# Volatility forecasting using global stochastic financial trends extracted from non-synchronous data

Lyudmila Grigoryeva[1] , Juan-Pablo Ortega[2,3], and Anatoly Peresetsky[4]

**Abstract**

A method based on various linear and nonlinear state space models used to extract global stochastic financial trends (GST) out of non-synchronous financial data is introduced. These models are constructed in order to take advantage of the intraday arrival of closing information coming from different international markets so that volatility description and forecasting is improved. A set of three major asynchronous international stock market indices is considered in order to empirically show that this forecasting scheme is capable of significant performance gains when compared to standard parametric models like the dynamic conditional correlation (DCC) family.

*Keywords:* Multivariate volatility modeling and forecasting, global stochastic trend, extended Kalman filter, dynamic conditional correlations (DCC), non-synchronous data.

## 1. Introduction

Many frameworks for the description of financial returns have as their first building block a factor model of the form

$$r_t = \alpha + \beta y_t + u_t \quad \text{with} \quad \{u_t\} \sim \mathrm{WN}(0, \sigma^2),$$

where the instantaneous returns $r_t$ at time $t$ of individual assets are presented as an affine function of a common factor $y_t$ additively perturbed with a stochastic stationary white noise process $\{u_t\}$ (not necessarily normal) with variance $\sigma^2$. This factor usually accounts for a common market feature to which all the assets under study are exposed. Consequently, this functional dependence allows to determine, for a given asset return $r_t$, how much of it has to do with the market situation (through the coefficient $\beta$, which is a function of the correlation between $r_t$ and $y_t$) and how much comes from an idiosyncratic

[1]Corresponding author. Department of Mathematics and Statistics. Universität Konstanz. Box 146. D-78457 Konstanz. Germany. `Lyudmila.Grigoryeva@uni-konstanz.de`

[2]Universität Sankt Gallen. Bodanstrasse 6. CH-9000 Sankt Gallen. Switzerland. `Juan-Pablo.Ortega@unisg.ch`

[3]Centre National de la Recherche Scientifique (CNRS). France.

[4]National Research University Higher School of Economics, Myasnitskaya str. 20. Moscow. Russia. `aperesetsky@hse.ru`

perturbation $u_t$ specifically related to the individual asset. In particular applications of this model, the factor values $y_t$ are sometimes computed by using an index constructed out of a set of assets that represent the class to which $r_t$ is naturally associated. An alternative to this approach consists of treating the common factor values as a non-observable variable and of extracting them using observed individual returns via a Kalman-type state-space model. This direction has already been profusely explored in the literature. In the early available works (see for instance Jeon and Chiang (1991); Kasa (1992); Chung and Liu (1994); Siklos and Ng (2001); Rangvid (2001); Rangvid and Sorensen (2002); Phengpis and Apilado (2004)) the authors consider low sampling frequencies in order to be able to neglect asynchronicity issues. In several more recent references Dungey et al. (2001); Choudhry et al. (2007); Chang et al. (2009); Lucey and Muckley (2011); Cartea and Karyampas (2011); Bae and Kim (2011); Felices and Wieladek (2012); Bentes (2015) daily quoted data are used but always using a synchronized approach.

In this work we use a different point of view introduced in Korhonen and Peresetsky (2013) and later on extended in Durdyev and Peresetsky (2014); Peresetsky and Yakubov (2015), in which the market returns are thought of as a non-observable global stochastic trend (GST) whose value is ruled by the arrival of information coming from different local markets. In this framework, the returns of the GST are estimated several times per day at time points that are synchronized with the closing times of the markets that are assumed to drive it. This approach is implemented by setting a state-space model in which the observation equation writes the different observable individual market returns that we are interested in as a stochastic perturbation of an affine function of the estimated GST return accumulated during the 24 hours that precede this quote. It is assumed that the observed returns are those that drive the GST and hence its returns are estimated as many times per day as different closing times are included in the list of markets considered.

This point of view has been studied in Korhonen and Peresetsky (2013) using three different setups, namely: three world indices (NIKKEI 225, MICEX, S&P 500) with three different closing times, five world indices (NIKKEI 225, MICEX, DAX, PX, S&P 500) with four different closing times, and ten world indices (NIKKEI 225, HSI, SENSEX, MICEX, DAX, PX, FTSE 100, IBOV, DJI, S&P 500) with seven different closing times. The estimates of the GST obtained in these different situations are remarkably similar. The robustness that these results indicate allowed the authors to identify, for each market, the relative importance of local with respect to global news in stock prices formation.

The main goal of our work is modifying this approach in order to make it amenable to volatility forecasting and to prove the pertinence of the resulting method when compared to more standard families of models designed to specifically carry out this task that do not take advantage of the intraday arrival of information. The rationale behind this attempt is that the error inherent to the filtering and forecasting of an unobserved variable like the GST is compensated by the more frequent information updates, that the use of asynchronous information carries in its wake.

Since the models introduced in Korhonen and Peresetsky (2013); Peresetsky and Yakubov (2015) are intrinsically homoscedastic, they are not appropriate to handle financial volatility modeling and forecasting. The heteroscedastic generalization needed for this purpose can be naturally implemented by using two different approaches. The simplest one consists of using the linear state-space approach in Korhonen and Peresetsky (2013) in a first step to estimate the GST and to subsequently model the volatility and conditional correlation of the resulting global trend and idiosyncratic term using an adapted multivariate correlation model; for this purpose, we consider in this work adapted scalar and non-scalar versions of the dynamic conditional correlation (DCC) model introduced in Engle (2002); Tse and Tsui (2002). The non-scalar models are estimated using the techniques introduced in Chrétien and Ortega (2014); Bauwens et al. (2016). The adjustments of these standard models for the handling of the GST are implemented at the level of the so called "deGARCHing" or "first estimation step" in which a model for the conditional variances of the assets of interest is chosen; in our context, we put to work in this step two different GARCH-type models that take into account in their specification the chronology with which the different intraday trend returns are quoted. A more sophisticated approach that we also study is the inclusion of the heteroscedasticity assumption on the GST returns directly in the formulation of the state-space model by using a GARCH-type and GST-adapted prescription of the type that we just described. The main complication that arises in this setup is the nonlinearity of the resulting modeling scheme that we handle using the extended Kalman filter (EKF) (see Durbin and Koopman (2012) and references therein).

In the two approaches that we just described we proceed in two stages, in the first one we filter the GST and eventually its conditional variance and, in a second one, we model conditional correlations. This way to operate is admittedly suboptimal but provides good empirical results in practice that outperform those obtained with a more elegant one-shot approach that, as we experienced in unreported simulations, faces important numerical estimation problems that make it less advisable.

The paper is organized as follows. In Section 2.1 we explain in detail the linear and nonlinear state-space models that we propose. We give details on how they handle the asynchronous character of the observable data and prove rigorous sufficient conditions that ensure their proper identification. Section 2.2 contains details on the Kalman filter based model estimation techniques that we use in the paper, as well as on the model specifications for the conditional variances incorporated in the nonlinear state-space model, together with the positivity and stationarity constraints that need to be imposed at the time of estimation. The GST-based volatility forecasting scheme is described in Section 4. Section 5 contains an empirical study using the adjusted closing values of three major indices (NIKKEI 225, FTSE 100, and S&P 500) that are quoted at different times due to the time zones in which they are geographically based. In this experiment, we use the model confidence set (MCS) approach of Hansen et al. (2003, 2011) and we implement it with loss functions constructed with the conditional covariance matrices implied by the different models under consideration. *The results show that the proposed forecasting scheme exhibits excellent and statistically significant performance improvements when compared to the use of standard multivariate parametric correlation models that ignore non-synchronicity.* Even though no simulation study has been carried out for the method that we propose, we have tested the robustness of the good empirical results by using another estimation period that does not include the Fall 2008 volatility events and an out-of-sample forecasting period that comprises the Great Recession (results reported on the supplementary material section) and by using another data set (not reported) in which the index FTSE 100 has been replaced by the Russian MICEX. Section 6 concludes the paper.

A Supplementary Material section is available that contains details about the notation and the conventions followed in the presentation and the proofs of various technical results. Additionally, the above mentioned replicate of the empirical study in Section 5 has been added in which the estimation period does not include the Fall 2008 volatility events and the out-of-sample forecasting period comprises the Great Recession; this experiment aims at illustrating the robustness of our empirical results with respect to the estimation period used and the pertinence in some situations of the nonlinear state-space model.

## 2. State-space models for the global stochastic trend

We start by recalling the description of the global stochastic trend as the state variable in a state-space model, as it was introduced in Korhonen and Peresetsky (2013). In order to keep the presentation simple, we consider only three different non-synchronous assets and our subsequent empirical analysis takes place in this framework. The generalization to more assets and quoting times is straightforward. Let $\mathbf{r}_t \in \mathbb{R}^3$ be a vector containing three non-synchronous stock market returns (typically based on adjusted closing prices and computed over a full day) quoted at different times of the same day $t \in \mathbb{N}$. The different intraday quoting times have typically to do with lags in the closing times of the different markets. The intraday moments of time $t_i$, $i \in \{1, 2, 3\}$ of the given day $t$ at which the components of $(r_{1,t}, r_{2,t}, r_{3,t})^\top$ of the daily returns vector $\mathbf{r}_t$ become available are labeled as $t_i := 3(t-1) + i$, $t \in \mathbb{N}$.

We now assume the existence of an underlying and non-observable global stochastic trend (GST) and we denote by $s_{t_i}$, $i \in \{1, 2, 3\}$, its intraday log-values for the given calendar date $t \in \mathbb{N}$. We now define $\boldsymbol{\varepsilon}_t \in \mathbb{R}^3$ as the vector that contains the intra-day stochastic trend log-return components of a given calendar day $t$, that is,

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix} := \begin{pmatrix} s_{t_1} - s_{(t-1)_3} \\ s_{t_2} - s_{t_1} \\ s_{t_3} - s_{t_2} \end{pmatrix}. \tag{2.1}$$

Figure 1 represents the chronology of the times at which the different returns are computed together with the associated information sets (see an explanation later on in Section 4). Following the factor model scheme discussed in the introduction, for every $t \in \mathbb{Z}$ we write the log-returns of the components of $\mathbf{r}_t$ as excess returns with respect to an affine function of the entries of the vector $\mathbf{y}_t \in \mathbb{R}^3$, which is constructed with the daily GST returns computed at the moments in which the components of $\mathbf{r}_t$ are quoted. More specifically $\mathbf{y}_t := (s_{t_1} - s_{(t-1)_1}, s_{t_2} - s_{(t-1)_2}, s_{t_3} - s_{(t-1)_3})^\top$ and

$$r_{i,t} = \alpha_i + \beta_i y_{i,t} + u_{i,t}, \;\; i = \{1, 2, 3\}, \;\; t \in \mathbb{Z}, \tag{2.2}$$

with the regression intercepts $\alpha_i \in \mathbb{R}$, $i = \{1, 2, 3\}$, and the parameters $\boldsymbol{\beta} := (\beta_1, \beta_2, \beta_3)^\top \in \mathbb{R}^3$. For the time being, in this relation we only assume that the residuals $\mathbf{u}_t$ are serially uncorrelated (they are a white noise) with mean zero and unconditional diagonal covariance matrix $\Sigma_{\mathbf{u}} \in \mathbb{S}_3^+$, that is $\{\mathbf{u}_t\} \sim \mathrm{WN}\left(\mathbf{0}_3, \Sigma_{\mathbf{u}}\right)$, $\Sigma_{\mathbf{u}} := \mathrm{diag}\left(\sigma_{u,1}^2, \sigma_{u,2}^2, \sigma_{u,3}^2\right)$ with $\sigma_{u,1}, \sigma_{u,2}, \sigma_{u,3} \in \mathbb{R}^+$.

Using the definition (2.1), the returns $\mathbf{r}_t$ in (2.2) can be written in terms of the non-observable GST returns in the preceding twenty-four hours as

$$\mathbf{r}_t = \boldsymbol{\alpha} + B\boldsymbol{e}_t + \mathbf{u}_t, \tag{2.3}$$

with $\boldsymbol{e}_t := \left(\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t}, \varepsilon_{2,t-1}, \varepsilon_{3,t-1}\right)^\top$, $\boldsymbol{\alpha} \in \mathbb{R}^3$, $\{\mathbf{u}_t\} \sim \mathrm{WN}\left(\mathbf{0}_3, \Sigma_{\mathbf{u}}\right)$ as in (2.2), and the matrix $B \in \mathbb{M}_{5,3}$ of the form

$$B := \begin{pmatrix} \beta_1 & 0 & 0 & \beta_1 & \beta_1 \\ \beta_2 & \beta_2 & 0 & 0 & \beta_2 \\ \beta_3 & \beta_3 & \beta_3 & 0 & 0 \end{pmatrix}. \tag{2.4}$$

In order to determine the values of the GST, we consider (2.3) as the observation equation of several linear and nonlinear state-space models. These particular models are estimated to be subsequently used

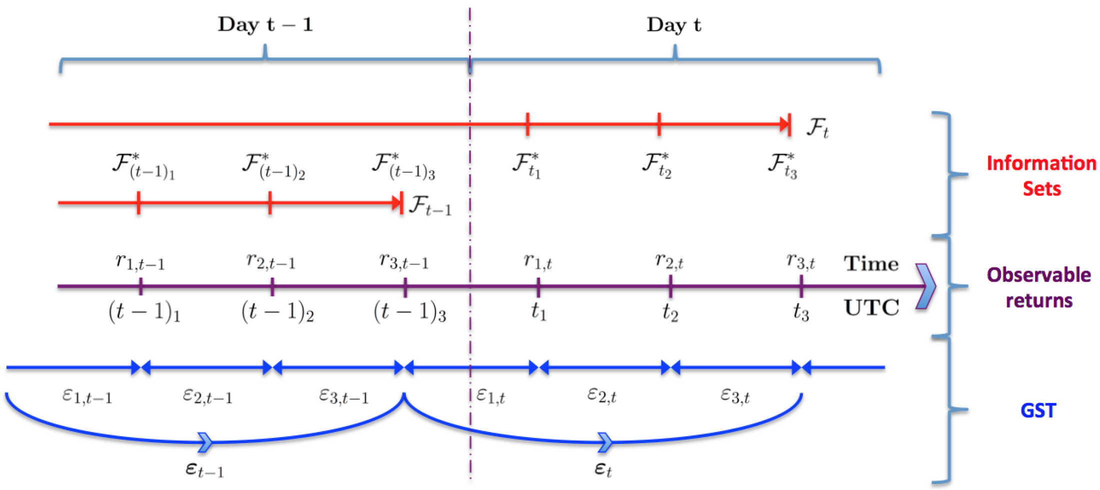


Figure 1: Diagram representing the different variables, time labels, and information sets used in the study as well as their chronology.

for volatility forecasting.

