample losses of HAR, HARQ, HAR_{log}, HARK, SHAR, SHARK models on S&P500 data and average out-of-sample losses of the same models estimated on 18 individual stock data. In the first case we show in parenthesis the p-values of the MCS at 90% c.l. computed using the MSE and QLIKE. The presence of an asterisk indicates that the model is included in $\widehat{\mathcal{M}}_{90\%}$. In the second case we report the number of times the model is included in $\widehat{\mathcal{M}}_{90\%}$.

Gains coming from modeling the residual nonlinearity through time-varying parameters are roughly equal to 2% on average, as can be seen by looking at the values of $\varphi^{(1)}_{\text{tvp}}$ and $\varphi^{(2)}_{\text{tvp}}$.

On stock data, the HARK, SHAR and SHARK improve significantly over the performance of the $\text{HAR}_{\log}$ model, as indicated by lower MSE and QLIKE and by the MCS analysis. With only two exceptions, the SHARK is always included in $\widehat{\mathcal{M}}_{90\%}$, while the HARK and SHAR are included more frequently than the remaining models.

The impact of measurement errors, as quantified by $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(2)}_{\text{err}}$, is slightly lower for $M = 39$ than $M = 78$. While $\varphi^{(1)}_{\text{err}}$ approaches 1 for $M = 390$, $\varphi^{(2)}_{\text{err}}$ remains low. This is due to the surprising improvement of the SHARK compared to all other models, as also indicated by the very small value of $\varphi^{(2)}_{\text{tvp}}$. The reason is that stock volatility exhibits more pronounced parameters dynamics which are not well captured by the SHAR, even in presence of small measurement errors.

The relative QLIKE of the HARQ is significantly larger than 1 for $M = 78, 390$. The performance of the HAR on stock data also deteriorates substantially compared to the $\text{HAR}_{\log}$. In both cases the reason of such a poor performance is due to the estimation window including the 2008 financial crisis, which strongly affects the OLS estimates of the two models. The $\text{HAR}_{\log}$, being estimated on $\log RV_t$ series, is robust to these huge peaks, as can be seen from the very low values of φ_{het}. Finally, forecast gains due to accounting for residual nonlinearity and assessed by $\varphi^{(1)}_{\text{tvp}}$ are of the same size of those found on S&P500 data.

Our empirical findings reveal substantial improvements coming from correcting for the three types of misspecification of the HAR. In particular, heteroskedasticity has deep impacts on HAR volatility forecasts, especially when the model is estimated on turbulent periods. Measurement errors are relevant at small and intermediate sampling frequencies, while modeling time-varying parameters results on forecast gains which are roughly equal to 2%.

We conclude this Section by examining in more detail what kind of misspecification is captured by the HARQ. Indeed, we have found on both simulated and empirical data that the HARQ corrects for both measurement errors and nonlinear dependencies. In order to quantify the extent to which this happens, we regress out-of-sample forecast errors of the HARQ against forecast errors provided by all other models. Table (6.3.7) shows the R^2 obtained from such regression on S&P500 data for $M = 39, 78, 390$. Forecasts errors of the HARQ are better described in terms of forecast errors of the SHARK, for $M = 39, 78$. Thus, at these sampling frequencies, the model is capturing both measurement errors and nonlinear dependencies. This is also confirmed by the fact that the R^2 resulting from regressing against the HARK and the SHAR is higher than the one resulting from regressing against the $\text{HAR}_{\log}$, where these effects are not taken into account. For $M = 390$, the highest R^2 is the one obtained by regressing against the SHAR, while the relative difference between regressions onto $\text{HAR}_{\log}$ and HARK

	φ_{het}	$\varphi^{(1)}_{\text{tvp}}$	$\varphi^{(2)}_{\text{tvp}}$	$\varphi^{(1)}_{\text{err}}$	$\varphi^{(2)}_{\text{err}}$
			S&P500		
			$M = 39$		
MSE	0.9927	0.9842	0.9803	0.9771	0.9732
QLIKE	0.8863	0.9752	0.9739	0.9890	0.9876
			$M = 78$		
MSE	0.9555	0.9872	0.9922	0.9715	0.9764
QLIKE	0.7844	0.9746	0.9817	0.9775	0.9846
			$M = 390$		
MSE	0.9781	0.9737	0.9898	0.9852	1.0015
QLIKE	0.9022	0.9787	0.9829	0.9876	0.9919
			Stocks		
			$M = 39$		
MSE	0.5639	0.9762	0.9897	0.9635	0.9768
QLIKE	0.7366	0.9831	0.9854	0.9904	0.9926
			$M = 78$		
MSE	0.6612	0.9869	0.9911	0.9809	0.9851
QLIKE	0.7071	0.9832	0.9829	0.9914	0.9911
			$M = 390$		
MSE	0.5437	0.9775	0.9165	1.0468	0.9815
QLIKE	0.6002	0.9904	0.9822	0.9951	0.9869

Table 6.3.6: Forecast gains φ_{het}, $\varphi^{(1)}_{\text{tvp}}$, $\varphi^{(2)}_{\text{tvp}}$, $\varphi^{(1)}_{\text{err}}$, $\varphi^{(2)}_{\text{err}}$ as defined in eq. (6.3.1), (6.3.2), (6.3.3) and computed using the MSE and QLIKE as loss measures. In case of stocks, we report the average gains.

is lower. This result is not surprising, since at large sampling frequencies measurement errors become less important and the forecast gains provided by the HARQ are entirely due to capturing nonlinear dependencies.

	HAR	HAR_{log}	HARK	SHAR	SHARK
			$M = 39$		
HARQ	0.9490	0.9129	0.9281	0.9627	0.9741
			$M = 78$		
HARQ	0.9321	0.9004	0.9235	0.9475	0.9737
			$M = 390$		
HARQ	0.9395	0.9405	0.9517	0.9779	0.9623

Table 6.3.7: R^2 resulting from regressing forecast errors of the HARQ against forecast errors of all other models.

6.3.3 Longer forecast horizons

In this Section we assess the effect of measurement errors and nonlinearity on out-of-sample volatility forecasts for longer time horizons. In particular, we choose a weakly ($j = 5$) and a monthly ($j = 22$) forecast horizon.

	HAR	HARQ	HAR_{log}	HARK	SHAR	SHARK
			j = 5			
			$M = 39$			
MSE	1.0000	0.9519	0.9302	0.9047*	0.9229	**0.8732***
	(0.000)	(0.001)	(0.008)	(0.276)	(0.091)	(1.000)
QLIKE	1.0000	0.9263	0.8253	0.8051*	0.8021*	**0.7729***
	(0.001)	(0.036)	(0.083)	(0.232)	(0.232)	(1.000)
			$M = 78$			
MSE	1.0000	0.9204	0.8853	0.8652*	0.8607*	**0.8460***
	(0.000)	(0.004)	(0.004)	(0.512)	(0.512)	(1.000)
QLIKE	1.0000	0.8917	0.7662	0.7452*	0.7387*	**0.7243***
	(0.000)	(0.010)	(0.023)	(0.544)	(0.544)	(1.000)
			$M = 390$			
MSE	1.0000	0.9422	0.9190	0.9084*	0.9140	**0.9033***
	(0.000)	(0.003)	(0.003)	(0.783)	(0.003)	(1.000)
QLIKE	1.0000	0.9387	0.8572	0.8426*	0.8375*	**0.8250***
	(0.000)	(0.003)	(0.003)	(0.568)	(0.568)	(1.000)
			j = 22			
			$M = 39$			
MSE	1.0000	1.0081	0.8686	0.8334*	0.8568*	**0.7936***
	(0.000)	(0.000)	(0.008)	(0.2350)	(0.2350)	(1.000)
QLIKE	1.0000	0.9755	0.8023	0.7630	0.8045	**0.6935***
	(0.001)	(0.001)	(0.006)	(0.074)	(0.074)	(1.000)
			$M = 78$			
MSE	1.0000	0.9896	0.8319	0.8039*	0.8050*	**0.7847***
	(0.000)	(0.000)	(0.001)	(0.852)	(0.852)	(1.000)
QLIKE	1.0000	0.9747	0.7697	0.7347*	0.7634*	**0.6886***
	(0.000)	(0.000)	(0.000)	(0.218)	(0.218)	(1.000)
			$M = 390$			
MSE	1.0000	1.0187	0.9409	0.9241*	0.9400	**0.9014***
	(0.000)	(0.000)	(0.000)	(0.461)	(0.000)	(1.000)
QLIKE	1.0000	1.0150	0.8712	0.8468*	0.8459*	**0.7993***
	(0.002)	(0.002)	(0.002)	(0.157)	(0.157)	(1.000)

Table 6.3.8: Relative out-of-sample losses of HAR, HARQ, HAR_{log}, HARK, SHAR, SHARK models on S&P500 data for weakly ($j = 5$) and monthly ($j = 22$) variance forecasts. Values in parenthesis denote p-values of the MCS at 90% c.l. computed using the MSE and QLIKE. The presence of an asterisk indicates that the model is included in $\widehat{\mathcal{M}}_{90\%}$.