### *2.1. The linear and nonlinear state-space models*

**The linear state-space model.** The first model that we present in this subsection is identical to the one originally considered in Korhonen and Peresetsky (2013):

$$\begin{cases} \mathbf{r}_t = \boldsymbol{\alpha} + B\boldsymbol{e}_t + \mathbf{u}_t, & \{\mathbf{u}_t\} \sim \text{WN}\left(\mathbf{0}_3, \Sigma_{\mathbf{u}}\right), \text{with } \Sigma_{\mathbf{u}} := \text{diag}\left(\sigma^2_{u,1}, \sigma^2_{u,2}, \sigma^2_{u,3}\right), \quad (2.5\text{a}) \\ \boldsymbol{e}_t = T\boldsymbol{e}_{t-1} + R\mathbf{v}_{t-1}, & \{\mathbf{v}_t\} \sim \text{WN}(\mathbf{0}_3, \mathbb{I}_3), \quad (2.5\text{b}) \end{cases}$$

where $\boldsymbol{e}_t \in \mathbb{R}^5$, $\boldsymbol{\alpha} \in \mathbb{R}^3$, the matrix $B \in \mathbb{M}_{3,5}$ is provided in (2.4), the matrices $R \in \mathbb{M}_{5,3}$ and $T \in \mathbb{M}_5$ are given by

$$R := \begin{pmatrix} \sigma_{v,1} & 0 & 0 \\ 0 & \sigma_{v,2} & 0 \\ 0 & 0 & \sigma_{v,3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad T := \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \tag{2.6}$$

with $\sigma_{u,i}, \sigma_{v,i} \in \mathbb{R}^+$ for $i \in \{1, 2, 3\}$. We emphasize that (2.5a)-(2.5b) constitutes a linear state-space model in which the dynamical behavior of the GST is prescribed by the corresponding state equation and where the observation equation establishes a relation between the time evolution of the GST and the observed returns.

**The nonlinear state-space model.** The dynamic specification (2.5b) does not introduce any dependence between the components $(\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})$ of the GST $\boldsymbol{\varepsilon}_t$, even though this feature is empirically observed. This has motivated the introduction in Durdyev and Peresetsky (2014); Peresetsky and Yakubov (2015) of a VAR-type prescription for the dynamics of $\boldsymbol{\varepsilon}_t$ that allows for correlation between its components while preserving the linearity of the state-space model. Since the focus of this work is volatility forecasting, we introduce instead the dependence between the components of $\boldsymbol{\varepsilon}_t$ via a specific nonlinear dynamic prescription for their conditional variances.

There are many different approaches that can be taken in order to implement this strategy. The most natural one consists in building into the state-space model the entire conditional covariance dynamics of both the GST $\boldsymbol{\varepsilon}_t$ and the residuals $\mathbf{u}_t$. In unreported numerical experiments, we observed that the complexity of the resulting specification makes its estimation difficult to implement. This limitation makes advisable a two-steps procedure in which the nonlinear state-space model prescribes the dynamics of the conditional variances and then conditional covariances are handled separately in a second step.

Consider the following nonlinear state-space model:

$$
\begin{cases}
\mathbf{r}_t = \boldsymbol{\alpha} + B\boldsymbol{e}_t + \mathbf{u}_t, & \{\mathbf{u}_t\} \sim \mathrm{WN}\left(\mathbf{0}_3, \Sigma_{\mathbf{u}}\right), \ \ \Sigma_{\mathbf{u}} := \mathrm{diag}\left(\sigma^2_{u,1}, \sigma^2_{u,2}, \sigma^2_{u,3}\right), \quad (2.7a) \\
\boldsymbol{e}_t = T\boldsymbol{e}_{t-1} + R_{t-1}(\underline{\boldsymbol{e}_{t-1}})\mathbf{v}_{t-1}, & \{\mathbf{v}_t\} \sim \mathrm{WN}(\mathbf{0}_3, \mathbb{I}_3), \quad (2.7b)
\end{cases}
$$

where $\boldsymbol{e}_t \in \mathbb{R}^5$, $\underline{\boldsymbol{e}_t} := \{\boldsymbol{e}_t, \boldsymbol{e}_{t-1}, \ldots, \boldsymbol{e}_0\}$, $\boldsymbol{\alpha} \in \mathbb{R}^3$, the matrices $B \in \mathbb{M}_{3,5}$ and $T \in \mathbb{M}_5$ are provided in (2.4) and (2.6), respectively, and $\sigma_{u,1}, \sigma_{u,2}, \sigma_{u,3} \in \mathbb{R}^+$. The main difference with the linear model (2.5a)-(2.5b) consists in the particular nonlinear specification of the matrix $R_{t-1}(\underline{\boldsymbol{e}_{t-1}}) \in \mathbb{M}_{5,3}$ whose structure is identical to $R$ in (2.6) but the non-zero entries are defined as $(R_{t-1}(\underline{\boldsymbol{e}_{t-1}}))_{ii} = \sigma_{i,t}(\underline{\boldsymbol{e}_{t-1}})$, $i \in \{1, 2, 3\}$. This specification of $R_{t-1}(\underline{\boldsymbol{e}_{t-1}})$ allows for a dynamic description of the conditional variances of the state variables (components of the GST) that we are ultimately interested in forecasting. We notice here that particular functional dependences for $\sigma_{i,t}(\underline{\boldsymbol{e}_{t-1}})$, $i \in \{1, 2, 3\}$ are a matter of choice. In this paper we propose two specifications which are consistent with the quoting chronology of the components of the GST $\boldsymbol{\varepsilon}_t$.

- **Model 1 for the conditional variances in the nonlinear state-space model.** In this first model we define recursively the values $\sigma_{i,t}(\underline{\boldsymbol{e}_{t-1}})$, $i \in \{1, 2, 3\}$, using a GARCH-type functional dependence adapted to the chronology of the GST components in the following way

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \varepsilon_{3,t-1}^2, \tag{2.8}$$

$$\sigma_{2,t}^2 = a_2 + \delta_2 \sigma_{1,t}^2, \tag{2.9}$$

$$\sigma_{3,t}^2 = a_3 + \delta_3 \sigma_{2,t}^2. \tag{2.10}$$

In order to insure the positivity of the elements $\sigma_{i,t}^2$, the model parameters are required to satisfy the constraints $\gamma_1 \geq 0$ and $a_i > 0$, $\delta_i \geq 0$, for all $i \in \{1,2,3\}$. The sufficient conditions for the stationarity of the process can be obtained using statements in Gouriéroux (1997) that result in the nonlinear inequality constraint $\delta_2\delta_3(\delta_1 + \gamma_1) < 1$ (see Supplementary material for details).

• **Model 2 for the conditional variances in the nonlinear state-space model.** Based on the same arguments that we used for Model 1, we consider another GARCH-type variant for the functions $\sigma_{i,t}(\underline{\boldsymbol{e}_{t-1}})$, $i \in \{1,2,3\}$, that determine the nonlinear state-space model (2.7a)-(2.7b) by allowing this time the possibility of autoregressive behavior in the volatilities and in the components of the GST. We set:

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \varepsilon_{3,t-1}^2 + \rho_1 \sigma_{1,t-1}^2 + \tau_1 \varepsilon_{1,t-1}^2, \tag{2.11}$$

$$\sigma_{2,t}^2 = a_2 + \delta_2 \sigma_{1,t}^2 + \rho_2 \sigma_{2,t-1}^2 + \tau_2 \varepsilon_{2,t-1}^2, \tag{2.12}$$

$$\sigma_{3,t}^2 = a_3 + \delta_3 \sigma_{2,t}^2 + \rho_3 \sigma_{3,t-1}^2 + \tau_3 \varepsilon_{3,t-1}^2, \tag{2.13}$$

where, again, we ensure positivity by requiring that $\gamma_1 \geq 0$ and $a_i > 0$, $\delta_i, \rho_i, \tau_i \geq 0$, for all $i \in \{1,2,3\}$. Two sets of sufficient stationarity conditions are discussed in detail in the Supplementary Material section. We provide here the one which consists of the three inequalities

$$\begin{cases} (\rho_1 + \tau_1)(1 + \delta_2(1 + \delta_3)) < 1, & \text{(2.14a)} \\ (\rho_2 + \tau_2)(1 + \delta_2(1 + \delta_3)) < 1 & \text{(2.14b)} \\ (\delta_1 + \gamma_1)(1 + \delta_2(1 + \delta_3)) + \tau_3 + \rho_3 < 1 & \text{(2.14c)} \end{cases}$$

and which we use in our empirical study in Section 5.

### *2.2. The linear and extended Kalman filters for state and parameter estimation*

We now recall the linear (LKF) and extended (EKF) Kalman filters corresponding to the models (2.5a)-(2.5b) and (2.7a)-(2.7b), respectively. An in-depth treatment of this topic can be found in Durbin and Koopman (2012).

Let $\mathbf{r} := \{\mathbf{r}_1, \ldots, \mathbf{r}_\mathrm{T}\}$ be a sample containing $T$ three-dimensional observed log-returns and for any $t \leq T$ denote by $\mathcal{F}_t$ the information set generated by the observed returns up to time $t$, that is, $\mathcal{F}_t = \sigma(\mathbf{r}_1, \ldots, \mathbf{r}_\mathrm{t})$. The Kalman recursions yield minimum variance linear unbiased estimates of the forecasted and updated (or filtered) state vectors and of their covariance matrices. We denote by $\boldsymbol{\epsilon}_{t|t} := \mathrm{E}\left[\boldsymbol{e}_t|\mathcal{F}_t\right]$ (respectively, $\boldsymbol{\epsilon}_{t+1} := \mathrm{E}\left[\boldsymbol{e}_{t+1}|\mathcal{F}_t\right]$) the **updated or filtered** (respectively, **forecasted**) state vector and by $P_{t|t} := \mathrm{Var}\left[\boldsymbol{e}_t|\mathcal{F}_t\right]$ (respectively, $P_{t+1} := \mathrm{Var}\left[\boldsymbol{e}_{t+1}|\mathcal{F}_t\right]$) the corresponding covariance matrices. Additionally, let $H_t := \mathrm{Var}[\mathbf{r}_t \mid \mathcal{F}_{t-1}]$ be the forecasted conditional covariance matrices of the returns. The elements that we just introduced can be recursively obtained out of the Kalman recursions (see Durbin and Koopman (2012)) once $\boldsymbol{\epsilon}_1$ and $P_1$ have been provided. More specifically:

$$\boldsymbol{\upsilon}_t = \mathbf{r}_t - (\boldsymbol{\alpha} + B\boldsymbol{\epsilon}_t), \qquad H_t = BP_tB^\top + \Sigma_\mathbf{u}, \tag{2.15}$$

$$\boldsymbol{\epsilon}_{t|t} = \boldsymbol{\epsilon}_t + K_t\boldsymbol{\upsilon}_t, \qquad P_{t|t} = P_t - K_tBP_t, \quad \text{with } K_t := P_tB^\top H_t^{-1}, \tag{2.16}$$

$$\boldsymbol{\epsilon}_{t+1} = T\boldsymbol{\epsilon}_{t|t}, \qquad P_{t+1} = TP_{t|t}T^\top + Q_t, \text{ with } Q_t := R_tR_t^\top. \tag{2.17}$$

Notice that in the linear case $R_t := R$ in (2.6), while in the nonlinear setup $R_t := R_t(\underline{\boldsymbol{e}_t})$ with $R_t(\underline{\boldsymbol{e}_t})$ defined as in (2.7b).

If the parameters $\boldsymbol{\theta} \in \mathbb{R}^s$ of the state-space model are known, the Kalman recursions make possible the filtering of the state vectors for a given observed sample $\mathbf{r} := \{\mathbf{r}_1, \ldots, \mathbf{r}_\mathrm{T}\}$. Otherwise, the vector $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^s$ of the model parameter estimates can be obtained via quasi-maximum likelihood method using a log-likelihood function

$$\log L(\mathbf{r};\boldsymbol{\theta}) = -\frac{n\mathrm{T}}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{\mathrm{T}}\left[\log(\det(H_t)) + \boldsymbol{\upsilon}_t^\top H_t^{-1}\boldsymbol{\upsilon}_t\right], \tag{2.18}$$

where the innovations $\{\boldsymbol{\upsilon}_t\}_{t\in\{1,\ldots,T\}}$ and the covariance matrices $\{H_t\}_{t\in\{1,\ldots,T\}}$ are computed using the Kalman recursions (2.15)-(2.17). The maximization of (2.18) is subjected to various constraints that depend on the particular specification of the state-space model and that we spell out in the following two paragraphs.

**The linear state-space model.** In the linear case (2.5a)-(2.5b), the only constraint that needs to be imposed at the time of estimation is associated to the proper identification of the model. In the following proposition, whose proof is provided in an Supplementary Material section, we provide the sufficient conditions that ensure that the parameters of this model are well identified.

**Proposition 2.1.** *The linear state-space model* (2.5a)-(2.5b) *is well identified if one of the elements of* *the vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ that define the matrix $B$ in (2.4) is constrained to be equal to a constant or, alternatively, when one of the unconditional variances $\sigma^2_{v,1}, \sigma^2_{v,2}, \sigma^2_{v,3}$ that define $R$ in (2.5b) is constrained to be equal to a positive constant.*

**The nonlinear state-space model.** In the nonlinear case, apart from the identification constraints that we specify below in Proposition 2.2, the volatility specifications $\sigma_{i,t}(\underline{\boldsymbol{e}_{t-1}})$, $i \in \{1, 2, 3\}$ need to yield positive values and it is required that the resulting process has stationary solutions. This obviously depends on the specific parametric dependence chosen to define the functions $\sigma_{i,t}(\underline{\boldsymbol{e}_{t-1}})$. Regarding the proper identification of the nonlinear state-space model, we state sufficient conditions in the next proposition.

**Proposition 2.2.** *The nonlinear state-space model (2.7a)-(2.7b) is well identified if one of the elements of the vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^\top$ that define the matrix $B$ in (2.4) is constrained to be equal to a constant or, alternatively, when one of the components of the vector $\boldsymbol{a} := (a_1, a_2, a_3)^\top$ that determine Model 1 or Model 2 is set equal to a positive constant before estimation.*

From the numerical point of view, parameter estimates are obtained via the constrained optimization of the likelihood (2.18). Most standard algorithms (interior point algorithm has been used in the empirical section) are stable in this setup and produce reliable results with limited computational effort when using sample sizes of a few thousand time steps. With standard codes and hardware, the linear model is estimated in a few seconds and its nonlinear counterpart in a few minutes.

## 3. Capturing time-varying correlations

The main objective of this paper is constructing one-step ahead forecasts for the conditional covariance matrices of the observed asset returns. The state-space models that we introduced in the previous section exhibit important limitations in that respect which motivates the need for a second modeling step before proceeding with the forecasting. In this section we explain in detail those limitations and propose a solution.

**Limitations of the state-space models for correlation forecasting**. The volatility forecast associated to the state-space models that we introduced in the previous section are naturally given by expression (2.15). More specifically, the conditional covariances $H_t := \mathrm{Var}[\mathbf{r}_t \mid \mathcal{F}_{t-1}]$ are given by $H_t = BP_tB^\top + \Sigma_{\mathbf{u}}$ and, moreover, expression (2.17) provides the conditional covariance $P_t$ for different components of the GST. Indeed, consider first $\widehat{\boldsymbol{\alpha}}$ and $\widehat{B}$ the parameters of the linear (2.5a)-(2.5b) or nonlinear (2.7a)-(2.7b) state-space model estimated via the maximization of the associated log-likelihood function (2.18). Second, the linear Kalman filter or the EKF, respectively, provide estimates $\{\widehat{\varepsilon}_t\}$ (with $\widehat{\varepsilon}_t := \boldsymbol{\epsilon}_{t|t}$) of the GST $\{\boldsymbol{\varepsilon}_t\}$ and $\{\widehat{\mathbf{u}}_t\}$ of the innovations $\{\mathbf{u}_t\}$ of the observation equations, that is, $\{\widehat{\mathbf{u}}_t\} := \left\{\mathbf{r}_t - (\widehat{\boldsymbol{\alpha}} + \widehat{B}\widehat{\varepsilon}_t)\right\}$. All these estimates can be subsequently used in order to obtain the forecasted conditional covariance matrices $H_t$ of the asset returns.

However, empirically we observe some properties of $\{\widehat{\varepsilon}_t\}$ and $\{\widehat{\mathbf{u}}_t\}$ which are not captured by the relations for the conditional covariances $P_t$ and $H_t$. We name here a few:

**(i)** The process $\{\widehat{\mathbf{u}}_t\}$ exhibits heteroscedasticity while both the linear and the nonlinear models capture by construction only the static unconditional covariance $\Sigma_{\widehat{\mathbf{u}}}$.

**(ii)** The process $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ also proves to be empirically heteroscedastic. These facts are captured by none of the proposed approaches at the first stage. More specifically:

(a) **Linear model** (2.5a)-(2.5b) is intrinsically homoscedastic, that is, the conditional covariances $H_t$ and $P_t$ of both the returns $\{\mathbf{r}_t\}$ and the associated GST are asymptotically constant in time. Indeed, the matrices $R$ and $\Sigma_{\mathbf{u}}$ lead the model to a steady state solution in which $P_t$ converges to the constant matrix $\overline{P}$ determined by (see (Durbin and Koopman, 2012, page 86)):

$$\overline{P} = T\overline{P}T^\top - T\overline{P}B^\top\overline{H}^{-1}B\overline{P}T^\top + RR^\top, \quad \text{and} \quad \overline{H} = B\overline{P}B^\top + \Sigma_{\mathbf{u}}.$$

(b) **Nonlinear model** (2.7a)-(2.7b) both in the case of Model 1 (2.8)-(2.10) and Model 2 (2.11)-(2.13), the matrix $Q_t$ in (2.17) has a non-trivial dynamical behavior and hence so does $P_t$. Nevertheless, a straightforward computation shows that $Q_t$ has zero off-diagonal entries which corresponds to zero correlation between the different components of the GST.

**Capturing time-varying correlations: the second stage**. These observations entail that the description of the GST associated to the models (2.5a)-(2.5b) or (2.7a)-(2.7b) is not complete enough to be used for multivariate volatility forecasting. We solve this limitation by using appropriate multivariate heteroscedastic models fitted to both the filtered estimates $\{\widehat{\varepsilon}_t\}$ of the GST $\{\boldsymbol{\varepsilon}_t\}$ and to the associated estimates $\{\widehat{\mathbf{u}}_t\}$ of the residuals $\{\mathbf{u}_t\}$ of the corresponding observation equations (2.5a) or (2.7a), respectively. There is an extensive variety of multivariate volatility models available in the literature which allow the modeling and forecasting of conditional covariances. Since we are concerned here about capturing the time-varying correlations in $\{\widehat{\varepsilon}_t\}$ and in $\{\widehat{\mathbf{u}}_t\}$, the Dynamic Conditional Correlation (DCC) model (see Engle (2002)) specification suggests itself. The DCC model was originally proposed to prescribe the dynamics of the conditional correlation matrix of returns that are standardized (or "deGARCHed") with the conditional variances of univariate GARCH models.

More specifically, let $\boldsymbol{\zeta}_t \in \mathbb{R}^n$ denote the returns vector standardized with some corresponding conditional standard deviations $\boldsymbol{\sigma}_t$. The DCC model prescribes the dynamics of the conditional covariance matrix $\Sigma_t$ through the correlation matrix $R_t$ of the standardized returns $\boldsymbol{\zeta}_t$ as:

$$\Sigma_t = D_t R_t D_t, \text{ with } D_t = \text{diag}(\sigma_{1,t}, \ldots, \sigma_{n,t}), \tag{3.1}$$

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \ Q_t^* := \mathbb{I}_3 \odot Q_t, \tag{3.2}$$

$$Q_t = (\mathbf{i}_3 \mathbf{i}_3^\top - A - B) \odot Q + A \odot (\boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}^\top) + B \odot Q_{t-1}, \tag{3.3}$$

where $\odot$ denotes the Hadamard (or component wise) matrix product, the parameter matrices $A$ and $B$ are symmetric of size three-by-three, $Q$ is a positive semidefinite parameter matrix of size three-by-three, and $\mathbf{i}_3$ is a column vector whose three components are all equal to one. Equation (3.3) is the most general DCC prescription proposed by Engle (2002); we call it the **Hadamard DCC model**. A simplified and much more parsimonious version of this model is the **scalar subfamily** in which all the elements of $A$ are considered identical and likewise those of $B$; in that case the expression (3.3) is replaced by

$$Q_t = (1 - a - b)Q + a(\boldsymbol{\zeta}_{t-1} \boldsymbol{\zeta}_{t-1}^\top) + bQ_{t-1}, \tag{3.4}$$

with $a, b \in \mathbb{R}^+$ such that $a + b < 1$. The matrix $Q$ is obtained following an approximate targeting procedure that consists in assuming that $Q = \mathrm{E}\left[\boldsymbol{\zeta}_t \boldsymbol{\zeta}_t^\top\right]$ and can thus be estimated by $\hat{Q} := \sum_{t=1}^{T} \boldsymbol{\zeta}_t \boldsymbol{\zeta}_t^\top / T$ prior to estimating the model parameters. Despite the fact that $Q$ is not equal to the second moment matrix of $\{\boldsymbol{\zeta}_t\}$ and, as a consequence, $\hat{Q}$ is not a consistent estimator of $Q$ (see Aielli (2013)), this targeting procedure is used in almost all applications of the DCC model which, according to simulation results in Aielli (2013), does not lead to strong biases in practice.

In our setup we will be fitting the DCC model to the filtered GST components $\{\widehat{\varepsilon}_t\}$ and the residuals $\{\widehat{\mathbf{u}}_t\}$ which need to be standardized accordingly. The asynchronicity of data and the particular specification of the state-space models used at the first stage require this "deGARCH"-ing procedure to be tackled in a non-standard way which we now discuss in detail separately for the case of the linear and the nonlinear state-space models.

### *3.1. Capturing dynamic correlations: the linear state-space GST model*

We start by noticing again that the linear state-space model is intrinsically homoscedastic which means that it does not produce any conditional variances that can be used in order to standardize either the filtered estimates $\{\widehat{\varepsilon}_t\}$ of the GST or the associated residuals $\{\widehat{\mathbf{u}}_t\}$. This calls for the need to model the conditional variances of the GST and the residuals keeping in mind that the parametric prescription to be proposed needs to respect the chronology with which for each given day $t$ the returns $r_{1,t}, r_{2,t}, r_{3,t}$ of the different markets become disclosed, or, equivalently, the chronology with which the filtered components $\widehat{\varepsilon}_{1,t}, \widehat{\varepsilon}_{2,t}, \widehat{\varepsilon}_{3,t}$ of the GST $\widehat{\boldsymbol{\varepsilon}}_t$ and the associated elements $\widehat{u}_{1,t}, \widehat{u}_{2,t}, \widehat{u}_{3,t}$ of the residuals vector $\widehat{\mathbf{u}}_t$ become available (see Figure 1). For the sake of clarity in what follows we work only with $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ but recall that the same technique needs to be applied to the residuals $\{\widehat{\mathbf{u}}_t\}$.

More specifically, we propose two parameter families of conditional variance dynamics that generalize Model 1 in (2.8)-(2.10) and Model 2 in (2.11)-(2.13) that we used in the context of the nonlinear state-space setup. However, it is important to emphasize that the EKF requires the nonlinear state-space model to be of Markov type at a daily level, while this time we can use intraday dependences that will allow us to update the information available more frequently. Consider the following specifications:

- **Model 1 for the conditional variances.**

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \widehat{\varepsilon}_{3,t-1}^2, \tag{3.5}$$

$$\sigma_{2,t}^2 = a_2 + \delta_2 \sigma_{1,t}^2 + \gamma_2 \widehat{\varepsilon}_{1,t}^2, \tag{3.6}$$

$$\sigma_{3,t}^2 = a_3 + \delta_3 \sigma_{2,t}^2 + \gamma_3 \widehat{\varepsilon}_{2,t}^2. \tag{3.7}$$

- **Model 2 for the conditional variances.**

$$\sigma_{1,t}^2 = a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \widehat{\varepsilon}_{3,t-1}^2 + \rho_1 \sigma_{1,t-1}^2 + \tau_1 \widehat{\varepsilon}_{1,t-1}^2, \tag{3.8}$$

$$\sigma_{2,t}^2 = a_2 + \delta_2 \sigma_{1,t}^2 + \gamma_2 \widehat{\varepsilon}_{1,t}^2 + \rho_2 \sigma_{2,t-1}^2 + \tau_2 \widehat{\varepsilon}_{2,t-1}^2, \tag{3.9}$$

$$\sigma_{3,t}^2 = a_3 + \delta_3 \sigma_{2,t}^2 + \gamma_3 \widehat{\varepsilon}_{2,t}^2 + \rho_3 \sigma_{3,t-1}^2 + \tau_3 \widehat{\varepsilon}_{3,t-1}^2. \tag{3.10}$$

**Positivity and stationarity of the conditional variance models.** We now study the conditions that need to be imposed in order to ensure that the models 1 and 2 exhibit second order stationary solutions and produce positive conditional variances. A way to approach this question consists of thinking of the three dimensional time series $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ as the one-dimensional process $\{\widehat{\varepsilon}_{t_i}\}$ obtained by ordering the components of each element $\widehat{\boldsymbol{\varepsilon}}_t$ according to the intraday time at which they have been disclosed. Using this point of view, the models 1 and 2 become one-dimensional GARCH models with time varying (periodic in this case) coefficients that are usually designated with the acronym tvGARCH (see Dahlhaus and Rao (2006); Cizek and Spokoiny (2009); Rohan and Ramanathan (2013), and references therein). More specifically, they can be considered as tvGARCH(1,1) and tvGARCH(3,3) models, respectively, if we rewrite them as:

$$\sigma_{t_i}^2 = a_{t_i} + \delta_{t_i}\sigma_{t_i-1}^2 + \gamma_{t_i}\widehat{\varepsilon}_{t_i-1}^2, \quad \text{and} \quad \sigma_{t_i}^2 = a_{t_i} + \delta_{t_i}\sigma_{t_i-1}^2 + \gamma_{t_i}\widehat{\varepsilon}_{t_i-1}^2 + \rho_{t_i}\sigma_{t_i-3}^2 + \tau_{t_i}\widehat{\varepsilon}_{t_i-3}^2, \tag{3.11}$$

with $i \in \{1, 2, 3\}$, $a_{t_i} := a_i$, $\delta_{t_i} := \delta_i$, $\gamma_{t_i} := \gamma_i$, $\rho_{t_i} := \rho_i$, $\tau_{t_i} := \tau_i$, and where $t_i - 1$ and $t_i - 3$ are defined by using recursively the convention

$$t_i - 1 := \begin{cases} (t-1)_3 & \text{when} \quad i = 1, \\ t_{i-1} & \text{when} \quad i \in \{2, 3\}. \end{cases}$$