Bias-corrected, multi-step-ahead forecasts of the HARK, SHAR and SHARK models can be evaluated as described in Section 6.1.8. In absence of a dynamic equation for the quarticity term in eq. (6.1.9), longer horizon forecasts for the HARQ model can only be evaluated directly, i.e. by replacing the daily RV_t on the left-hand-side of eq. (6.1.9) with the variance aggregated at different frequencies. Correspondingly, a correction term is introduced to adjust the lag at the specific forecast horizon (see Bollerslev et al. 2016a). Ghysels et al. (2009) point out that direct approaches are in general inferior to methods based on iterated forecasts.

Table (6.3.8) shows the relative out-of-sample MSE and QLIKE obtained on S&P500 data while table (6.3.9) reports the corresponding forecast gains. The SHARK provides smaller MSE and QLIKE on all the scenarios and it is always included in $\widehat{\mathcal{M}}_{90\%}$. With only one exception, the HARK and the SHAR outperform the HAR_{log} and are included in $\widehat{\mathcal{M}}_{90\%}$ in most of the cases. The HARQ improves over the HAR but is outperformed by the HAR_{log}, which provides better forecasts, especially on the monthly horizon.

Measurement errors have a similar effect as in daily forecasts, with $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(2)}_{\text{err}}$ approaching one as M increases. However, note that forecast gains increase as j increases, as longer horizon forecasts benefit from the higher persistence estimated by both the HARK and SHARK. A similar dependence on the forecast horizon is observed when modeling time-varying parameters. Figure (6.3.5) shows MSE-based forecast gains φ_{het}, $\varphi^{(1)}_{\text{err}}$, $\varphi^{(1)}_{\text{tvp}}$, for $j = 2, 4, \ldots, 22$ and $M = 78$. Both $\varphi^{(1)}_{\text{err}}$ and $\varphi^{(1)}_{\text{tvp}}$ exhibit a weak dependence on j, with the two measures declining slowly as the forecast horizon increases. In contrast, φ_{het} has a strong declining pattern that is due to the accumulation of significant forecast errors over large horizons.

The statistical significance of these forecasting gains is jointly tested through a pairwise version of the multi-horizon test recently proposed by Quaedvlieg (2017). In particular, as shown in table (6.3.10), we find strong evidence of uniform superior predictive ability for φ_{het}, by testing HAR$-\text{HAR}_{\text{log}}$ MSE losses, and for $\varphi^{(1)}_{\text{err}}$, by testing $\text{HAR}_{\text{log}}-$HARK MSE losses. In these cases the test achieves basically zero p-values. For φ_{tvp}, the test rejects the null hypothesis at the 5% c.l. when testing $\text{HAR}_{\text{log}}-$SHAR MSE losses.

6.4 Robustness checks

The discussion in Section 6.1.5 and the in-sample analysis of Section 6.3 showed that, even on $\log RV_t$ series, the HAR is misspecified, as its residuals are weakly heteroskedastic and parameters are time-varying. The HARK corrects HAR_{log} forecasts for measurement errors through the Kalman filter. However, being the latter based on a linear state-space representation, it is useful to quantify the misspecification of the HARK due to the above-mentioned effects. The forecast gain $\varphi^{(2)}_{\text{tvp}}$ in eq. (6.3.2) is a possible indicator,

	φ_{het}	$\varphi^{(1)}_{\text{tvp}}$	$\varphi^{(2)}_{\text{tvp}}$	$\varphi^{(1)}_{\text{err}}$	$\varphi^{(2)}_{\text{err}}$
			j = 5		
			$M = 39$		
MSE	0.9302	0.9922	0.9652	0.9726	0.9461
QLIKE	0.8253	0.9719	0.9600	0.9755	0.9636
			$M = 78$		
MSE	0.8853	0.9834	0.9778	0.9773	0.9717
QLIKE	0.7662	0.9667	0.9720	0.9726	0.9779
			$M = 390$		
MSE	0.9190	0.9946	0.9944	0.9885	0.9829
QLIKE	0.8572	0.9770	0.9791	0.9830	0.9851
			j = 22		
			$M = 39$		
MSE	0.8686	0.9864	0.9523	0.9595	0.9263
QLIKE	0.8023	1.0028	0.9088	0.9511	0.8620
			$M = 78$		
MSE	0.8319	0.9676	0.9762	0.9662	0.9749
QLIKE	0.7697	0.9919	0.9372	0.9546	0.9020
			$M = 390$		
MSE	0.9409	0.9990	0.9754	0.9822	0.9589
QLIKE	0.8712	0.9710	0.9439	0.9720	0.9449

Table 6.3.9: Forecast gains for weakly ($j = 5$) and monthly ($j = 22$) forecast horizons.

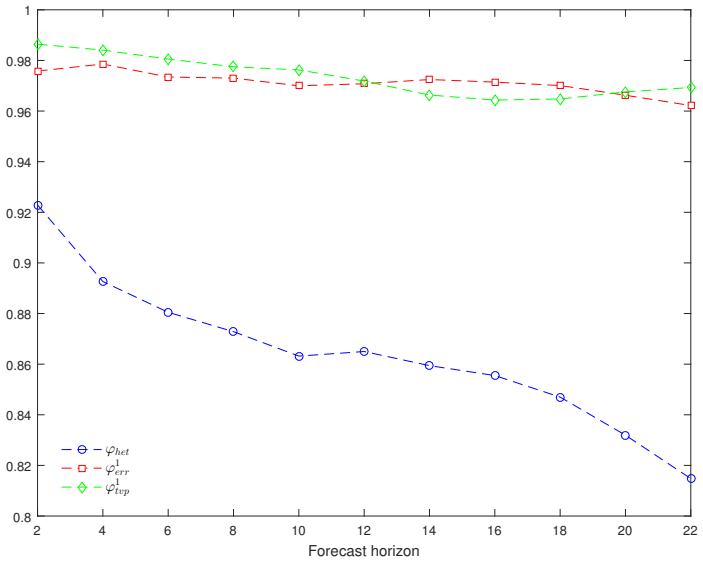


Figure 6.3.5: MSE-based forecast gains φ_{het}, φ^1_{err}, φ^1_{tvp} at different forecast horizons.

	φ_{het}	φ_{err^1}	φ_{tvp^1}
	uSPA test		
t_{uSPA}	3.5273	2.8827	1.8295
p-value	0	0	0.0365

Table 6.3.10: MSE based uniform superior predictive ability test (uSPA) of Quaedvlieg (2017). We show the t_{uSPA} test statistics and related p-values based on a multi-horizon bootstrap with 5000 replications.

since it is nothing but the gain resulting from modelling heteroskedasticity and time-varying parameters in the HARK.

Another possible method is to observe that the assumption $H_t = V_t$ in Section 6.1.5 is fair only if the HARK is correctly specified. Thus, one can use a more flexible specification for H_t and see, on real data, the extent to which the latter deviates from the assumption based on correct specification. In particular, we have examined the following dynamic specifications for H_t:

1. $H_t = \alpha V_t + (1-\alpha)V_{t-1}$
2. $H_t = \alpha V_t + (1-\alpha)\text{mean}(V_t)$
3. $H_t = \alpha V_t + (1-\alpha)\text{median}(V_t)$
4. $H_t = \beta V_t$

where $0 \leq \alpha \leq 1$, $\beta > 0$ and $\text{mean}(\cdot)$ and $\text{median}(\cdot)$ denote the sample mean and median, respectively. In the first case, H_t is exponentially smoothed with past observations of V_t and the parameter α indicates the level of smoothing. In the second and third case, if $\alpha < 1$, H_t is shrunk towards its mean or median level. This can also be viewed as an heteroskedasticity test on H_t, in the sense that, if $\alpha > 0$, we reject the null assumption that H_t is homoskedastic, at a given confidence level. Finally, in the last case, β is a scaling constant that fine-tunes H_t in case the model is misspecified.

	$M = 39$	$M = 78$	$M = 390$
	β		
HARK	2.0595	2.9399	6.9714
	(0.2106)	(0.2856)	(2.7879)
SHARK	0.8969	1.2051	2.3594
	(0.2511)	(0.3036)	(1.8479)

Table 6.4.1: Maximum-likelihood estimates of the scaling parameter β in the dynamic specification 4. in Section 6.4. Standard errors are shown in parenthesis.

We estimated the HARK and SHARK models, with the dynamic specifications $\{1, 2, 3, 4\}$ for H_t and combinations of them, on the same S&P500 sample used in the in-sample analysis of Section 6.3. For each specification involving the parameter α, the latter turns out to be very close to one. For instance, for

specification 2, we found $\alpha = 0.999994$, and similar values have been recovered for other specifications. This indicates that today's V_t is the best proxy for H_t and that there are not significant advantages in using lagged V_t's. Also, the null assumption of homoskedasticity for H_t is rejected, and this is in accordance with with what Bollerslev et al. (2016a) have found on RV_t series.

Table (6.4.1) shows maximum likelihood estimates of the parameter β for $M = 39, 78, 390$. We see that the estimated β is larger than one for the HARK, indicating significant departures from the assumption of correct specification. In particular, the fact that the estimated signal-to-noise ratio turns out to be lower, is due to deviations from the linearity assumption, which are spuriously interpreted by the Kalman filter as an excess of noise. It is worth underling that these more flexible specifications do not lead to significant out-of-sample improvements compared to the standard HARK model.

In the case of the SHARK, β is close to one for all the three sampling frequencies, in agreement with the assumption of correct specification. This result is in accordance with what we have found in Section 6.1.7 when looking at the normalized residuals of the SHAR: implementing score-driven parameters in the HAR_{log} results in removing the misspecification due to heteroskedasticity and time-varying parameters. The SHARK, which in turn corrects SHAR forecasts for measurement errors, provides therefore the most complete specification for realized volatility.

Chapter 7

Conclusions

In this thesis, we contributed to the econometric literature on time-varying parameter models. The first contribution is related to the estimation of nonlinear non-Gaussian state-space models. We showed how to recover approximate smoothed estimates through a fully observation-driven framework. More specifically, we extended popular score-driven models to include actual and future observations in a setting where they are viewed as misspecified filters for nonlinear non-Gaussian state-space models. The SDS is general, as it can be applied to any observation density. It is also computationally simple, as it consists of a backward recursion following the standard forward score-driven recursion. At the same time, it provides estimates that are very close to those of correctly specified parameter-driven models, which, instead, require computationally demanding simulation-based techniques.

The rest of the thesis has been focused on the investigation of important empirical properties of high-frequency financial data through the introduction of suitable time-varying parameter models. In particular, (i) we concentrated on the estimation of high-frequency lead-lag correlation and motivated their existence through a micro-founded model of lagged price adjustment; (ii) we introduced a conditional correlation model for noisy and asynchronous observations to study the dynamic dependence structure of high-frequency prices; finally, (iii) we proposed several extensions of the HAR model of realized volatility to keep into account and disentangle the effects of measurement errors and nonlinear dynamics. In the following, we discuss in more detail the results obtained in our empirical analysis.

In Chapter 4, we have introduced the MLA, a multi-asset model of price generation that extends standard univariate microstructure models of lagged price adjustment. Lead-lag effects are naturally incorporated in this framework through nonzero non-diagonal coefficients in the speed of adjustment matrix Ψ. The latter captures lagged dissemination of information among assets. Using extensive Monte-Carlo experiments, we have shown that, as opposed to alternative estimators, the MLA is robust to spurious correlations arising from asynchronous trading. Also, we have tested the performance of the estimator in presence of misspecified DGP's and found that, when covariances are time-varying,

parameter estimates are still unbiased.

We have tested the MLA on a cross-section of NYSE assets. The analysis provides empirical evidence for the existence of a multi-asset price formation mechanism. In particular, we observed strong deviations from the null assumption of a standard random walk plus noise process. The non-diagonal elements in the VAR matrix are found to be significantly different from zero. As such, the speed of adjustment matrix contains non-diagonal elements that are responsible for lagged dissemination of information across assets. Cross-asset effects are more pronounced in periods of high volatility. We argue that this empirical fact can be explained by the behavior of high-frequency trading strategies. Indeed, the latter tend to exploit short living cross autocorrelations that are likely to appear in periods of high uncertainty.

The LLSD, introduced in Chapter 5, is a dynamic conditional correlation model that is particularly appealing when observations of the underlying process are noisy and asynchronous. Both noise and asynchronicity are handled by writing the model in a state-space form with missing values. By allowing the parameters of the model to be driven by the score of the predictive likelihood, the state-space representation turns out to be conditionally Gaussian, implying that the Kalman filter can be applied in order to write down the likelihood in closed form. The TAP parametrization allows recovering positive definite estimates of the conditional covariance matrix.

Through extensive Monte-Carlo experiments, we showed that the LLSD provides better estimates compared to standard pre-averaging and synchronization techniques that are typically used to deal with noisy and asynchronous high-frequency financial data. This is mainly due to two reasons. The first is that the Kalman filter optimally estimates the latent state variable and the second is that the LLSD is able to employ all the available data, thus providing more efficient estimates. By applying the model to a cross-section of NYSE transaction data, we showed that the LLSD can recover the typical intraday patterns of volatilities and correlations and, at the same time, it instantaneously captures the effect of real-time, fast changes of volatilities and correlations following e.g. macro-news announcements. Building a sequence of intraday optimal portfolios, we found that the LLSD provides superior out-of-sample forecasts and large efficiency gains over the standard DCC model.

In Chapter 6, we proposed new HAR extensions aimed at removing the three main evidences of misspecification of the HAR that are imputable to neglecting measurement errors on integrated variance estimates and nonlinear dependencies in the dynamics of integrated variance.

In particular, starting from the HAR_{log} model, which provides a better dynamic specification for realized variance, we devised three models that progressively take into account all evidences of misspecification. The HARK model, being based on the Kalman filter, corrects HAR_{log} parameters for the attenuation bias induced by measurement errors through a time-varying Kalman gain driven by realized quarticity. The SHAR, having time-varying parameters, removes the heteroskedasticity that is found on

HAR_{log} residuals. Indeed, it features i.i.d. standardized residuals when estimated on real data. Finally, the SHARK is a generalization of the HARK that allows for time-varying parameters. It provides more responsive parameter dynamics, as they are not affected by measurement errors. The relative improvement of a model compared to another can thus be used to quantify the importance of removing each form of misspecification.

We have provided simulation and empirical evidence that the effects captured by these new models are statistically significant, as they translate into superior out-of-sample forecasts compared to the basic HAR_{log} model and other competing approaches. For instance, on real data, the analysis based on the model confidence set includes our new extensions on the confidence set and excludes the remaining models in the vast majority of cases. As a final result of this analysis, we concluded that (i) measurement errors are relevant at small and intermediate sampling frequencies and the corresponding forecast gains slightly increase with the forecast horizon; (ii) accounting for heteroskedasticity always provides significant gains, especially at longer forecast horizons; (iii) forecast gains due to time-varying parameters are independent on the sampling frequency and slightly increase with the forecast horizon.

It is worth emphasizing that, even though the MLA, LLSD and SHAR(K) have been developed in the field of high-frequency financial econometrics, they are rather general and can be potentially useful for a wide range of applications. For instance, mixed frequency and noisy data are quite common in macroeconomics, when one wants to jointly model daily and quarterly data affected by measurement errors. In this case the MLA and LLSD can be used to study lead-lag correlations, as well as the evolution of the dependence structure over time. On the other hand, the SHAR(K) is a general framework to model long-memory processes exhibiting non-linear dynamics and measurement errors.

Bibliography

A. P. Dempster, N. M. Laird, D. B. R., 1977. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39 (1), 1–38.

Abadir, K., Magnus, J., 2005. Matrix Algebra. Econometric Exercises. Cambridge University Press.

Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High-frequency covariance estimates with noisy and asynchronous financial data. J. Amer. Statist. Assoc. 105 (492), 1504–1517.

Allez, R., Bouchaud, J.-P., 2011. Individual and collective stock dynamics: intra-day seasonalities. New Journal of Physics 13 (2), 025010.

Amihud, Y., Mendelson, H., 1987. Trading mechanisms and stock returns: An empirical investigation. The Journal of Finance 42 (3), 533–553.

Andersen, T., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4 (2-3), 115–158.

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

Andrade, S., Chang, C., Seasholes, M., 5 2008. Trading imbalances, predictable reversals, and cross-stock price pressure. Journal of Financial Economics 88 (2), 406–423.

Asai, M., McAleer, M., Medeiros, M. C., 2012. Modelling and forecasting noisy realized volatility. Computational Statistics & Data Analysis 56 (1), 217 – 230.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Journal of Econometrics 162 (2), 149 – 169.

Barndorff-Nielsen, O. E., Hansen, P. R., Shephard, A. L. N., 2009. Realized kernels in practice: trades and quotes. The Econometrics Journal 12 (3), C1–C32.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (2), 253–280.

Barndorff-Nielsen, O. E., Shephard, N., 2004. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. Econometrica 72 (3), 885–925.

Bekierman, J., Manner, H., Jul. 2016. Improved forecasting of realized variance measures. Working paper Available at https://ssrn.com/abstract=2812586.

Bernhardt, D., Taub, B., 2008. Cross-asset speculation in stock markets. The Journal of Finance 63 (5), 2385–2427.

Bibinger, M., 2012. An estimator for the quadratic covariation of asynchronously observed itô processes with noise: Asymptotic distribution theory. Stochastic Processes and their Applications 122 (6), 2411 – 2453.

Bibinger, M., Hautsch, N., Malec, P., Reiss, M., 2014. Estimating the spot covariation of asset prices: Statistical theory and empirical evidence. CFS Working Paper Series 477, Center for Financial Studies (CFS).

Blasques, F., Koopman, S. J., Lucas, A., Jun. 2014a. Maximum Likelihood Estimation for correctly Specified Generalized Autoregressive Score Models: Feedback Effects, Contraction Conditions and Asymptotic Properties. Tinbergen Institute Discussion Papers 14-074/III, Tinbergen Institute.

Blasques, F., Koopman, S. J., Lucas, A., 2014b. Optimal formulations for nonlinear autoregressive processes. Tinbergen Institute Discussion Papers 14-103/III, Tinbergen Institute.

Blasques, F., Koopman, S. J., Lucas, A., Aug. 2014c. Optimal formulations for nonlinear autoregressive processes. Working paper Available at https://ssrn.com/abstract=2478575.

Blasques, F., Koopman, S. J., Lucas, A., 2014d. Stationarity and ergodicity of univariate generalized autoregressive score processes. Electron. J. Statist. 8 (1), 1088–1112.