The positivity of the conditional variances implied by these models can be obtained by using only positive coefficients in the expressions that define them. Regarding stationarity, a sufficient condition of widespread use in the tvGARCH context (see for example Rohan and Ramanathan (2013)) is that $\delta_{t_i} + \gamma_{t_i} < 1$ for the model 1, and that $\delta_{t_i} + \gamma_{t_i} + \rho_{t_i} + \tau_{t_i} < 1$ for the model 2, with $i \in \{1, 2, 3\}$. Numerical experiments show that, in our particular situation, these conditions lack sharpness and produce mediocre estimation results. In the following proposition, whose proof is provided in the Supplementary Material section, we establish less restrictive stationarity solutions that take advantage of the periodicity of the GARCH coefficients. In order to formulate them, we need to introduce the matrices $A_{t_i}$ associated to the Markov representations of the recursions in (3.11) corresponding to the second model, as well as their expectations $A_i := \mathrm{E}[A_{t_i}]$ (see Section 2.2.2 in Francq and Zakoian (2010) for the details). Let $\{\mathbf{v}_t\} \sim \mathrm{WN}(\mathbf{0}_3, \mathbb{I}_3)$ be the innovations introduced in the definition of the state-space model (2.5b). Then:

$$A_{t_i} := \begin{pmatrix} \gamma_{t_i} v_{t_i}^2 & 0 & \tau_{t_i} v_{t_i}^2 & \delta_{t_i} v_{t_i}^2 & 0 & \rho_{t_i} v_{t_i}^2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \gamma_{t_i} & 0 & \tau_{t_i} & \delta_{t_i} & 0 & \rho_{t_i} \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad A_i := \mathrm{E}[A_{t_i}] = \begin{pmatrix} \gamma_i & 0 & \tau_i & \delta_i & 0 & \rho_i \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \gamma_i & 0 & \tau_i & \delta_i & 0 & \rho_i \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

**Proposition 3.1.** *Consider the GARCH models with time-varying coefficients defined by the recursions in the expression* (3.11)*. If the innovations* $\{\mathbf{v}_t\}$ *that drive them are independent, then the following conditions imply the existence of a unique periodic (with period equal to three) stationary solution:*

**(i)** *For Model 1:* $(\delta_1 + \gamma_1)(\delta_2 + \gamma_2)(\delta_3 + \gamma_3) < 1$.

**(ii)** *For Model 2:* $\rho(A_3A_2A_1) < 1$*, where* $\rho(\cdot)$ *denotes the spectral radius.*

The stationarity condition $\rho(A_3A_2A_1) < 1$ for Model 2 cannot be implemented as such at the time of estimation due to the convoluted analytic expression of the spectrum of $A_3A_2A_1$. Indeed, a straightforward computation shows that:

$$A_3A_2A_1 = \begin{pmatrix} \gamma_1\zeta_2\zeta_3 + \tau_3 & \tau_2\zeta_3 & \tau_1\zeta_2\zeta_3 & \delta_1\zeta_2\zeta_3 + \rho_3 & \rho_2\zeta_3 & \rho_1\zeta_2\zeta_3 \\ \gamma_1\zeta_2 & \tau_2 & \tau_1\zeta_2 & \delta_1\zeta_2 & \rho_2 & \rho_1\zeta_2 \\ \gamma_1 & 0 & \tau_1 & \delta_1 & 0 & \rho_1 \\ \gamma_1\zeta_2\zeta_3 + \tau_3 & \tau_2\zeta_3 & \tau_1\zeta_2\zeta_3 & \delta_1\zeta_2\zeta_3 + \rho_3 & \rho_2\zeta_3 & \rho_1\zeta_2\zeta_3 \\ \gamma_1\zeta_2 & \tau_2 & \tau_1\zeta_2 & \delta_1\zeta_2 & \rho_2 & \rho_1\zeta_2 \\ \gamma_1 & 0 & \tau_1 & \delta_1 & 0 & \rho_1 \end{pmatrix},$$

with $\zeta_i := \delta_i + \gamma_i,\ i = \{2, 3\}$. We now use the fact that for any matrix norm $||\cdot||$ the inequality $\rho(A_3A_2A_1) \leq ||A_3A_2A_1||$ is satisfied and hence it suffices to require that $||A_3A_2A_1|| < 1$ to ensure that $\rho(A_3A_2A_1) < 1$. We implement this condition by using, for example, the maximum row sum norm (see Horn and Johnson (2013)), in which case the inequality $||A|| < 1$ amounts to the following three conditions:

$$\begin{cases} \delta_1 + \gamma_1 + \rho_1 + \tau_1 < 1, & \text{(3.12a)} \\ (\delta_2 + \gamma_2)(\delta_1 + \gamma_1 + \rho_1 + \tau_1) + \rho_2 + \tau_2 < 1, & \text{(3.12b)} \\ (\delta_3 + \gamma_3)((\delta_2 + \gamma_2)(\delta_1 + \gamma_1 + \rho_1 + \tau_1) + \rho_2 + \tau_2) + \tau_3 + \rho_3 < 1. & \text{(3.12c)} \end{cases}$$

**Constructing the DCC model**. Models 1 and 2 subjected to the appropriate parameter constraints deliver conditional variances for the filtered components of the GST $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ that can be used to construct the standardized (or "deGARCH"-ed) corresponding vectors. More specifically, we denote the standardized components of the GST as $\boldsymbol{\zeta}_t \in \mathbb{R}^3$ and obtain them via the component-wise assignment $\zeta_{i,t} := \widehat{\varepsilon}_{i,t}/\sigma^{\widehat{\boldsymbol{\varepsilon}}}_{i,t},\ i \in \{1, 2, 3\}$ with $\sigma^{\widehat{\boldsymbol{\varepsilon}}}_{i,t}$ the conditional deviations provided by Model 1 in (3.5)-(3.7) or Model 2 in (3.8)-(3.10) for the filtered components of the GST $\{\widehat{\boldsymbol{\varepsilon}}_t\}$. Additionally, we set $D^{\widehat{\boldsymbol{\varepsilon}}}_t := \mathrm{diag}(\sigma^{\widehat{\boldsymbol{\varepsilon}}}_{1,t}, \sigma^{\widehat{\boldsymbol{\varepsilon}}}_{2,t}, \sigma^{\widehat{\boldsymbol{\varepsilon}}}_{3,t})$ in both the Hadamard DCC and the scalar DCC models in (3.1)-(3.3) and (3.4), respectively. This approach provides via (3.1) time-varying conditional covariance matrices $\{\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_t\}$. The same strategy needs to be put to work in the case of the residuals $\{\widehat{\mathbf{u}}_t\}$ that yields conditional covariance matrices $\{\Sigma^{\widehat{\mathbf{u}}}_t\}$ in an analogous way.

*3.2. Capturing dynamic correlations: the nonlinear state-space GST model*

As we mentioned earlier, the nonlinear state-space model (2.7a)-(2.7b) is of Markov type at a daily level and yields conditional variances for the date $t$ that are measurable with respect to $\mathcal{F}_{t-1}$. Indeed, the relation (2.7b) implies that

$$\mathrm{Var}\left[\widehat{\varepsilon}_{i,t} \mid \mathcal{F}_{t-1}\right] = \sigma_{i,t}^{\widehat{\boldsymbol{\varepsilon}}\,2}\left(\underline{\widehat{\boldsymbol{e}}_{t-1}}\right), \quad \mathrm{Cov}\left[\widehat{\varepsilon}_{i,t}, \widehat{\varepsilon}_{j,t} \mid \mathcal{F}_{t-1}\right] = 0, \quad \text{for any} \quad i, j \in \{1,2,3\}, i \neq j, \tag{3.13}$$

where the functional prescription $\sigma_{i,t}^{\widehat{\boldsymbol{\varepsilon}}}\left(\underline{\widehat{\boldsymbol{e}}_{t-1}}\right)$ is given by one of the models (2.8)-(2.10) or (2.11)-(2.13) under consideration. The second identity in (3.13) shows that this model neglects the conditional correlation between the components of the GST that is nevertheless empirically observed Durdyev and Peresetsky (2014); Peresetsky and Yakubov (2015) and that can be captured by the DCC models in (3.1)-(3.3) or in (3.4). This strategy introduces time-varying correlation between the components of the GST while preserving the conditional variance (3.13) captured by the non-linear state-space model.

In this instance we construct the standardized vector $\boldsymbol{\zeta}_t \in \mathbb{R}^3$ via the component-wise assignment $\zeta_{i,t} := \widehat{\varepsilon}_{i,t}/\sigma_{i,t}^{\widehat{\boldsymbol{\varepsilon}}}\left(\underline{\widehat{\boldsymbol{e}}_{t-1}}\right)$, $i \in \{1,2,3\}$, and we subsequently estimate the corresponding DCC models. As for the residuals $\{\widehat{\mathbf{u}}_t\}$, we need to standardize them with the standard conditional deviations provided by a GARCH(1,1) model and then fit a relevant Hadamard or scalar DCC model as in (3.1)-(3.3) or (3.4), respectively.

## 4. GST-based volatility forecasting

In the following paragraphs we provide the implementation details of the intraday forecasting scheme for both the linear and the nonlinear state-space models. The empirical performances of the proposed forecasting strategies are evaluated later on in Section 5.

**The extended filtration**. We start by constructing for each trading date $t$, three different filtrations whose elements $\mathcal{F}^*_{t_i}$ are the pseudo-information sets generated by the observed returns and the filtered values $\widehat{\boldsymbol{\varepsilon}}_t$ of the GST up to time $t_i$, $i \in \{1,2,3\}$. We call these the **intraday extended information sets** and define them as

$$\mathcal{F}^*_{t_1} := \sigma\left(\{\mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\} \cup \{\widehat{\boldsymbol{\varepsilon}}_1, \ldots, \widehat{\boldsymbol{\varepsilon}}_{t-1}, \widehat{\varepsilon}_{t_1}\}\right), \tag{4.1}$$

$$\mathcal{F}^*_{t_2} := \sigma\left(\{\mathbf{r}_1, \ldots, \mathbf{r}_{t-1}\} \cup \{\widehat{\boldsymbol{\varepsilon}}_1, \ldots, \widehat{\boldsymbol{\varepsilon}}_{t-1}, \widehat{\varepsilon}_{t_1}, \widehat{\varepsilon}_{t_2}\}\right) = \mathcal{F}^*_{t_1} \cup \sigma\left(\widehat{\varepsilon}_{t_2}\right), \tag{4.2}$$

$$\mathcal{F}^*_{t_3} := \sigma\left(\{\mathbf{r}_1, \ldots, \mathbf{r}_t\} \cup \{\widehat{\boldsymbol{\varepsilon}}_1, \ldots, \widehat{\boldsymbol{\varepsilon}}_t\}\right) = \mathcal{F}^*_t. \tag{4.3}$$

**GST-based intraday forecasting using the extended filtration**. We now construct for each date $t$ three different intraday forecasts for the conditional covariance matrix $H_t$ of the returns, based on the information available at $(t-1)_3 = t_1 - 1$, $t_1 = t_2 - 1$, and $t_2 = t_3 - 1$. We denote these intraday forecasts by $H_{t|t_i-1}$, $i = \{1, 2, 3\}$ and recall that by (2.5a)–(2.5b) or (2.7a)–(2.7b) we have that

$$H_{t|t_i-1} := \mathrm{Var}[(\mathbf{r}_t - \boldsymbol{\alpha}) \mid \mathcal{F}^*_{t_i-1}] = BP^*_{t|t_i-1}B^\top + \Sigma^{\widehat{\mathbf{u}}\,*}_{t|t_i-1}, \quad i = \{1, 2, 3\}\,. \tag{4.4}$$

where

$$P^*_{t|t_i-1} = \mathrm{Var}\left[\widehat{\boldsymbol{e}}_t \mid \mathcal{F}^*_{t_i-1}\right] = \left(\begin{array}{c|c} \Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t|t_i-1} & 0 \\ \hline 0 & \left(\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t-1}\right)_{2:3,2:3} \end{array}\right),$$

with $\Sigma^{\widehat{\mathbf{u}}}_{t|t_i-1}$ and $\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t|t_i-1}$ the conditional covariance matrices of $\{\widehat{\mathbf{u}}_t\}$ and $\{\widehat{\boldsymbol{\varepsilon}}_t\}$, respectively, computed at time $t$ with respect to the elements of the extended filtration associated to the intraday instants $t_i$, $i = \{1, 2, 3\}$. More specifically, $\Sigma^{\widehat{\mathbf{u}}}_{t|t_i-1} := \mathrm{Var}\left[\widehat{\mathbf{u}}_t \mid \mathcal{F}^*_{t_i-1}\right]$ and $\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t|t_i-1} := \mathrm{Var}\left[\widehat{\boldsymbol{\varepsilon}}_t \mid \mathcal{F}^*_{t_i-1}\right]$. The symbol $\left(\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t-1}\right)_{2:3,2:3}$ denotes the two-by-two block in the lower right corner of the matrix $\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t-1}$.

We now address in detail the construction of the forecasts $\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t|t_i-1}$ and note that an identical exercise needs to be carried out for the conditional covariance matrices $\Sigma^{\widehat{\mathbf{u}}}_{t|t_i-1}$ in order to obtain the forecasts (4.4).