Blasques, F., Koopman, S. J., Lucas, A., 2015. Information-theoretic optimality of observation-driven time series models for continuous responses. Biometrika 102 (2), 325.

Blasques, F., Koopman, S. J., Lucas, A., 2017. Maximum likelihood estimation for score-driven models. Tinbergen Institute Discussion Papers 14-029/III, Tinbergen Institute.

Blasques, F., Koopman, S. J., Łasak, K., Lucas, A., 2016. In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. International Journal of Forecasting 32 (3), 875 – 887.

Bollerslev, T., April 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–327.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics 69 (3), 542–547.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2016a. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192 (1), 1 – 18.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., Apr. 2016b. Modeling and Forecasting (Un)Reliable Realized Covariances for More Reliable Financial Decisions. CREATES Research Papers 2016-10, Department of Economics and Business Economics, Aarhus University.

Buccheri, G., , Corsi, F., 2018a. Hark the shark: Realized volatility modelling with measurement errors and nonlinear dependencies. Working Paper, R&R at Journal of Financial Econometrics. Available at https://ssrn.com/abstract=3089929.

Buccheri, G., , Corsi, F., Peluso, S., 2018b. High-frequency lead-lag effects and cross-asset linkages: a multi-asset lagged adjustment model. Working Paper, submitted. Available at https://ssrn.com/abstract=2938619.

Buccheri, G., Bormetti, G., Corsi, F., Lillo, F., 2017a. A General Class of Score-Driven Smoothers. Working Paper, submitted. Available at https://ssrn.com/abstract=3139666.

Buccheri, G., Bormetti, G., Corsi, F., Lillo, F., 2017b. A score-driven conditional correlation model for noisy and asynchronous data: an application to high-frequency covariance dynamics. Working Paper, R&R at Journal of Business and Economic Statistics. Available at https://ssrn.com/abstract=2912438.

Buccheri, G., Bormetti, G., Corsi, F., Lillo, F., 2018c. Assessing filtering uncertainty in observation-driven models. Mimeo.

Caballé, J., Krishnan, M., 1994. Imperfect competition in a multi-security market with risk neutrality. Econometrica 62 (3), 695–704.

Caivano, M., Harvey, A., Luati, A., Mar 2016. Robust time series models with trend and seasonal components. SERIEs 7 (1), 99–120.

Carnero, M. A., Peña, D., Ruiz, E., 2004. Persistence and kurtosis in garch and stochastic volatility models. Journal of Financial Econometrics 2 (2), 319–342.

Cespa, G., Focault, T., Apr. 2011. Learning from Prices, Liquidity Spillovers, and Market Segmentation. CSEF Working Papers 284, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.

Chan, K., 1992. A further analysis of the lead–lag relationship between the cash market and stock index futures market. The Review of Financial Studies 5 (1), 123–152.

Chen, Y., Härdle, W. K., Pigorsch, U., 2010. Localized realized volatility modeling. Journal of the American Statistical Association 105 (492), 1376–1393.

Chiao, C., Hung, K., Lee, C. F., 2004. The price adjustment and lead-lag relations between stock returns: microstructure evidence from the taiwan stock market. Journal of Empirical Finance 11 (5), 709 – 731.

Cipollini, F., Gallo, G., Otranto, E., Sep. 2017. On heteroskedasticity and regimes in volatility forecasting. Working paper Available at https://ssrn.com/abstract=3037550.

Cogley, T., Sargent, T., 2001. Evolving Post-World War II U.S. Inflation Dynamics. Working Papers 2132872, Department of Economics, W. P. Carey School of Business, Arizona State University.

Cogley, T., Sargent, T. J., 2005. The conquest of us inflation: Learning and robustness to model uncertainty. Review of Economic Dynamics 8 (2), 528 – 563, monetary Policy and Learning.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7 (2), 174.

Corsi, F., Mittnik, S., Pigorsch, C., Pigorsch, U., 2008. The volatility of realized volatility. Econometric Reviews 27 (1-3), 46–78.

Corsi, F., Peluso, S., Audrino, F., 2015. Missing in asynchronicity: A kalman-em approach for multivariate realized covariance estimation. Journal of Applied Econometrics 30 (3), 377–397.

Corsi, F., Renò, R., 2009. Har volatility modelling with heterogeneous leverage and jumps. Available at SSRN 1316953.

Cox, D., 1981. Statistical analysis of time series: Some recent developments [with discussion and reply]. Scandinavian Journal of Statistics 8 (2), 93–115.

Creal, D., Koopman, S. J., Lucas, A., 2008. A general framework for observation driven time-varying parameter models. SSRN Electronic Journal.

Creal, D., Koopman, S. J., Lucas, A., 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. Journal of Business & Economic Statistics 29 (4), 552–563.

Creal, D., Koopman, S. J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28 (5), 777–795.

Creal, D., Schwaab, B., Koopman, S. J., Lucas, A., 2014. Observation-driven mixed-measurement dynamic factor models with an application to credit risk. The Review of Economics and Statistics 96 (5), 898–915.

Damodaran, A., 1993. A simple measure of price adjustment coefficients. The Journal of Finance 48 (1), 387–400.

de Jong, F., Nijman, T., 1997. High frequency analysis of lead-lag relationships between financial markets. Journal of Empirical Finance 4 (2), 259 – 277.

Delle Monache, D., Petrella, I., 2017. Adaptive models and heavy tails with an application to inflation forecasting. International Journal of Forecasting 33 (2), 482 – 501.

Delle Monache, D., Petrella, I., Venditti, F., Jul. 2015. Common faith or parting ways? A time varying parameters factor analysis of euro-area inflation. Birkbeck Working Papers in Economics and Finance 1515, Birkbeck, Department of Economics, Mathematics & Statistics.

Delle Monache, D., Petrella, I., Venditti, F., 2016. Adaptive state space models with applications to the business cycle and financial stress. CEPR Discussion Paper (DP11599).

Dobrev, D., Schaumburg, E., 2017. High-Frequency Cross-Market Trading: Model Free Measurement and Applications. Working paper.

Durbin, J., Koopman, S., 2000. Time series analysis of non gaussian observations based on state space models from both classical and bayesian perspectives. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62 (1), 3–56.

Durbin, J., Koopman, S., 2012. Time Series Analysis by State Space Methods: Second Edition. Oxford Statistical Science Series. OUP Oxford.

Durbin, J., Koopman, S. J., 1997. Monte carlo maximum likelihood estimation for non-gaussian state space models. Biometrika 84 (3), 669–684.

Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50 (4), 987–1007.

Engle, R., 2002a. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics 20 (3), 339–50.

Engle, R., 2002b. New frontiers for arch models. Journal of Applied Econometrics 17 (5), 425–446.

Engle, R., 2002c. New frontiers for arch models. Journal of Applied Econometrics 17 (5), 425–446.

Engle, R., Colacito, R., 2006. Testing and valuing dynamic correlations for asset allocation. Journal of Business & Economic Statistics 24, 238–253.

Engle, R., Russell, J., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66 (5), 1127–1162.

Engle, R. F., Gallo, G. M., 2006. A multiple indicators model for volatility using intra-daily data. Journal of Econometrics 131 (1), 3 – 27.

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

Ghysels, E., Harvey, A. C., Renault, E., 1996. 5 stochastic volatility. In: Statistical Methods in Finance. Vol. 14 of Handbook of Statistics. Elsevier, pp. 119 – 191.

Ghysels, E., Rubia, A., Valkanov, R., 2009. Multi-period forecasts of volatility: Direct, iterated, and mixed-data approaches.

Griffin, J. E., Oomen, R. C., 2011. Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Journal of Econometrics 160 (1), 58 – 68, realized Volatility.

Hamilton, J., 1994. Time series analysis. Princeton Univ. Press, Princeton, NJ.

Hansen, P. R., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business & Economic Statistics 24 (2), 127–161.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Hansen, R., Janus, P., Koopman, S., 2016. Realized wishart-garch: A score-driven multi-asset volatility model. WorkingPaper 16-061/III, Tinbergen Institute.

Harford, J., Kaul, A., 2005. Correlated order flow: Pervasiveness, sources, and pricing effects. The Journal of Financial and Quantitative Analysis 40 (1), 29–55.

Harvey, A., 1991. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Harvey, A., Chakravarty, T., Sep. 2008. Beta-t-(E)GARCH. Cambridge Working Papers in Economics 0840, Faculty of Economics, University of Cambridge.

Harvey, A., Luati, A., 2014. Filtering with heavy tails. Journal of the American Statistical Association 109 (507), 1112–1122.

Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. The Review of Economic Studies 61 (2), 247–264.

Harvey, A. C., 2013. Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series. Econometric Society Monographs. Cambridge University Press.

Hasbrouck, J., 1996. Modeling market microstructure time series. SSRN Electronic Journal NYU Working Paper No. FIN-95-024.

Hasbrouck, J., Ho, T. S. Y., 1987. Order arrival, quote behavior, and the return-generating process. Journal of Finance 42 (4), 1035–48.

Hasbrouck, J., Seppi, D. J., 2001. Common factors in prices, order flows, and liquidity. Journal of Financial Economics 59 (3), 383 – 411.

Hayashi, T., Koike, Y., Dec. 2016. Wavelet-based methods for high-frequency lead-lag analysis. Papers 1612.01232, arXiv.org.

Hayashi, T., Koike, Y., Aug. 2017. Multi-scale analysis of lead-lag relationships in high-frequency financial markets. Papers 1708.03992, arXiv.org.

Hayashi, T., Yoshida, N., 04 2005. On covariance estimation of non-synchronously observed diffusion processes. Bernoulli 11 (2), 359–379.

Hoffmann, M., Rosenbaum, M., Yoshida, N., 05 2013. Estimation of the lead-lag parameter from non-synchronous data. Bernoulli 19 (2), 426–461.

Huang, X., Tauchen, G., 2005. The relative contribution of jumps to total price variance. Journal of Financial Econometrics 3 (4), 456–499.

Huth, N., Abergel, F., 2014. High frequency lead/lag relationships — empirical facts. Journal of Empirical Finance 26, 41 – 58.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Processes and their Applications 119 (7), 2249 – 2276.

Jaeckel, P., Rebonato, R., 1999. The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. Journal of risk 2 (2).

Koopman, S. J., Lit, R., Lucas, A., Mar. 2015a. Intraday Stock Price Dependence using Dynamic Discrete Copula Distributions. Tinbergen Institute Discussion Papers 15-037/III/DSF90, Tinbergen Institute.

Koopman, S. J., Lucas, A., Scharth, M., 2015b. Numerically accelerated importance sampling for non-linear non-gaussian state-space models. Journal of Business & Economic Statistics 33 (1), 114–127.

Koopman, S. J., Lucas, A., Scharth, M., March 2016. Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models. The Review of Economics and Statistics 98 (1), 97–110.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59 (2), 347–370.

Nelson, D. B., 1992. Filtering and forecasting with misspecified arch models i: Getting the right variance with the wrong model. Journal of Econometrics 52 (1), 61 – 90.

Nelson, D. B., 1996. Asymptotically optimal smoothing with arch models. Econometrica 64 (3), 561–573.

Nelson, D. B., Foster, D. P., 1994. Asymptotic filtering theory for univariate arch models. Econometrica 62 (1), 1–41.

Nelson, D. B., Foster, D. P., 1995. Filtering and forecasting with misspecified arch models ii: Making the right forecast with the wrong model. Journal of Econometrics 67 (2), 303 – 335.

Oh, D. H., Patton, A. J., 2017. Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. Journal of Business & Economic Statistics 0 (0), 1–15.

Opschoor, A., Janus, P., Lucas, A., Dijk, D. V., 2017. New heavy models for fat-tailed realized covariances and returns. Journal of Business & Economic Statistics 0 (0), 1–15.

Pasquariello, P., Vega, C., 2015. Strategic cross-trading in the u.s. stock market*. Review of Finance 19 (1), 229.

Patton, A. J., 2006. Modelling asymmetric exchange rate dependence. International Economic Review 47 (2), 527–556.

Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160 (1), 246 – 256, realized Volatility.

Patton, A. J., Sheppard, K., 2009. Evaluating Volatility and Correlation Forecasts. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 801–838.

Protter, P., 1992. Stochastic Integration and Differential Equation, 2nd Edition. Springer-Verlag, Berlin, Heidelberg.

Quaedvlieg, R., Jun. 2017. Multi-horizon forecast comparison. Working paper Available at https://ssrn.com/abstract=2979352.

Rapisarda, F., Brigo, D., Mercurio, F., 2007. Parameterizing correlations: a geometric interpretation. IMA Journal of Management Mathematics 18 (1), 55–73.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. The Journal of Finance 39 (4), 1127–1139.

Sandmann, G., Koopman, S. J., 1998. Estimation of stochastic volatility models via monte carlo maximum likelihood. Journal of Econometrics 87 (2), 271–301.

Shephard, N. (Ed.), 2005. Stochastic Volatility: Selected Readings. Oxford University Press.

Shephard, N., Xiu, D., Dec. 2016. Econometric analysis of multivariate realised qml: Estimation of the covariation of equity prices under asynchronous trading. Chicago booth research paper no. 12-14.

Shumway, R. H., Stoffer, D. S., 1982. An approach to time series smoothing and forecasting using the em algorithm. Journal of Time Series Analysis 3 (4), 253–264.

Shumway, R. H., Stoffer, D. S., 2015. Time series analysis and its applications : with R examples. Springer texts in statistics. Springer, New York.

Simon, D., 2006. Kalman filter generalizations. John Wiley & Sons, Inc., pp. 183–227.

Stock, J. H., Watson, M. W., 2007. Why has u.s. inflation become harder to forecast? Journal of Money, Credit and Banking 39 (s1), 3–33.

Tauchen, G., Pitts, M., 1983. The price variability-volume relationship on speculative markets. Econometrica 51 (2), 485–505.

Tookes, H. E., 2008. Information, trading, and product market interactions: Cross-sectional implications of informed trading. The Journal of Finance 63 (1), 379–413.

Tsay, R. S., 2005. Analysis of financial time series. Wiley series in probability and statistics. Wiley-Interscience, Hoboken (N.J.).

Wansbeek, T., Meijer, E., 2000. Measurement Error and Latent Variables in Econometrics, 1st Edition. Elsevier.

Wu, C. F. J., 03 1983. On the convergence properties of the em algorithm. Ann. Statist. 11 (1), 95–103.

Zamojski, M., 2016. Filtering with confidence: In-sample confidence bands for garch filters. Working paper.

Zhang, F., Dec. 2010. High-frequency trading, stock volatility, and price discovery. Working paper, available at https://ssrn.com/abstract=1691679.

Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100 (472), 1394–1411.

Zu, Y., Boswijk, H. P., 2014. Estimating spot volatility with high-frequency financial data. Journal of Econometrics 181 (2), 117 – 135.

Appendix A

A.1 Notation

I_n denotes the $n \times n$ identity matrix. We use $\otimes$ to denote the Kronecker product between two matrices. The operator vec$(\cdot)$, applied to an $m \times n$ matrix A, stacks the columns of A into an mn column vector while the operator vech$(\cdot)$, applied to a symmetric $n \times n$ matrix B, stacks all the $n(n-1)/2$ upper (or lower) diagonal elements into a column vector. We also introduce the commutation matrix K_{mn}, i.e. the $mn \times mn$ matrix such that $K_{mn}\text{vec}A = \text{vec}A'$ for every $m \times n$ matrix A. The derivative of an $m \times n$ matrix function $F(X)$ with respect to the $p \times q$ matrix X is defined as in Abadir and Magnus (2005), i.e. as the $mn \times pq$ matrix computed as $\partial \text{vec}(F(X))/\partial \text{vec}(X)'$.

A.2 Proof of Proposition 1

We first compute the score:

$$\left[\frac{\partial \log p(y_t|\mathcal{F}_{t-1})}{\partial a_t'}\right]' = \left[\frac{\partial \log p(y_t|\mathcal{F}_{t-1})}{\partial v_t'}\frac{\partial v_t}{\partial a_t'}\right]' = [v_t' F_t^{-1} Z]' = Z' F_t^{-1} v_t \tag{A.2.1}$$

The information matrix is then computed as:

$$\mathcal{I}_{t|t-1} = \text{E}_{t-1}[\nabla_t \nabla_t'] = Z' F_t^{-1} Z \tag{A.2.2}$$

Thus, we can re-write recursions for $a_{t|t}$ and a_{t+1} as:

$$a_{t|t} = a_t + P_t \nabla_t \tag{A.2.3}$$

$$a_{t+1} = c + T a_t + T P_t \nabla_t \tag{A.2.4}$$

and the backward recursion for $\hat{\alpha}_t$ as

$$r_{t-1} = \nabla_t + L_t' r_t \tag{A.2.5}$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \tag{A.2.6}$$

where $L_t = T - T P_t \mathcal{I}_{t|t-1}$. In the steady state, P_t converges to the solution $\bar{P}$ of the matrix Riccati equation (3.1.5) (Harvey 1991, Durbin and Koopman 2012). By defining $R = T\bar{P}$, $\bar{F} = Z'\bar{P}Z + H$ and $\mathcal{I} = Z'\bar{F}^{-1}Z$, we can re-write the Kalman filtering and smoothing recursions for the mean in the steady state as:

$$a_{t|t} = a_t + T^{-1} R \nabla_t \tag{A.2.7}$$

$$a_{t+1} = c + T a_t + R \nabla_t \tag{A.2.8}$$

and

$$r_{t-1} = \nabla_t + (T - R\mathcal{I})' r_t \tag{A.2.9}$$

$$\hat{\alpha}_t = a_t + T^{-1} R r_{t-1} \tag{A.2.10}$$

Q.E.D.