**(i) The forecast $H_{t|t_1-1}$:** we produce a forecast for the covariance matrix $H_t$ using the information available at the intraday instant $t_1 - 1 = (t-1)_3$. Since all the DCC models used in the second stage in Section 3 produce $\mathcal{F}_t$–predictable covariance matrices and in this case $\mathcal{F}^*_{(t-1)_3} = \mathcal{F}_{t-1}$, it is clear that the forecast $\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t|t_1-1}$ that we are interested in is given by

$$\Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t|t_1-1} = \Sigma^{\widehat{\boldsymbol{\varepsilon}}}_{t|t-1} = D^{\widehat{\boldsymbol{\varepsilon}}}_{t|t-1} R^{\widehat{\boldsymbol{\varepsilon}}}_{t|t-1} D^{\widehat{\boldsymbol{\varepsilon}}}_{t|t-1}. \tag{4.5}$$

The entries of the matrix $R^{\widehat{\boldsymbol{\varepsilon}}}_{t|t-1}$ in this expression are determined by the equations (3.2)–(3.3) and those of $D^{\widehat{\boldsymbol{\varepsilon}}}_{t|t-1}$ depend on the specific GARCH prescription used to model the conditional variances of the GST. In the nonlinear case, $D^{\widehat{\boldsymbol{\varepsilon}}}_{t|t-1}$ is determined by (2.8)-(2.10) for Model 1 and (2.11)-(2.13) for Model 2. In the case of the linear state-space model, those will be (3.5)-(3.7) for the models of the type 1 and (3.8)-(3.10) for the type 2. More specifically, for Model 1 in

(3.5)-(3.7) we have

$$\sigma_{1,t}^{\widehat{\varepsilon}\,2} = a_1 + \delta_1 \sigma_{3,t-1}^{\widehat{\varepsilon}\,2} + \gamma_1 \widehat{\varepsilon}_{3,t-1}^2, \tag{4.6}$$

$$\sigma_{2,t}^{\widehat{\varepsilon}\,2} = a_2 + (\delta_2 + \gamma_2) \sigma_{1,t}^{\widehat{\varepsilon}\,2}, \tag{4.7}$$

$$\sigma_{3,t}^{\widehat{\varepsilon}\,2} = a_3 + (\delta_3 + \gamma_3) \sigma_{2,t}^{\widehat{\varepsilon}\,2}, \tag{4.8}$$

and for Model 2 in (3.8)-(3.10).

$$\sigma_{1,t}^{\widehat{\varepsilon}\,2} = a_1 + \delta_1 \sigma_{3,t-1}^{\widehat{\varepsilon}\,2} + \gamma_1 \widehat{\varepsilon}_{3,t-1}^2 + \rho_1 \sigma_{1,t-1}^{\widehat{\varepsilon}\,2} + \tau_1 \widehat{\varepsilon}_{1,t-1}^2, \tag{4.9}$$

$$\sigma_{2,t}^{\widehat{\varepsilon}\,2} = a_2 + (\delta_2 + \gamma_2) \sigma_{1,t}^{\widehat{\varepsilon}\,2} + \rho_2 \sigma_{2,t-1}^{\widehat{\varepsilon}\,2} + \tau_2 \widehat{\varepsilon}_{2,t-1}^2, \tag{4.10}$$

$$\sigma_{3,t}^{\widehat{\varepsilon}\,2} = a_3 + (\delta_3 + \gamma_3) \sigma_{2,t}^{\widehat{\varepsilon}\,2} + \rho_3 \sigma_{3,t-1}^{\widehat{\varepsilon}\,2} + \tau_3 \widehat{\varepsilon}_{3,t-1}^2. \tag{4.11}$$

**(ii) The forecast $H_{t|t_1}$:** we produce a forecast for the covariance matrix $H_t$ using the information available at the intraday instant $t_1$. In this case we assume that the index return $r_{1,t}$ quoted at the instant $t_1$ of day $t$ has already been observed and that the corresponding GST return $\widehat{\varepsilon}_{1,t}$ has been filtered and is available. Now, the forecast $H_{t|t_1}$ depends again on the kind of state-model used for the GST modeling. First, in nonlinear case, the model is predictable and is not able to take advantage of intraday information; indeed, the forecast $H_{t|t_1}$ coincides with the one obtained in the previous point and spelled out in (4.5). In the linear state-space model case we use the approximations

$$\Sigma_{t|t_1}^{\widehat{\boldsymbol{\varepsilon}}} := D_{t|t_1}^{\widehat{\boldsymbol{\varepsilon}}} R_{t|t-1}^{\widehat{\boldsymbol{\varepsilon}}} D_{t|t_1}^{\widehat{\boldsymbol{\varepsilon}}} \text{ with } D_{t|t_1}^{\widehat{\boldsymbol{\varepsilon}}} := \operatorname{diag}(\sigma_{1,t}^{\widehat{\varepsilon}}, \sigma_{2,t}^{\widehat{\varepsilon}}, \sigma_{3,t}^{\widehat{\varepsilon}}), \tag{4.12}$$

and the models of type 1 (3.5)-(3.7) or of type 2 (3.8)-(3.10) for the conditional variance of $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ in order to produce the forecasts in the diagonal entries of the matrix $D_{t|t_1}^{\widehat{\boldsymbol{\varepsilon}}}$. More specifically, for Model 1 in (3.5)-(3.7) we have

$$\sigma_{1,t}^{\widehat{\varepsilon}\,2} = a_1 + \delta_1 \sigma_{3,t-1}^{\widehat{\varepsilon}\,2} + \gamma_1 \widehat{\varepsilon}_{3,t-1}^2, \tag{4.13}$$

$$\sigma_{2,t}^{\widehat{\varepsilon}\,2} = a_2 + \delta_2 \sigma_{1,t}^{\widehat{\varepsilon}\,2} + \gamma_2 \widehat{\varepsilon}_{1,t}^2, \tag{4.14}$$

$$\sigma_{3,t}^{\widehat{\varepsilon}\,2} = a_3 + (\delta_3 + \gamma_3) \sigma_{2,t}^{\widehat{\varepsilon}\,2}, \tag{4.15}$$

while for Model 2 in (3.8)-(3.10):

$$\sigma_{1,t}^{\widehat{\varepsilon}\,2} = a_1 + \delta_1 \sigma_{3,t-1}^{\widehat{\varepsilon}\,2} + \gamma_1 \widehat{\varepsilon}_{3,t-1}^2 + \rho_1 \sigma_{1,t-1}^{\widehat{\varepsilon}\,2} + \tau_1 \widehat{\varepsilon}_{1,t-1}^2, \tag{4.16}$$

$$\sigma_{2,t}^{\widehat{\varepsilon}\,2} = a_2 + \delta_2 \sigma_{1,t}^{\widehat{\varepsilon}\,2} + \gamma_2 \widehat{\varepsilon}_{1,t}^2 + \rho_2 \sigma_{2,t-1}^{\widehat{\varepsilon}\,2} + \tau_2 \widehat{\varepsilon}_{2,t-1}^2, \tag{4.17}$$

$$\sigma_{3,t}^{\widehat{\varepsilon}\,2} = a_3 + (\delta_3 + \gamma_3) \sigma_{2,t}^{\widehat{\varepsilon}\,2} + \rho_3 \sigma_{3,t-1}^{\widehat{\varepsilon}\,2} + \tau_3 \widehat{\varepsilon}_{3,t-1}^2. \tag{4.18}$$

**(iii) The forecast $H_{t|t_2}$:** we produce a forecast for the covariance matrix $H_t$ using the information available at the intraday instant $t_2$. In this case we assume that the index return $r_{2,t}$ quoted at the instant $t_2$ of day $t$ has already been observed and that the corresponding GST return $\widehat{\varepsilon}_{2,t}$ has been filtered and is available. Now, the forecast $H_{t|t_2}$ depends again on the kind of state-model used for the GST modeling. In nonlinear case, we again have that $H_{t|t_2}$ coincides with (4.5). In the linear state-space model case we use the approximations

$$\Sigma_{t|t_2}^{\widehat{\varepsilon}} = D_{t|t_2}^{\widehat{\varepsilon}} R_{t|t-1}^{\widehat{\varepsilon}} D_{t|t_2}^{\widehat{\varepsilon}} \text{ with } D_{t|t_2}^{\widehat{\varepsilon}} := \operatorname{diag}(\sigma_{1,t}^{\widehat{\varepsilon}}, \sigma_{2,t}^{\widehat{\varepsilon}}, \sigma_{3,t}^{\widehat{\varepsilon}}), \tag{4.19}$$

and we use the models of type 1 (3.5)-(3.7) or of type 2 (3.8)-(3.10) for the conditional variance of $\{\widehat{\varepsilon}_t\}$ in order to produce the forecasts in the diagonal entries of the matrix $D_{t|t_1}^{\widehat{\varepsilon}}$.

## 5. Empirical performance of the GST-based volatility forecasting schemes

In this section we carry out an empirical study to assess the one-step ahead volatility forecasting performances of the proposed two-stage modeling approaches. More specifically, first, we construct the linear and nonlinear state-space models for the log-returns of three major market indices with non-synchronous closing times; second, we fit to the filtered state variables and to the model residuals (residuals of the observation equations) multivariate conditional correlation models in order to account for the correlation dynamics that is empirically observed. Finally, we use the estimated models for one-step ahead volatility forecasting and compare their performances with standard benchmarks in the literature that do not take into account the arrival of intraday information.

### *5.1. Dataset, competing models, and volatility forecasting study*

**Dataset.** We use as dataset[5] the daily closing values of three major stock market indices, namely, NIKKEI 225, FTSE 100, and S&P 500[6]. These markets are geographically located in different time zones and have asynchronous closing times: NIKKEI 225 is an index based on the quotes of the Tokyo Stock Exchange that closes at 6:00 UTC. FTSE 100 and S&P 500 are based on the quotes of the London and the New York stock exchanges that close at 16:30 and 21:00 UTC, respectively (see Figure 2 for a diagram representing the chronology of the three markets that are chosen for the empirical study). The closing values are adjusted for dividend payments and stock splits and the resulting data is synchronized by taking into account all the holidays of the different markets. The daily log-returns for the three indices are computed between January 5, 1996 and April 1, 2015 which yields a dataset with $T := 4581$ observations. The whole log-returns sample is demeaned and it is divided into two parts. The first one corresponds to the period between January 5, 1996 and April 1, 2013; it has length $T_{\text{est}} := 4095$ and it is reserved for estimation purposes. The remaining $T_{\text{out}} := 486$ observations from April 2, 2013 to April 1, 2015 are reserved for an out-of-sample study consisting on one-day ahead volatility forecasting.

In order to illustrate the robustness of the results obtained in this empirical study, we have included in the supplementary material in Section Appendix A.6 a similar analysis based on the same dataset but using a shorter estimation period (January 5, 1996 – December 4, 2006) that does not contain the volatility events in the Fall 2008. The out-of-sample study in that case comprises the entire Great Recession (December 5, 2006 – April 1, 2015).

**Competing models.** We consider three groups of models whose one-step ahead volatility forecasting performance will be assessed, namely:

**(i) Scalar and Hadamard DCC daily models**. These families of models are designed and widely used for volatility forecasting and we hence choose them to serve as a benchmark for the forecasting tasks that we perform in this empirical section.

**Estimation:** We proceed in a standard way by constructing and estimating on $T_{\text{est}}$ observations both scalar and Hadamard three dimensional models that use exclusively the daily quoted information on the closing values of the indices under consideration and, as it is customary, ignore their non-synchronicity. The first stage of the model construction is common for both the scalar

[5] In our empirical exercises we used also another dataset that consisted of NIKKEI 225, MICEX, and S&P 500. The conclusions were analogous but we however decided to consider the FTSE 100 index instead of MICEX since it belongs to the same group of the major stock market indices of the developed markets.

[6] Data were downloaded from the Yahoo Finance database

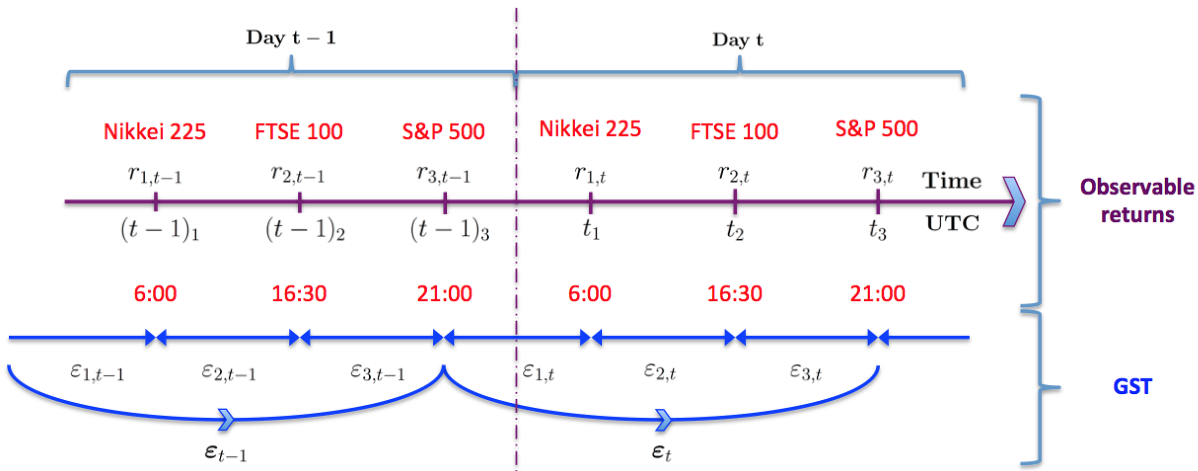


Figure 2: Diagram representing the different variables, time labels, and chronology corresponding to the indices used in the empirical study.

and the Hadamard setups. The deGARCHing of the returns is accomplished with the conditional deviations provided by standard one-dimensional GARCH(1,1) models that are estimated on the individual daily returns. The estimation procedure in the case of the scalar DCC (3.4) is straightforward (see for instance Engle (2009)), while the Hadamard prescription (3.3) presents some complications due to the presence of positivity constraints to which the model parameters are subjected and that we handle using the tools presented in Bauwens et al. (2016).

**Volatility forecasting**: We implement volatility forecasting using the $T_{\text{out}}$ observations reserved for the out-of-sample study. The value of the one-day ahead forecast of the conditional covariance matrix $H_t$ of the returns $(r_{1,t}, r_{2,t}, r_{3,t})$, $t \in \{T_{\text{est}}+1, \ldots, T_{\text{est}}+T_{\text{out}}\}$, with respect to the information set $\mathcal{F}_{t-1}$ is computed by setting $H_t := D_t R_t D_t$, with $R_t$ given by (3.1)-(3.3) and $D_t := \text{diag}(\sigma_{1,t}, \sigma_{2,t}, \sigma_{3,t})$ a diagonal matrix containing the conditional standard deviations obtained out of the GARCH(1,1) model that has been previously fit to the log-returns during the first stage of the DCC model construction. The models do not take into account the asynchronicity of the data arrival, hence the forecasting exercise provides one conditional covariance matrix $H_t$ of the returns $(r_{1,t}, r_{2,t}, r_{3,t})$ for each day $t \in \{T_{\text{est}}+1, \ldots, T_{\text{est}}+T_{\text{out}}\}$.