A.3 DGP for time-varying covariances

We generate daily time series of realized covariance matrices and log-returns through the following method: (a) using daily log-returns of $k = 5, 10, 15$ randomly selected NYSE stocks, we compute the sample covariance matrix C over the period from 03-01-2006 to 31-12-2014, (b) consider the decomposition $C = UDU'$, where $D = \text{diag}[\lambda_1, \ldots, \lambda_k]$ is a diagonal matrix of decreasing eigenvalues of C and U is the corresponding matrix of normalized eigenvectors, (c) let the first r eigenvalues of C evolve through an AR(1) process:

$$\log \lambda_{t,j} = c_j + \phi_j \log \lambda_{t-1,j} + \beta_{t,j}, \quad \beta_{t,j} \sim \text{N}(0, q_j) \tag{A.3.1}$$

for $t = 2, \ldots, T$ and $j = 1, \ldots, r$. We therefore consider the sequence of matrices $V_t = UD_tU'$, where:

$$D_t = \text{diag}[\lambda_{t,1}, \ldots, \lambda_{t,r}, \lambda_{r+1}, \ldots, \lambda_k] \tag{A.3.2}$$

for $t = 1, \ldots T$. We set $r = 3$ and choose the parameters of the AR(1) model as: $\phi_1 = 0.99$, $q_1 = 0.05$, $\phi_2 = 0.95$, $q_2 = 0.05$, $\phi_1 = 0.90$, $q_1 = 0.5$ while constants c_j are chosen by setting the unconditional mean equal to $\log \lambda_j$, for $j = 1, 2, 3$. We then use the simulated covariance matrices V_t to generate observations of daily returns r_t realized covariance matrices X_t through the observation equations (3.2.20), (3.2.21). For each k, we set $\nu = k + 1$.

As loss measures, we use the root mean square error (RMSE) and the quasi-likelihood (Qlike). The former is defined as:

$$L_t^{\text{RMSE}} = \sqrt{\text{Tr}[(\widehat{V}_t - V_t)'(\widehat{V}_t - V_t)]} \tag{A.3.3}$$

while the Qlike loss is given by:

$$L_t^{\text{Qlike}} = \log |\widehat{V}_t| + \text{Tr}[\widehat{V}_t^{-1} V_t] \tag{A.3.4}$$

where $\widehat{V}_t$ generically denotes the estimate obtained through the SDF or the SDS. See Patton (2011) for a more detailed discussion on the use of these loss measures in the field of (co)variance estimation.

A.4 Computation of $\bar{P}^n_{t,t-1}$ in the MLA

We follow Shumway and Stoffer (1982). The set of Kalman filter recursions for the state-space model (4.1.9), (4.1.10) are given by:

$$\bar{X}_t^{t-1} = \phi \bar{X}_{t-1}^{t-1} \tag{A.4.1}$$

$$\bar{P}_t^{t-1} = \phi \bar{P}_{t-1}^{t-1} \phi' + \bar{Q} \tag{A.4.2}$$

$$K_t = \bar{P}_t^{t-1} M' (M \bar{P}_t^{t-1} M' + H)^{-1} \tag{A.4.3}$$

$$\bar{X}_t^t = \bar{X}_t^{t-1} + K_t (Y_t - M \bar{X}_t^{t-1}) \tag{A.4.4}$$

$$\bar{P}_t^t = \bar{P}_t^{t-1} - K_t H \bar{P}_t^{t-1} \tag{A.4.5}$$

for $t = 1, \ldots, n$. The set of backward smoothing recursions are given by:

$$J_{t-1} = \bar{P}_{t-1}^{t-1}\phi'(\bar{P}_t^{t-1})^{-1} \tag{A.4.6}$$

$$\bar{X}_{t-1}^n = \bar{X}_{t-1}^{t-1} + J_{t-1}(X_t^n - \phi\bar{X}_{t-1}^{t-1}) \tag{A.4.7}$$

$$\bar{P}_{t-1}^n = \bar{P}_{t-1}^{t-1} + J_{t-1}(\bar{P}_t^n - \bar{P}_t^{t-1})J_{t-1}' \tag{A.4.8}$$

for $t = n, \ldots, 1$. The covariance $\bar{P}_{t,t-1}^n$ in eq. (4.1.19) can be computed using the following backward recursion:

$$\bar{P}_{t-1,t-2}^n = \bar{P}_{t-1}^{t-1}J_{t-2}' + J_{t-1}(\bar{P}_{t,t-1}^n - \phi\bar{P}_{t-1}^{t-1})J_{t-2}' \tag{A.4.9}$$

where $t = n, \ldots 2$ and $\bar{P}_{n,n-1}^n = (I - K_nM)\phi\bar{P}_{n-1}^{n-1}$.

A.5 Proof of Proposition 2

We will use the following matrix differentiation rules:

$$\nabla_A \mathrm{tr}(AB) = B' \tag{A.5.1}$$

$$\nabla_A \mathrm{tr}(ABA'C) = CAB + C'AB' \tag{A.5.2}$$

$$\nabla_A |A| = |A|(A^{-1})' \tag{A.5.3}$$

where A, B and C are matrices of appropriate dimensions.

Let us re-write $G_1(F, Q)$ as:

$$G_1(F, Q) = -\frac{1}{2}\mathrm{Tr}[Q^{-1}(MCM' - \tilde{B}\tilde{\phi}' - \tilde{\phi}\tilde{B}' + \tilde{\phi}A\tilde{\phi}')] \tag{A.5.4}$$

where we have defined $\tilde{B} = MB$ and $\tilde{\phi} = M\phi$. Let us compute explicitly the terms in $G_1(F, Q)$ depending on F:

$$\begin{aligned}
\tilde{B}\tilde{\phi}' &= B_{11}(\mathcal{I} + F') - B_{12}F' \\
\tilde{\phi}\tilde{B}' &= (\mathcal{I} + F)B_{11}' - FB_{12}' \\
\tilde{\phi}A\tilde{\phi}' &= (\mathcal{I} + F)A_{11}(\mathcal{I} + F') - FA_{21}(\mathcal{I} + F') \\
&\quad - (\mathcal{I} + F)A_{12}F' + FA_{22}F'
\end{aligned}$$

Therefore, we need to solve $\nabla_F \overline{G}_1(F) = 0$, where:

$$\begin{aligned}
\overline{G}_1(F) = \mathrm{Tr}[Q^{-1}(&-B_{11}(\mathcal{I} + F') + B_{12}F' - (\mathcal{I} + F)B_{11}' + FB_{12}' \\
&+ (\mathcal{I} + F)A_{11}(\mathcal{I} + F') - FA_{21}(\mathcal{I} + F') - (\mathcal{I} + F)A_{12}F' + FA_{22}F')]
\end{aligned}$$

This can be done using eq. (A.5.1) and (A.5.2). One obtains:

$$\begin{aligned}
\nabla_F \overline{G}_1(F) = Q^{-1}[&-2(B_{11} - B_{12} - A_{11} + A_{12}) \\
&+ 2F(A_{11} + A_{22} - A_{21} - A_{12})]
\end{aligned} \tag{A.5.5}$$

and therefore:

$$\hat{F} = \Gamma\Theta^{-1} \tag{A.5.6}$$

Q.E.D.

A.6 Proof of Proposition 3

First, we solve $\nabla_{Q^{-1}} G_2(\hat{F}_r, Q) = 0$. We obtain:

$$\begin{aligned} &\nabla_{Q^{-1}} G_2(\hat{F}_r, Q) = \\ &= \nabla_{Q^{-1}} \left[-\frac{n}{2} \log |Q| - \frac{1}{2} \mathrm{Tr}(Q^{-1} \hat{\Upsilon}) \right] \\ &= \frac{n}{2} Q - \frac{1}{2} \hat{\Upsilon}' \end{aligned} \tag{A.6.1}$$

and therefore, since $\hat{\Upsilon}' = \hat{\Upsilon}$:

$$\hat{Q} = \frac{\hat{\Upsilon}}{n} \tag{A.6.2}$$

We now solve $\nabla_H G_3(\hat{F}_r, Q) = 0$. Note that, since H is diagonal, we can write:

$$\begin{aligned} &\nabla_H G_3(H) = \\ &= \nabla_H \left[-\frac{n}{2} \log |H| - \frac{1}{2} \mathrm{Tr}(H^{-1} \mathrm{diag}(\Lambda)) \right] \\ &= \frac{n}{2} H - \frac{1}{2} \Lambda \end{aligned} \tag{A.6.3}$$

and therefore:

$$\hat{H} = \frac{\mathrm{diag}(\Lambda)}{n} \tag{A.6.4}$$

Q.E.D.

A.7 Computation of lead-lag correlations in the MLA

In order to compute lead-lag correlations, we first compute the j-th order autocovariance matrix, which is defined as:

$$S_j = \mathbb{E}[\Delta X_t \Delta X'_{t-j}] \tag{A.7.1}$$

It can be evaluated from the estimated matrices $\hat{F}$ and $\hat{Q}$ as:

$$\hat{S}_j = \hat{F} \hat{S}_{j-1}, \quad j = 1, 2, \dots \tag{A.7.2}$$

where the covariance matrix $S_0 = \mathbb{E}[\Delta X_t \Delta X'_t]$ is estimated as:

$$\mathrm{vec}(\hat{S}_0) = (\mathcal{I}_{d^2} - \hat{F} \otimes \hat{F})^{-1} \mathrm{vec}(\hat{Q}) \tag{A.7.3}$$

see e.g. Hamilton (1994). Lead-lag correlations are finally obtained by normalizing the autocovariances with the diagonal elements of $\hat{S}_0$.