**(ii) Linear state space model combined with Models 1 and Model 2 (LSS Model 1/LSS Model 2)**.

**Estimation:** We use the linear state-space setup (2.5a)-(2.5b) and we estimate the components $\{\widehat{\varepsilon}_t\}$ of the GST via the Kalman recursions (2.15)-(2.17), together with the model parameters by minimizing minus the log-likelihood function provided in (2.18) associated with the considered $T_{\text{est}}$ in-sample observations. The model has been properly identified using Proposition 2.1 by setting $\beta_1 = 0.4355$[7]. We proceed by fitting scalar and Hadamard DCC models on the filtered estimates $\{\widehat{\varepsilon}_t\}$ of the GST and on the associated linear state-space model residuals $\{\widehat{\mathbf{u}}_t\}$. In order to standardize the estimates $\{\widehat{\varepsilon}_t\}$ that are needed to construct the DCC model, we use either Model 1 in (3.5)-(3.7) or Model 2 in (2.11)-(2.13) whose parameters are estimated under the associated positivity and stationarity constraints (see Section 3.1 for the detailed discussion and, in particular, Proposition 3.1). In order to standardize the residuals $\{\widehat{\mathbf{u}}_t\}$, we follow the analogous approach. Additionally, we add to the competing models the case when these residuals are "deGARCHed" with the conditional deviations coming from the standard individual GARCH(1,1) models. Scalar and Hadamard DCC (3.3) models are used at the time of modeling the conditional variances of both for the GST components $\{\widehat{\varepsilon}_t\}$ and of the residuals $\{\widehat{\mathbf{u}}_t\}$.

**Volatility forecasting**: We construct the volatility forecasts for the $T_{\text{out}}$ observations reserved for the out-of-sample study following the description provided in Section 4. For each day $t \in \{T_{\text{est}}+1, \ldots, T_{\text{est}}+T_{\text{out}}\}$ LSS Model 1 and LSS Model 2 provide three forecasts for the conditional covariance matrix of the returns $(r_{1,t}, r_{2,t}, r_{3,t})$. More specifically at the end of the day before, the third market closes and we construct $H_{t|(t-1)_3} = H_{t|t_1-1}$ (see point **(i)** in Section 4), after the closing of the first market, $H_{t|t_1}$ becomes available (see point **(ii)** in Section 4) and, finally, as soon as the second market is closed, $H_{t|t_2}$ is obtained (see point **(iii)** in Section 4).

**(iii) Nonlinear state space model combined with Models 1 and 2 (NSS Model 1/NSS Model 2)**: A procedure identical to the one presented in the previous point is followed but, in this case, using the nonlinear state-space model (2.7a)-(2.7b) for the GST components $\{\widehat{\varepsilon}_t\}$ in which the Model 1 (2.8)-(2.10) or the Model 2 (2.11)-(2.13) prescribe their conditional variances using a GARCH-type functional dependence adapted to their chronology. The models are properly identified using Proposition 2.2 by setting $\beta_1 = 0.4355$[8] and, additionally, the associated positivity and stationarity constraints are imposed (see the discussion below (2.8)-(2.10) and the relations (2.14a)-(2.14c), respectively). As in the previous case, scalar and Hadamard DCC models are then used for both the filtered $\{\widehat{\varepsilon}_t\}$ and the residuals $\{\widehat{\mathbf{u}}_t\}$. The standardizing of the estimated GST components $\{\widehat{\varepsilon}_t\}$ is performed by using the conditional deviations implied by the Model 1 in (2.8)-(2.10) or by the Model 2 (2.11)-(2.13). In the case of construction of the DCC model for the

[7] This choice of $\beta_1$ is arbitrary and is taken close to the estimate of $\beta_1$ obtained in Korhonen and Peresetsky (2013)
[8] See footnote 7.

residuals $\{\widehat{\mathbf{u}}_t\}$, these are standardized with the conditional deviations determined with standard univariate GARCH(1,1) models.

**Volatility forecasting**: The volatility forecasting is carried out using the $T_{\text{out}}$ observations reserved for the out-of-sample study following the description provided in Section 4. Since in this case we cannot take advantage of the intraday information, the forecasting exercise provides one conditional covariance matrix $H_t$ of the returns $(r_{1,t}, r_{2,t}, r_{3,t})$ for each day $t \in \{T_{\text{est}}+1, \ldots, T_{\text{est}} + T_{\text{out}}\}$.

**In-sample goodness-of-fit.** The table 5.1 contains the values of the log-likelihood function for the corresponding models at the estimated parameters. For the conventional DCC models in **(i)** fitted to the returns these values are obtained in a standard way. For the two other groups of models that we put forward in this work, namely **(ii)** and **(iii)**, the log-likelihood function incorporates the two stages (state-space and DCC) and is evaluated using the expression (2.18) with $T = T_{\text{est}}$ and $H_t$ given by

$$H_t = BP_tB^\top + \Sigma_t^{\widehat{\mathbf{u}}} \tag{5.1}$$

with

$$P_t = \left( \begin{array}{c|c} \Sigma_t^{\widehat{\boldsymbol{\varepsilon}}} & 0 \\ \hline 0 & \left(\Sigma_{t-1}^{\widehat{\boldsymbol{\varepsilon}}}\right)_{2:3,2:3} \end{array} \right), \quad \Sigma_t^{\widehat{\boldsymbol{\varepsilon}}} = D_t^{\widehat{\boldsymbol{\varepsilon}}} R_t^{\widehat{\boldsymbol{\varepsilon}}} D_t^{\widehat{\boldsymbol{\varepsilon}}\top}, \quad \text{and} \quad \Sigma_t^{\widehat{\mathbf{u}}} = D_t^{\widehat{\mathbf{u}}} R_t^{\widehat{\mathbf{u}}} D_t^{\widehat{\mathbf{u}}\top}.$$

Table 5.1 reports for each of the competing models the total number of parameters, the associated values of the Akaike (AIC), consistent Akaike (cAIC), and Bayesian information criteria (BIC). The results reveal that the worst performing models for all the statistics considered are the standard DCC models that ignore the asynchronous intraday arrival of information. The best fit is exhibited by the linear state space model with the second prescription (3.8)–(3.10) for the conditional variances of the GST and the residuals, combined with a scalar DCC model for the corresponding conditional correlations.

| | DCC models | | Linear state-space models | | | | | | | | Nonlinear state-space models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Model 1 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ | | | | Model 2 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ | | | | Model 1 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ | | Model 2 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ | |
| | | | Model 1 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | Model 2 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | |
| DCC models | S | H | S | H | S | H | S | H | S | H | S | H | S | H |
| $k$ | 11 | 21 | 30 | 50 | 30 | 50 | 42 | 62 | 36 | 56 | 25 | 45 | 31 | 51 |
| AIC | -76274.88$^{13}$ | -76255.11$^{14}$ | -86500.10$^{5}$ | -86459.79$^{6}$ | -86412.08$^{7}$ | -86370.17$^{8}$ | **-87160.88**$^{1}$ | -87120.84$^{2}$ | -86764.80$^{3}$ | -86723.17$^{4}$ | -80486.32$^{11}$ | -80477.04$^{12}$ | -82035.41$^{9}$ | -81999.33$^{10}$ |
| BIC | -76205.39$^{13}$ | -76122.44$^{14}$ | -86310.57$^{5}$ | -86143.91$^{7}$ | -86222.55$^{6}$ | -86054.29$^{8}$ | **-86895.54**$^{1}$ | -86729.15$^{2}$ | -86537.37$^{3}$ | -86369.39$^{4}$ | -80328.38$^{11}$ | -80192.75$^{12}$ | -81839.56$^{9}$ | -81677.13$^{10}$ |
| cAIC | -76274.82$^{13}$ | -76254.88$^{14}$ | -86499.64$^{5}$ | -86458.53$^{6}$ | -86411.62$^{7}$ | -86368.91$^{8}$ | **-87159.99**$^{1}$ | -87118.90$^{2}$ | -86764.14$^{3}$ | -86721.59$^{4}$ | -80486.00$^{11}$ | -80476.01$^{12}$ | -82034.92$^{9}$ | -81998.01$^{10}$ |
| $\log L$ | 38148.44$^{14}$ | 38148.55$^{13}$ | 43280.05$^{5}$ | 43279.89$^{6}$ | 43236.04$^{7}$ | 43235.09$^{8}$ | **43622.44**$^{1}$ | 43622.42$^{2}$ | 43418.40$^{3}$ | 43417.59$^{4}$ | 40268.16$^{12}$ | 40283.52$^{11}$ | 41048.70$^{10}$ | 41050.66$^{9}$ |

Table 5.1: Total number of model parameters $k$, values of the log-likelihood function $\log L$, and associated AIC, cAIC, and BIC statistics. The largest values of the log-likelihood function and the smallest values of the information criteria are displayed in red bold. Exponents of the values at each row indicate the rank of the model from 14 (the worse) to 1 (the best).

*5.2. Model confidence sets based on covariance and KLIC loss functions*

The different models are compared using the model confidence set (MCS) approach introduced in Hansen et al. (2003, 2011) with loss functions that involve the daily log-returns of the three indices under consideration and the forecasts of the conditional covariance matrices associated to each of the competing models.

**Covariance loss functions.** We use three different covariance loss functions in the implementation of the MCS approach depending on the specific intraday extended information set used at the time of forecasting, namely:

$$d^{\mathrm{Cov}}_{t|t_i-1} := \frac{1}{6} \sum_{j \leq k = 1,2,3} \left( r_{j,t} r_{k,t} - (H_{t|t_i-1})_{j,k} \right)^2, \ \ i = 1, 2, 3, \tag{5.2}$$

where $t \in \{T_{\mathrm{est}} + 1, \ldots, T_{\mathrm{est}} + T_{\mathrm{out}}\}$ and $(H_{t|t_i-1})_{j,k}$ are the $(j,k)$-entries of the corresponding model dependent forecasts for the conditional covariance matrices at $t$. More specifically, when using the scalar/Hadamard DCC model, we will consider the conditional covariance matrix $H_t$ at $t$ with respect to the information set $\mathcal{F}_{t-1}$.

Since the nonlinear state-space model is predictable and is not able to take advantage of intraday information, we hence consider the covariance matrix $H_{t|t-1}$ associated to $\mathcal{F}^*_{t-1}$ and determined by (4.4) and (4.5). In these two groups of models the values of the loss functions will have the same value regardless the intraday moment of time and independently from the value of $i = 1, 2, 3$. Finally, for the linear state-space model instance, we will use the three forecasts $H_{t|t_i-1}$, $i = 1, 2, 3$, based on the three different intraday extended information sets and provided in points **(i)-(iii)** in Section 4.

The MCS approach identifies, from a set of competing models, the subset of models that are statistically equivalent in terms of out-of-sample conditional covariance predictive ability and which outperform all the other models at a considered significance level $\alpha$ for the so called equivalence test. We set this significance level at 10% and 25%, and use 100 000 block bootstrap replicates with block length two in order to obtain the distribution of the relevant test statistic under the null of equal predictive ability.

Tables 5.2 and 5.3 contain the MCS results associated to the values of the covariance loss functions obtained in 36 different out-of-sample time intervals of the form $\{T_{\mathrm{est}} + 1, \ldots, T_{\mathrm{est}} + 136 + 10k\}$ with $k = \{0, 1, \ldots, 35\}$. The first 136 elements in the out-of-sample period are included in all these intervals in order to ensure that there are enough values available for the bootstrapping process that is necessary in the estimation of the distribution of the model equivalence test statistic. The date corresponding to the end of this offset interval is October 18, 2013.

These tables report, for each model, the number of times that it is included in the model confidence

| | DCC models | | Linear state-space models | | | | | | | | Nonlinear state-space models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | |
| | | | Model 1 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | Model 2 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | |
| DCC models | S | H | S | H | S | H | S | H | S | H | S | H | S | H |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{(t-1)_3}}$ | **36** | **36** | 27 | 0 | **36** | **36** | **36** | **36** | **36** | **36** | 0 | 0 | **36** | **36** |
| Sum $p$-vals | **29.090** | **36.000** | 4.388 | 1.204 | **28.153** | **28.260** | **25.607** | **17.969** | **28.892** | **28.892** | 0.109 | 0.109 | **20.816** | **22.426** |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{t_1}}$ | 23 | 23 | 2 | 0 | 23 | 23 | **36** | 23 | 7 | 7 | 0 | 0 | 16 | 17 |
| Sum $p$-vals | 11.804 | 14.634 | 1.449 | 0.465 | 7.050 | 7.050 | **32.679** | 7.050 | 2.176 | 2.176 | 0.121 | 0.121 | 6.397 | 6.977 |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{t_2}}$ | 14 | 16 | 2 | 0 | 8 | 8 | **36** | 13 | **36** | **36** | 0 | 0 | 9 | 10 |
| Sum $p$-vals | 12.987 | 15.561 | 1.665 | 0.457 | 2.745 | 2.745 | **19.799** | 5.911 | **31.858** | **35.238** | 0.117 | 0.117 | 5.739 | 6.525 |

Table 5.2: Model confidence sets (MCS) constructed using the covariance based loss functions (5.2) for 36 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\mathrm{est}} + 136 + 10k$, $k \in \{0, 1, \ldots, 35\}$. The letters 'S' and 'H' stand for 'Scalar' and 'Hadamard', respectively. For each model and information set under consideration, the corresponding value indicates the number of times that the model has been included in the MCS at a 90% confidence level; the value underneath indicates the sum of all the MCS $p$-values obtained by a given model in the 36 tests. The best performing models (determined by the number of times included in the MCS) for the considered information set are marked in bold red.

| | DCC models | | Linear state-space models | | | | | | | | Nonlinear state-space models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | |
| | | | Model 1 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | Model 2 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | |
| DCC models | S | H | S | H | S | H | S | H | S | H | S | H | S | H |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{(t-1)_3}}$ | **36** | **36** | 0 | 0 | **36** | **36** | **36** | 30 | **36** | **36** | 0 | 0 | 26 | 30 |
| Sum $p$-vals | **29.023** | **36.000** | 4.399 | 1.202 | **28.136** | **28.136** | **25.618** | 17.965 | **28.828** | **28.828** | 0.107 | 0.107 | 20.808 | 22.366 |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{t_1}}$ | 13 | 13 | 0 | 0 | 7 | 7 | **36** | 7 | 0 | 0 | 0 | 0 | 7 | 7 |
| Sum $p$-vals | 11.821 | 14.641 | 1.449 | 0.467 | 7.072 | 7.072 | **32.688** | 7.072 | 2.179 | 2.179 | 0.120 | 0.120 | 6.420 | 7.005 |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{t_2}}$ | 13 | 15 | 0 | 0 | 0 | 0 | **36** | 7 | **36** | **36** | 0 | 0 | 7 | 7 |
| Sum $p$-vals | 12.989 | 15.566 | 1.650 | 0.453 | 2.728 | 2.728 | **19.816** | 5.909 | **31.856** | **35.238** | 0.116 | 0.116 | 5.722 | 6.508 |

Table 5.3: Model confidence sets (MCS) constructed using the covariance based loss functions (5.2) for 36 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\mathrm{est}} + 136 + 10k$, $k \in \{0, 1, \ldots, 35\}$. The letters 'S' and 'H' stand for 'Scalar' and 'Hadamard', respectively. For each model and information set under consideration, the corresponding value indicates the number of times that model has been included in the MCS at a 75% confidence level; the value underneath indicates the sum of all the MCS $p$-values obtained by a given model in the 36 tests. The best performing models (determined by the number of times included in the MCS) for the considered information set are marked in bold red.

set with the a significance level of 10% or 25%. The second figure represents the total sum of the 36 obtained MCS $p$-values corresponding to each model. These results show that:

**(i)** *The group of Kalman-based linear models significantly outperforms the standard DCC models as soon as the intraday information is taken into account.*

**(ii)** *The linear state-space Model 2 significantly outperforms other competing models as soon as the intraday extended information sets are involved.*

The supplementary material section contains the results of an analogous experiment with a shorter estimation period that does not contain the high volatility events of the Fall 2008 period. In that situation the empirical study shows that: first, the conclusion **(i)** above holds and, second, in **(ii)** it is the linear state-space Model 1 that has the superior performance.

**KLIC loss functions.** We also implement the MCS approach using loss functions based on the Kullback-Leibler Information Criterion (KLIC) Kullback and Leibler (1951). We recall that the KLIC divergence $D_{T_{\rm est},t}(\phi\|\psi)$ of a density $\psi$ that depends on the parameters $\boldsymbol{\theta}$ with respect to another density $\phi$, is defined as:

$$D_{T_{\rm est},t}(\phi\|\psi) = \frac{1}{t-T_{\rm est}} \sum_{i=T_{\rm est}+1}^{t} \ln\left[\frac{\phi_i(\boldsymbol{v}_i)}{\psi_i(\boldsymbol{v}_i;\boldsymbol{\theta})}\right], \quad t \in \{T_{\rm est}+1,\ldots,T_{\rm est}+T_{\rm out}\} \tag{5.3}$$

where $\phi_i(\boldsymbol{v}_i)$ is the real underlying conditional density associated to the data under consideration, $\psi_i(\boldsymbol{v}_i;\boldsymbol{\theta})$ is the one coresponding to the competing model of interest, and $\boldsymbol{v}_i$ stands in our case either for the Kalman residuals, in the case of the state-space based models, or for the returns in the case of the benchmark DCC models.

We use this information criterion in order to construct a loss function to evaluate the out-of-sample density forecasting abilities of the considered models, that we subsequently use in the MCS context (see Bao et al. (2006) and Banulescu et al. (2015)). Since the terms having to do with the real density $\phi(\mathbf{r})$ are common to all the models and appear as an additive constant, we then disregard the numerator in (5.3) at the time of constructing the KLIC loss functions. Additionally, in order to account for the specific intraday extended information sets used at the time of forecasting, we use again three different KLIC loss functions adapted to these different filtrations, namely:

$$d^{\rm KLIC}_{t|t_i-1} := -\ln\left[\frac{1}{(2\pi)^{3/2}\det(H_{t|t_i-1})^{1/2}}\exp\left(-\frac{1}{2}\boldsymbol{v}_i^\top(H_{t|t_i-1})^{-1}\boldsymbol{v}_i\right)\right], \tag{5.4}$$

where $t \in \{T_{\rm est}+1,\ldots,T_{\rm est}+T_{\rm out}\}$ and $h_{ij,t}$ are the $(i,j)$-entries of the model dependent forecasts for the conditional covariance matrices at $t$ explained above in the context of the covariance loss functions (5.2).

Table 5.4 contains the MCS results at significance levels 10% and 25% (they are identical) associated to the values of the covariance loss functions obtained in the 36 different out-of-sample time intervals of the form $\{T_{\text{est}}+1,\ldots,T_{\text{est}}+136+10k\}$ with $k=\{0,1,\ldots,35\}$.

| | DCC models | | Linear state-space models | | | | | | | | Nonlinear state-space models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Model 1 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ | | | | Model 2 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ | | | | Model 1 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ | | Model 2 for $\{\widehat{\boldsymbol{\varepsilon}}_t\}$ | |
| | | | Model 1 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | Model 2 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | |
| DCC models | S | H | S | H | S | H | S | H | S | H | S | H | S | H |
| MCS $d^{\text{KLIC}}_{\mathcal{F}^*_{(t-1)_3}}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **36** | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | **36.000** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| MCS $d^{\text{KLIC}}_{\mathcal{F}^*_{t_1}}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **36** | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | **36.000** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| MCS $d^{\text{KLIC}}_{\mathcal{F}^*_{t_2}}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **36** | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | **36.000** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 5.4: Model confidence sets (MCS) constructed using the KLIC loss functions (5.4), for 36 different out-of-sample lengths $l(k)$, namely for $l(k)=T_{\text{est}}+136+10k$, $k\in\{0,1,\ldots,35\}$. The letters 'S' and 'H' stand for 'Scalar' and 'Hadamard', respectively. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at the 90% and 75% confidence levels (they are identical); the value underneath indicates the sum of all the MCS $p$-values obtained by a given model in the 36 tests. The best performing models for the considered information set are marked in bold red.

The results are robust with respect to the number of the out-of-sample observations and as for the previous MCS experiment we can conclude that:

**(i)** *The group of Kalman-based linear models significantly outperform the standard DCC benchmarks regardless the information sets involved.*

**(ii)** *The linear state-space Model 2 significantly outperforms other competing models regardless the information sets involved.*

The supplementary material section contains the results of an analogous experiment with a shorter estimation period that does not contain the high volatility events of the Fall 2008 period. In that situation the empirical study shows that: first, *the conclusion (i) above holds and, second, the linear state-space Model 1 and Model 2 significantly outperform other competing models regardless the information sets involved.*

## 6. Conclusions

In this work we have used linear and nonlinear state-space models that extract global stochastic financial trends out of asynchronous daily data. These models are specifically constructed to take advantage of the intraday arrival of closing information coming from different international markets located in lagged time zones in order to enhance volatility and correlation forecasting performance.

The state-space models considered incorporate nonlinearities at various levels capable of capturing the heteroscedasticity that global trends empirically exhibit. This feature is of much importance since correlation forecasting is the main application developed. The identification of these models, as well as the constraints that their parameters need to satisfy in order to exhibit stationary solutions and positive semidefinite conditional correlation matrices, are carefully studied.

A volatility forecasting empirical study using the adjusted closing values of three major indices (NIKKEI 225, FTSE 100, and S&P 500) has been conducted using the models introduced in the theoretical part and two different estimation periods. In this experiment, we use the model confidence set (MCS) approach of Hansen et al. (2003, 2011) implemented with loss functions constructed with the conditional covariance matrices implied by the different models under consideration. The results show that *the proposed Kalman-based forecasting scheme exhibits statistically significant performance improvements* when compared to the use of standard multivariate parametric correlation models (scalar and non-scalar DCC).

## 7. Acknowledgments

LG acknowledges partial financial support from the Faculty for the Future Program of the Schlumberger Foundation. LG and JPO acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493) as well as the hospitality of the International Laboratory of Quantitative Finance of the National Research University Higher School of Economics, Moscow, that made possible the collaborative work that lead to the results presented in this paper. The research of AP was partially supported by the Russian Federation government grant 14.A12.31.0007.

## References

Aielli, G.P., 2013. Dynamic Conditional Correlation: on Properties and Estimation. Journal of Business & Economic Statistics 31, 282–299.

Ambroževičit, D., Klivečka, A., 2008. On the tvGARCH(1,1) model: Existence, CLT, and tail index. Lithuanian Mathematical Journal 48, 1–16.

Bae, B.Y., Kim, D.H., 2011. Global and Regional Yield Curve Dynamics and Interactions: The Case of Some Asian Countries. International Economic Journal 25, 717–738.

Banulescu, G.D., Hansen, P.R., Huang, Z., 2015. Volatility during the financial crisis through the lens of high frequency data: a realized GARCH approach.

Bao, Y., Lee, T.H., Saltoglu, B., 2006. Comparing density forecast models.

Bauwens, L., Grigoryeva, L., Ortega, J.P., 2016. Estimation and empirical performance of non-scalar dynamic conditional correlation models. Computational Statistics and Data Analysis 100, 17–36.

Bentes, S.R., 2015. On the integration of financial markets: How strong is the evidence from five international stock markets? Physica A: Statistical Mechanics and its Applications 429, 205–214.

Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital asset pricing model with time varying covariances. Journal of Political Economy 96, 116–131.

Cartea, Á., Karyampas, D., 2011. Volatility and covariation of financial assets: A high-frequency analysis. Journal of Banking & Finance 35, 3319–3334.

Chang, Y., Isaac Miller, J., Park, J.Y., 2009. Extracting a common stochastic trend: Theory with some applications. Journal of Econometrics 150, 231–247.

Choudhry, T., Lu, L., Peng, K., 2007. Common stochastic trends among Far East stock prices: Effects of the Asian financial crisis. International Review of Financial Analysis 16, 242–261.

Chrétien, S., Ortega, J.P., 2014. Multivariate GARCH estimation via a Bregman-proximal trust-region method. Computational Statistics and Data Analysis 76, 210–236.

Chung, P.J., Liu, D.J., 1994. Common stochastic trends in pacific rim stock markets. The Quarterly Review of Economics and Finance 34, 241–259.

Cizek, P., Spokoiny, V., 2009. Varying coefficient GARCH models, in: Mikosch, T., Kreiß, J.P., Davis, R.A., Andersen, T.G. (Eds.), Handbook of Financial Time Series. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 169–185.

Dahlhaus, R., Rao, S.S., 2006. Statistical inference for time-varying ARCH processes. The Annals of Statistics 34, 1075–1114.

Dungey, M., Martin, V.L., Pagan, A.R., 2001. A multivariate latent factor decomposition of international bond yield spreads. Journal of Applied Econometrics 15, 697–715.

Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods. Oxford University Press.

Durdyev, R.I., Peresetsky, A., 2014. Autocorrelation in the global stochastic trend. Applied Econometrics 35, 81–100.

Engle, R., 2009. Anticipating Correlations. Princeton University Press, Princeton, NJ.

Engle, R.F., 2002. Dynamic conditional correlation -a simple class of multivariate GARCH models. Journal of Business and Economic Statistics 20, 339–350.

Felices, G., Wieladek, T., 2012. Are emerging market indicators of vulnerability to financial crises decoupling from global factors? Journal of Banking & Finance 36, 321–331.

Francq, C., Zakoian, J.M., 2010. GARCH Models: Structure, Statistical Inference and Financial Applications. Wiley.

Gouriéroux, C., 1997. ARCH models and financial applications. Springer Series in Statistics, Springer-Verlag, New York.

Hansen, P.R., Lunde, A., Nason, J.M., 2003. Choosing the best volatility models: the model confidence set approach. Oxford Bulletin of Economics and Statistics 65, 839–861.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Horn, R.A., Johnson, C.R., 2013. Matrix Analysis. Cambridge University Press. second edition.

Jeon, B.N., Chiang, T.C., 1991. A system of stock prices in world stock exchanges: Common stochastic trends for 19751990. Journal of Economics and Business 43, 329–338.

Kasa, K., 1992. Common stochastic trends in international stock markets. Journal of Monetary Economics 29, 95–124.

Korhonen, I., Peresetsky, A., 2013. Extracting global stochastic trend from non-synchronous data. Bank of Finland, BOFIT Discussion Papers 15/2013.

Kullback, L., Leibler, R.A., 1951. On information and sufficiency. Annals of Mathematical Statistics 22, 79–86.

Lucey, B.M., Muckley, C., 2011. Robust global stock market interdependencies. International Review of Financial Analysis 20, 215–224.

Peresetsky, A., Yakubov, R., 2015. Autocorrelation in an unobservable global trend: does it help to forecast market returns? To appear in International Journal of Computational Economics and Econometrics .

Phengpis, C., Apilado, V.P., 2004. Economic interdependence and common stochastic trends: A comparative analysis between EMU and non-EMU stock markets. International Review of Financial Analysis 13, 245–263.

Rangvid, J., 2001. Increasing convergence among European stock markets? Economics Letters 71, 383–389.

Rangvid, J., Sorensen, C., 2002. Convergence in the ERM and Declining Numbers of Common Stochastic Trends. Journal of Emerging Market Finance 1, 183–213.

Rohan, N., Ramanathan, T.V., 2013. Nonparametric estimation of a time-varying GARCH model. Journal of Nonparametric Statistics 25, 33–52.

Siklos, P.L., Ng, P., 2001. Integration Among Asia-Pacific and International Stock Markets: Common Stochastic Trends and Regime Shifts. Pacific Economic Review 6, 89–110.

Tse, Y.K., Tsui, A.K.C., 2002. A multivariate GARCH with time-varying correlations. Journal of Business and Economic Statistics 20, 351–362.

## Appendix A. Supplementary material

This section contains technical appendices that are meant for publication as a companion to the paper.