A.8 The univariate LLSD model

We show here how the time-varying parameters in H_t and D_t can be estimated separately using the univariate version of the LLSD model. If $n = 1$, model (5.1.3)-(5.1.4) becomes a univariate local level model with time-varying variances:

$$Y_{t,i} = X_{t,i} + \epsilon_{t,i}, \quad \epsilon_t \sim \mathrm{N}(0, \sigma^2_{t,i}) \tag{A.8.1}$$

$$X_{t+1,i} = X_{t,i} + \eta_{t,i}, \quad \eta_t \sim \mathrm{N}(0, \xi^2_{t,i}) \tag{A.8.2}$$

for $i = 1, \ldots, n$. This model is the same as the one introduced by Creal et al. (2008). The time-varying parameters are $\sigma^2_{t,i}$ and $\xi^2_{t,i}$. In order to recover positive estimates of the variance terms, we consider the two time-varying parameter vectors:

$$f_{t,i} = \begin{pmatrix} \log \sigma^2_{t,i} \\ \log \xi^2_{t,i} \end{pmatrix}, \quad \tilde{f}_{t,i} = \begin{pmatrix} \sigma^2_{t,i} \\ \xi^2_{t,i} \end{pmatrix} \tag{A.8.3}$$

which are related by the following link-function:

$$\tilde{f}_{t,i} = \mathcal{L}(f_{t,i}) = \begin{pmatrix} \exp f^1_{t,i} \\ \exp f^2_{t,i} \end{pmatrix} \tag{A.8.4}$$

The Kalman filter recursions are simply given as a special case of (2.4.5)-(2.4.10):

$$v_{t,i} = Y_{t,i} - x_{t,i}, \quad F_{t,i} = P_{t,i} + \tilde{f}^1_{t,i}, \quad K_{t,i} = \frac{P_{t,i}}{F_{t,i}} \tag{A.8.5}$$

$$x_{t|t,i} = x_{t,i} + K_{t,i} v_{t,i}, \quad P_{t|t,i} = P_{t,i}(1 - K_{t,i}) \tag{A.8.6}$$

$$x_{t+1,i} = x_{t|t,i}, \quad P_{t+1,i} = P_{t,i}(1 - K_{t,i}) + \tilde{f}^2_{t,i} \tag{A.8.7}$$

The conditional log-likelihood at time t is computed as:

$$l_{t,i} = -\frac{1}{2} \log 2\pi - \frac{1}{2} \left(\log |F_{t,i}| + \frac{v^2_{t,i}}{F_{t,i}} \right) \tag{A.8.8}$$

In case at time t observation $Y_{t,i}$ is missing, the above Kalman filter recursions and the expression (A.8.8) of the conditional log-likelihood are still valid, provided that one takes $K_{t,i} = 0$, $v_{t,i} = 0$ and $F_{t,i} = 1$.

The vector $f_{t,i}$ is driven by the scaled score of the predictive likelihood:

$$s_{t,i} = S_{t,i} \nabla_{t,i} \tag{A.8.9}$$

where $\nabla_{t,i} = (\partial \log l_{t,i})/(\partial f'_{t,i})$ and $S_{t,i} = (\mathcal{I}^i_{t|t-1})^{-1/2}$, $\mathcal{I}^i_{t|t-1} = \mathrm{E}_{t-1}[\nabla_{t,i} \nabla'_{t,i}]$. Denoting by $J_{\mathcal{L}}$ the Jacobian matrix of the link function

$$J_{\mathcal{L}} = \begin{pmatrix} \exp f^1_{t,i} & 0 \\ 0 & \exp f^2_{t,i} \end{pmatrix} \tag{A.8.10}$$

we have:

$$\nabla_{t,i} = J_{\mathcal{L}} \tilde{\nabla}_{t,i}, \quad \mathcal{I}^i_{t|t-1} = J_{\mathcal{L}} \tilde{\mathcal{I}}^i_{t|t-1} J_{\mathcal{L}} \tag{A.8.11}$$

where $\tilde{\nabla}_{t,i} = (\partial \log l_{t,i})/(\partial \tilde{f}'_{t,i})$ and $\tilde{\mathcal{I}}^i_{t|t-1} = \mathrm{E}_{t-1}[\tilde{\nabla}_{t,i} \tilde{\nabla}'_{t,i}]$. The latter are computed as special cases of (5.2.2) and (5.2.3):

$$\tilde{\nabla}_{t,i} = \frac{1}{2} \left[\dot{F}_{t,i} \left(\left(\frac{v_{t,i}}{F_{t,i}} \right)^2 - \frac{1}{F_{t,i}} \right) - 2 \dot{v}'_{t,i} \frac{v_{t,i}}{F_{t,i}} \right] \tag{A.8.12}$$

$$\tilde{I}_{t|t-1,i} = \frac{1}{2} \left[\frac{\dot{F}'_{t,i} \dot{F}_{t,i}}{F^2_t} + 2 \frac{\dot{v}'_{t,i} \dot{v}_{t,i}}{F_t} \right] \tag{A.8.13}$$

Note that now dotted variables denote derivatives with respect to $\tilde{f}_t$. Similarly to the multivariate case, the latter are computed by deriving the Kalman filter recursions (A.8.5), (A.8.6) and (A.8.7). One obtains:

$$\dot{v}_{t,i} = -\dot{x}_{t,i} \tag{A.8.14}$$

$$\dot{F}_{t,i} = \dot{P}_{t,i} + (1, 0) \tag{A.8.15}$$

$$\dot{K}_{t,i} = \frac{\dot{P}_{t,i} - K_{t,i} \dot{F}_{t,i}}{F_{t,i}} \tag{A.8.16}$$

$$\dot{x}_{t+1,i} = \dot{x}_{t,i}(1 - K_{t,i}) + \dot{K}_{t,i} v_{t,i} \tag{A.8.17}$$

$$\dot{P}_{t+1,i} = \dot{P}_{t,i}(1 - K_{t,i}) - P_{t,i} \dot{K}_{t,i} + (0, 1) \tag{A.8.18}$$

Therefore, the score-driven update equation in the univariate LLSD is given by:

$$f_{t+1,i} = \omega_i + A_i s_{t,i} + B_i f_{t,i} \tag{A.8.19}$$

where ω_i is a two-dimensional column vector, while A_i and B_i are two-dimensional square matrices. They are part of the vector Θ_i of static parameters that is estimated by maximum likelihood. Constraints on ω_i, A_i and B_i depend on the specific data at hand and will be discussed in detail in our empirical application.

As in the correlation model, the likelihood is maximized numerically. Once all of the n univariate models have been estimated, one can construct the two matrices H_t and D_t as:

$$H_t = \begin{pmatrix} \exp f_{t,1}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp f_{t,n}^1 \end{pmatrix}, \quad D_t = \begin{pmatrix} (\exp f_{t,1}^2)^{1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (\exp f_{t,n}^2)^{1/2} \end{pmatrix} \tag{A.8.20}$$

A.9 Computation of ∇_t and $\mathcal{I}_{t|t-1}$ in the SHAR model

It is convenient to introduce the auxiliary vector of time-varying parameters:

$$\tilde{f}_t = \left(\beta_{0,t}, \beta_{1,t}, \beta_{2,t}, \beta_{3,t}, q_t\right)' \tag{A.9.1}$$

The latter is related to f_t by the following link-function:

$$\tilde{f}_t = \mathcal{L}(f_t) = \left[f_t^1, f_t^2, f_t^3, f_t^4, \exp(f_t^5)\right]' \tag{A.9.2}$$

The Jacobian of the transformation is:

$$J_{\mathcal{L}} = \frac{\partial \tilde{f}_t}{\partial f_t'} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \exp(f_t^5) \end{pmatrix} \tag{A.9.3}$$

Note that, using the chain rule, ∇_t and $\mathcal{I}_{t|t-1}$ can be expressed as:

$$\nabla_t = J_{\mathcal{L}} \tilde{\nabla}_t, \quad \mathcal{I}_{t|t-1} = J_{\mathcal{L}} \tilde{\mathcal{I}}_{t|t-1} J_{\mathcal{L}} \tag{A.9.4}$$

where:

$$\tilde{\nabla}_t = \left[\frac{\partial \log p(RV_t|\tilde{f}_t, \mathcal{B}_{t-1}, \Theta)}{\partial \tilde{f}_t'}\right]', \quad \tilde{\mathcal{I}}_{t|t-1} = \mathrm{E}[\tilde{\nabla}_t \tilde{\nabla}_t'] \tag{A.9.5}$$

Thus, it is simpler to compute $\tilde{\nabla}_t$ and $\tilde{\mathcal{I}}_{t|t-1}$ and then using eq. (A.9.5) to recover ∇_t and $\mathcal{I}_{t|t-1}$. By direct differentiation of the conditional log-likelihood (6.1.23), we have:

$$\tilde{\nabla}_t = \begin{pmatrix} \frac{\eta_t}{q_t} \\ \frac{\eta_t}{q_t} RV_{t-1} \\ \frac{\eta_t}{q_t} RV_{t-2|t-n_w-1} \\ \frac{\eta_t}{q_t} RV_{t-n_w-2|t-n} \\ -\frac{1}{2q_t} + \frac{1}{2}\frac{\eta_t^2}{q_t^2} \end{pmatrix} \tag{A.9.6}$$

where $\eta_t = RV_t - \mu_{t|t-1}$. The information matrix $\tilde{\mathcal{I}}_{t|t-1}$ is easily computed by noting that $\mathrm{E}_{t|t-1}[\eta_t^2] = q_t$ and $\mathrm{E}_{t|t-1}[\eta_t^4] = 3q_t^2$. We have:

$$\tilde{\mathcal{I}}_{t|t-1} = \frac{1}{q_t}\begin{pmatrix} 1 & RV_{t-1} & RV_{t-2|t-n_w-1} & RV_{t-n_w-2|t-n} & 0 \\ RV_{t-1} & RV_{t-1}^2 & RV_{t-1}RV_{t-2|t-n_w-1} & RV_{t-1}RV_{t-n_w-2|t-n} & 0 \\ RV_{t-2|t-n_w-1} & RV_{t-1}RV_{t-2|t-n_w-1} & RV_{t-2|t-n_w-1}^2 & RV_{t-2|t-n_w-1}RV_{t-n_w-2|t-n} & 0 \\ RV_{t-n_w-2|t-n} & RV_{t-1}RV_{t-n_w-2|t-n} & RV_{t-2|t-n_w-1}RV_{t-n_w-2|t-n} & RV_{t-n_w-2|t-n}^2 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2q_t} \end{pmatrix} \tag{A.9.7}$$

A.10 Computation of $\dot{v}_t$ and $\dot{F}_t$ in the SHARK model

As done with the SHAR, in order to obtain positive variances, we introduce the auxiliary vector of time-varying parameters:

$$\tilde{f}_t = \mathcal{L}(f_t) = \left[f_t^1, f_t^2, f_t^3, f_t^4, \exp(f_t^5)\right]' \tag{A.10.1}$$

The Jacobian $J_{\mathcal{L}}$ is the same as in eq. (A.9.3) and eq. (A.9.4) still hold. Consequently, we first compute $\tilde{\nabla}_t$ and $\tilde{\mathcal{I}}_{t|t-1}$ defined as in eq. (A.9.5). They have the same form as in eq. (6.1.32) but now $\dot{v}_t = \partial v_t/\partial \tilde{f}_t'$ and $\dot{F}_t = \partial \mathrm{vec}(F_t)/\partial \tilde{f}_t'$ denote derivatives with respect to $\tilde{f}_t$. Following Delle Monache et al. (2016), they can be computed through the following set of recursions:

$$\dot{v}_t = -Z\dot{a}_t \tag{A.10.2}$$

$$\dot{F}_t = (Z \otimes Z)\dot{P}_t \tag{A.10.3}$$

$$\dot{a}_{t+1} = \dot{c}_t + T_t\dot{a}_t + (a_t' \otimes \mathbb{I}_n)\dot{T}_t + (v_t' \otimes \mathbb{I}_n)\dot{K}_t + K_t\dot{v}_t \tag{A.10.4}$$

$$\begin{aligned}\dot{P}_{t+1} = {} & (T_t \otimes T_t)\dot{P}_t + (\mathbb{I}_n \otimes T_tP_t)C_{n,n}\dot{T}_t + (T_tP_t' \otimes I_n)\dot{T}_t - (K_tZP_t \otimes \mathbb{I}_n)\dot{T}_t \\ & - (K_tZ \otimes T_t)\dot{P}_t - (\mathbb{I}_n \otimes T_tP_tZ')C_{n,1}\dot{K}_t + \dot{Q}_t\end{aligned} \tag{A.10.5}$$

$$\dot{K}_t = (F_t^{-1}ZP_t \otimes \mathbb{I}_n)\dot{T}_t + (F_t^{-1}Z \otimes T_t)\dot{P}_t - K_tF_t^{-1}\dot{F}_t \tag{A.10.6}$$

where $\dot{c}_t = \frac{\partial a_t}{\partial \tilde{f}_t'}$ is a $n \times 5$ matrix and $\dot{T}_t = \frac{\partial \mathrm{vec}T_t}{\partial \tilde{f}_t'}$, $\dot{Q}_t = \frac{\partial \mathrm{vec}Q_t}{\partial \tilde{f}_t'}$ are $n^2 \times 5$. They are selection matrix of the form:

$$\{\dot{A}_t\}_{i,j} = \begin{cases} 1, & \text{if } \{\mathrm{vec}A_t\}_i = \tilde{f}_t^j \\ 0, & \text{else} \end{cases} \tag{A.10.7}$$

where A_t generically denotes c_t, T_t and Q_t.

A.11 Proof of proposition 4

The proof is made by induction on j. First, let us prove the formula for $j = 2$. We have:

$$\begin{aligned}\mathrm{E}_t[\exp(Z\overline{RV}_{t+2})] &= \mathrm{E}_t[\exp(Z(c_l + T_l\overline{RV}_{t+1} + \overline{\eta}_{t+1}))] \\ &= \mathrm{E}_t[\exp(Z(c_l + T_l(c_l + T_l\overline{RV}_t + \overline{\eta}_t) + \overline{\eta}_{t+1}))] \\ &= \exp\left[Z((\mathbb{I}_n + T_l)c_l + T^2\overline{RV}_t) + \frac{1}{2}Z(Q_l + T_lQ_lT_l')Z'\right]\end{aligned}$$

where we have used the moment generating function of the multivariate normal distribution in the last line. So the formula is true for $j = 2$. Now let us assume that it is true for $j = k$ and prove that it is also true for $j = k+1$:

$$\begin{aligned}
&\mathrm{E}_t[\exp(Z\overline{RV}_{t+k+1})] = \\
&= \mathrm{E}_t[\exp(Z(c_l + T_l\overline{RV}_{t+k} + \overline{\eta}_{t+k}))] \\
&= \exp(Zc_l)\mathrm{E}_t[\exp(ZT_l\overline{RV}_{t+k})]\mathrm{E}_t[\exp(Z\overline{\eta}_{t+l})] \\
&= \exp(Zc_l)\exp\left\{Z\left[(T_l + \cdots + T_l^k)c_l + T_l^{k+1}\overline{RV}_t^l\right] + \frac{1}{2}Z\left[T_lQ_lT_l' + \cdots + T^kQ_l(T^k)'\right]Z'\right\}\exp(ZQ_lZ')
\end{aligned}$$

where we have used the induction hypothesis to compute $\exp(ZT_l\overline{RV}_{t+k})$. Upon multiplication of the three terms we get:

$$\begin{aligned}
&\mathrm{E}_t[\exp(Z\overline{RV}_{t+k+1})] = \\
&= \exp\left\{Z\left[(\mathbb{I}_n + \cdots + T_l^k)c_l + T_l^j\overline{RV}_t^{k+1}\right] + \frac{1}{2}Z\left[Q_l + \cdots + T^kQ_l(T^k)'\right]Z'\right\}
\end{aligned}$$

which is the formula in Proposition 1 for $j = k + 1$

Q.E.D.

A.12 Proof of proposition 5

The proof follows exactly the same steps as those in Proposition 4Z Let us prove the formula for $j = 2$. We have:

$$\begin{aligned}
\mathrm{E}_t[\exp(Z\alpha_{t+2})] &= \mathrm{E}_t[\exp(Z(c + T\alpha_{t+1} + \eta_{t+1}))] \\
&= \exp(Zc)\mathrm{E}_t[\exp(ZT\alpha_{t+1})]\mathrm{E}_t[\exp(Z\eta_{t+1})]
\end{aligned}$$

where the last step follows from independence between α_{t+1} and η_{t+1}. Using the fact that $\alpha_{t+1}|\mathcal{F}_t \sim \mathrm{N}(a_{t+1}, P_{t+1})$, we have:

$$\begin{aligned}
&\mathrm{E}_t[\exp(Z\alpha_{t+2})] = \\
&= \exp(Zc)\exp\left(ZTa_{t+1} + \frac{1}{2}ZTP_{t+1}T'Z'\right)\exp(ZQZ') \\
&= \exp\left[Z(c + Ta_{t+1}) + \frac{1}{2}Z(TP_{t+1}T' + Q)Z'\right]
\end{aligned}$$

which is the formula in Proposition 2 for $j = 2$. We now assume that the formula is true for $j = k$ and prove that it is also true for $j = k + 1$:

$$\begin{aligned}
&\mathrm{E}_t[\exp(Z\alpha_{t+k+1})] = \\
&= \mathrm{E}_t[\exp(Z(c + T\alpha_{t+k} + \eta_{t+k}))] \\
&= \exp(Zc)\mathrm{E}_t[\exp(ZT\alpha_{t+k})]\mathrm{E}_t[\exp(Z\eta_{t+k})]
\end{aligned}$$

The last term is equal to $\exp(\frac{1}{2}ZQZ')$. Based on the induction hypothesis, the second term is given by:

$$\begin{aligned}
\mathrm{E}_t[\exp(ZT\alpha_{t+k})] = \exp\big\{&Z\left[(T + \cdots + T^{k-1})c + T^ka_{t+1}\right] + \\
&\frac{1}{2}Z\left[T^kP_{t+1}(T^k)' + TQT' + \cdots + T^{k-1}Q(T^{k-1})'\right]Z'\Big\}
\end{aligned}$$

By summing the three exponents we end up with:

$$\mathrm{E}_t[\exp(ZT\alpha_{t+k})] = \exp\left\{Z\left[(\mathbb{I}_n + \cdots + T^{k-1})c + T^k a_{t+1}\right] + \frac{1}{2}Z\left[T^k P_{t+1}(T^k)' + Q + \cdots + T^{k-1}Q(T^{k-1})'\right]Z'\right\}$$

which is the formula in Proposition 2 for $j = k + 1$

Q.E.D.