### *Appendix A.1. Notation and conventions*

Column vectors are denoted by a bold lower case symbol like $\mathbf{v}$ and $\mathbf{v}^\top$ indicates its transpose. Given a vector $\mathbf{v} \in \mathbb{R}^n$, we denote its components by $v_i$, with $i \in \{1, \ldots, n\}$; we also write $\mathbf{v} = (v_i)_{i \in \{1,\ldots,n\}}$. The symbols $\mathbf{i}_n, \mathbf{0}_n \in \mathbb{R}^n$ stand for the vectors of length $n$ consisting of ones and zeros, respectively. We denote by $\mathbb{M}_{n,m}$ the space of real $n \times m$ matrices with $m, n \in \mathbb{N}$. When $n = m$, we use the symbols $\mathbb{M}_n$ and $\mathbb{D}_n$ to refer to the space of square and diagonal matrices of order $n$, respectively. Given a matrix $A \in \mathbb{M}_{n,m}$, we denote its components by $A_{ij}$ and we write $A = (A_{ij})$, with $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots m\}$. We use $\mathbb{S}_n$ to denote the subspace $\mathbb{S}_n \subset \mathbb{M}_n$ of symmetric matrices:

$$\mathbb{S}_n = \left\{ A \in \mathbb{M}_n \mid A^\top = A \right\},$$

and we use $\mathbb{S}_n^+$ (respectively $\mathbb{S}_n^-$) to refer to the cone $\mathbb{S}_n^+ \subset \mathbb{S}_n$ (respectively $\mathbb{S}_n^- \subset \mathbb{S}_n$) of positive (respectively negative) semidefinite matrices. When $A \in \mathbb{S}_n^+$ (respectively, $A \in \mathbb{S}_n^-$) we write $A \succeq 0$ (respectively, $A \preceq 0$). The symbol $\mathbb{I}_n \in \mathbb{D}_n$ denotes the identity matrix. Given two matrices $A, B \in \mathbb{M}_{n,m}$, we denote by $A \odot B \in \mathbb{M}_{n,m}$ their elementwise multiplication matrix or Hadamard product, that is:

$$(A \odot B)_{ij} := A_{ij} B_{ij} \ \text{ for all } \ i \in \{1, \ldots, n\}, \ \ j \in \{1, \ldots, m\}. \tag{A.1}$$

We denote as Diag the operator $\mathrm{Diag} : \mathbb{M}_n \longrightarrow \mathbb{D}_n$ that sets equal to zero all the components of a square matrix except for those that are on the main diagonal. The operator $\mathrm{diag} : \mathbb{R}^n \longrightarrow \mathbb{D}_n$ takes a given vector and constructs a diagonal matrix with its entries in the main diagonal.

### *Appendix A.2. Stationarity conditions for Model 1 in the nonlinear state-space setup*

In order to provide sufficient conditions for the stationarity of the process, we rewrite (2.8)-(2.10) as

$$\begin{aligned}
\sigma_{1,t}^2 &= a_1 + \delta_1 \sigma_{3,t-1}^2 + \gamma_1 \varepsilon_{3,t-1}^2, \\
\sigma_{2,t}^2 &= (a_2 + a_1 \delta_2) + \delta_1 \delta_2 \sigma_{3,t-1}^2 + \gamma_1 \delta_2 \varepsilon_{3,t-1}^2, \\
\sigma_{3,t}^2 &= (a_3 + a_2 \delta_3 + \alpha_1 \delta_2 \delta_3) + \delta_1 \delta_2 \delta_3 \sigma_{3,t-1}^2 + \gamma_1 \delta_2 \delta_3 \varepsilon_{3,t-1}^2.
\end{aligned}$$

Consider these relations as those defining a VEC model (see Bollerslev et al. (1988)), then stationarity can be ensured by imposing that the spectral radius of the matrix $A$ given by

$$A := \begin{pmatrix} 0 & 0 & \delta_1 + \gamma_1 \\ 0 & 0 & \delta_2(\delta_1 + \gamma_1) \\ 0 & 0 & \delta_2\delta_3(\delta_1 + \gamma_1) \end{pmatrix}$$

is smaller than one Gouriéroux (1997). It is easy to verify that this results in the inequality

$$\delta_2\delta_3(\delta_1 + \gamma_1) < 1. \tag{A.2}$$

*Appendix A.3. Stationarity conditions for Model 2 in the nonlinear state-space setup*

In order to find sufficient stationarity conditions, we proceed by rewriting Model 2 as

$$\begin{aligned}
\sigma_{1,t}^2 =& a_1 + \delta_1\sigma_{3,t-1}^2 + \gamma_1\varepsilon_{3,t-1}^2 + \rho_1\sigma_{1,t-1}^2 + \tau_1\varepsilon_{1,t-1}^2, \\
\sigma_{2,t}^2 =& (a_2 + a_1\delta_2) + \delta_1\delta_2\sigma_{3,t-1}^2 + \gamma_1\delta_2\varepsilon_{3,t-1}^2 + \rho_1\delta_2\sigma_{1,t-1}^2 + \tau_1\delta_2\varepsilon_{1,t-1}^2 + \rho_2\sigma_{2,t-1}^2 + \tau_2\varepsilon_{2,t-1}^2, \\
\sigma_{3,t}^2 =& (a_3 + a_2\delta_3 + a_1\delta_2\delta_3) + (\delta_1\delta_2\delta_3 + \rho_3)\sigma_{3,t-1}^2 + (\gamma_1\delta_2\delta_3 + \tau_3)\varepsilon_{3,t-1}^2 + \rho_1\delta_2\delta_3\sigma_{1,t-1}^2 \\
& + \tau_1\delta_2\delta_3\varepsilon_{1,t-1}^2 + \rho_2\delta_3\sigma_{2,t-1}^2 + \tau_2\delta_3\varepsilon_{2,t-1}^2.
\end{aligned}$$

Proceeding in the way analogous to Model 1 in Section Appendix A.2, we ensure stationarity by requiring that the spectral radius $\rho(A)$ of the matrix $A$ defined by:

$$A := \begin{pmatrix} \rho_1 + \tau_1 & 0 & \delta_1 + \gamma_1 \\ \delta_2(\rho_1 + \tau_1) & \rho_2 + \tau_2 & \delta_2(\delta_1 + \gamma_1) \\ \delta_2\delta_3(\rho_1 + \tau_1) & \delta_3(\rho_2 + \tau_2) & \delta_2\delta_3(\delta_1 + \gamma_1) + \tau_3 + \rho_3 \end{pmatrix},$$

is smaller than one. Since in this case the general expression of the eigenvalues of $A$ is very convoluted, we take advantage of the fact that for any matrix norm $||\cdot||$ the inequality $\rho(A) \leq ||A||$ is satisfied and hence it suffices to require that $||A|| < 1$ to ensure that $\rho(A) < 1$. We implement this condition by using the so called maximum column and row sum norms (see Horn and Johnson (2013)). In the case of the maximum column sum norm, the inequality $||A|| < 1$ amounts to the following three conditions

$$\begin{cases} (\rho_1 + \tau_1)(1 + \delta_2(1 + \delta_3)) < 1, & \text{(A.3a)} \\ (\rho_2 + \tau_2)(1 + \delta_2(1 + \delta_3)) < 1, & \text{(A.3b)} \\ (\delta_1 + \gamma_1)(1 + \delta_2(1 + \delta_3)) + \tau_3 + \rho_3 < 1, & \text{(A.3c)} \end{cases}$$

while the use of the maximum row sum norm results in three other different conditions, namely,

$$\begin{cases} \delta_1 + \gamma_1 + \rho_1 + \tau_1 < 1, & \text{(A.4a)} \\ \delta_2(\delta_1 + \gamma_1 + \rho_1 + \tau_1) + \rho_2 + \tau_2 < 1, & \text{(A.4b)} \\ \delta_2\delta_3(\delta_1 + \gamma_1 + \rho_1 + \tau_1) + \delta_3(\rho_2 + \tau_2) + \tau_3 + \rho_3 < 1. & \text{(A.4c)} \end{cases}$$

Any of these two sets of inequalities may be used as parameter constraints at the time of model estimation in order to ensure stationarity. Nevertheless, imposing different parameter constraints may obviously produce different estimation results.

*Appendix A.4. Proof of Proposition 2.1*

Let $A$ be an element of the general linear group of order five, that is, $A \in GL_5(\mathbb{R})$. Consider the model prescription (2.5a)-(2.5b) and transform it according to the following prescription. First, replace the parameter $B$ by $BA^{-1}A$ in (2.5a) and second, apply $A$ to both sides of (2.5b). This yields:

$$\begin{cases} \mathbf{r}_t = \boldsymbol{\alpha} + BA^{-1}A\boldsymbol{e}_t + \mathbf{u}_t, & \text{(A.5a)} \\ A\boldsymbol{e}_t = AT\boldsymbol{e}_{t-1} + AR\mathbf{v}_{t-1}. & \text{(A.5b)} \end{cases}$$

The model (2.5a)-(2.5b) remains invariant under this transformation if the following conditions hold:

**(i)** $BA^{-1}$ has the same entries structure as $B$.

**(ii)** The matrices $A$ and $T$ commute, that is, $AT = TA$.

**(iii)** $\bar{Q} := AQA^\top$ with $Q := RR^\top$ is a matrix of the same entries structure as $Q$, namely, $\bar{Q} = \mathrm{diag}(\bar{\sigma}_{v,1}^2, \bar{\sigma}_{v,2}^2, \bar{\sigma}_{v,3}^2, 0, 0)$, for some $\bar{\sigma}_{v,1}^2, \bar{\sigma}_{v,2}^2, \bar{\sigma}_{v,3}^2 \in \mathbb{R}^+$.

Indeed, under these hypotheses, the transformed equations (A.5a)-(A.5b) become:

$$\begin{cases} \mathbf{r}_t = \boldsymbol{\alpha} + (BA^{-1})(A\boldsymbol{e}_t) + \mathbf{u}_t, & \text{(A.6a)} \\ (A\boldsymbol{e}_t) = T(A\boldsymbol{e}_{t-1}) + AR\mathbf{v}_{t-1}. & \text{(A.6b)} \end{cases}$$

It is hence easy to see that the model (A.6a)-(A.6b) has the same structure as the original model (2.5a)-(2.5b) with the variables $\boldsymbol{e}_t$ replaced by $(A\boldsymbol{e}_t)$, provided that $BA^{-1}$ has the same entries structure as $B$ and that the covariance matrix $\Sigma(\bar{\mathbf{v}}_t)$ of $\bar{\mathbf{v}}_t := AR\mathbf{v}_t$ is of the form $\bar{Q} = \mathrm{diag}(\bar{\sigma}_{v,1}^2, \bar{\sigma}_{v,2}^2, \bar{\sigma}_{v,3}^2, 0, 0)$, with some $\bar{\sigma}_{v,1}, \bar{\sigma}_{v,2}, \bar{\sigma}_{v,3} \in \mathbb{R}^+$. This covariance matrix equals

$$\Sigma(\bar{\mathbf{v}}_t) := \mathrm{E}\left[\bar{\mathbf{v}}_t\bar{\mathbf{v}}_t^{\top}\right] = \mathrm{E}\left[AR\mathbf{v}_t\mathbf{v}_t^{\top}R^\top A^\top\right] = ARR^\top A^\top = AQA^\top \tag{A.7}$$

with $Q := RR^\top$.

We first study what the implications that conditions **(i)-(iii)** have in the structure of $A \in GL_5(\mathbb{R})$. First, by point **(iii)** suppose that $A \in GL_5(\mathbb{R})$ is such that for any $\sigma_{v,1}, \sigma_{v,2}, \sigma_{v,3} \in \mathbb{R}^+$ there exist $\bar{\sigma}_{v,1}, \bar{\sigma}_{v,2}, \bar{\sigma}_{v,3} \in \mathbb{R}^+$ such that

$$A \cdot \begin{pmatrix} \sigma_{v,1}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{v,2}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{v,3}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot A^\top = \begin{pmatrix} \bar{\sigma}_{v,1}^2 & 0 & 0 & 0 & 0 \\ 0 & \bar{\sigma}_{v,2}^2 & 0 & 0 & 0 \\ 0 & 0 & \bar{\sigma}_{v,3}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{A.8}$$

Define now $\Sigma_v := \begin{pmatrix} \sigma_{v,1}^2 & 0 & 0 \\ 0 & \sigma_{v,2}^2 & 0 \\ 0 & 0 & \sigma_{v,3}^2 \end{pmatrix}$, $\bar{\Sigma}_v := \begin{pmatrix} \bar{\sigma}_{v,1}^2 & 0 & 0 \\ 0 & \bar{\sigma}_{v,2}^2 & 0 \\ 0 & 0 & \bar{\sigma}_{v,3}^2 \end{pmatrix}$, and let $K, P \in \mathbb{M}_3$, $C, W \in \mathbb{M}_{3,2}$, $D, X \in \mathbb{M}_{2,3}$, $E, U \in \mathbb{M}_2$ be such that $A = \left(\begin{array}{c|c} K & C \\ \hline D & E \end{array}\right)$, $A^\top = \left(\begin{array}{c|c} K^\top & D^\top \\ \hline C^\top & E^\top \end{array}\right)$, and $A^{-1} = \left(\begin{array}{c|c} P & W \\ \hline X & U \end{array}\right)$. Condition (A.8), namely, $AQA^\top = \bar{Q}$ is equivalent to $A^{-1}AQA^\top = A^{-1}\bar{Q}$ or to $QA^\top = A^{-1}\bar{Q}$ which in the notation that we just introduced amounts to

$$\left(\begin{array}{c|c} \Sigma_v & 0 \\ \hline 0 & 0 \end{array}\right) \left(\begin{array}{c|c} K^\top & D^\top \\ \hline C^\top & E^\top \end{array}\right) = \left(\begin{array}{c|c} P & W \\ \hline X & U \end{array}\right) \left(\begin{array}{c|c} \bar{\Sigma}_v & 0 \\ \hline 0 & 0 \end{array}\right). \tag{A.9}$$

Expression (A.9) is equivalent to the following three conditions:

$$\Sigma_v K^\top = P\bar{\Sigma}_v, \tag{A.10}$$

$$\Sigma_v D^\top = 0, \tag{A.11}$$

$$X\bar{\Sigma}_v = 0. \tag{A.12}$$

We continue by noticing that since $\Sigma_v$ and $\bar{\Sigma}_v$ are invertible, the expressions (A.11) and (A.12) amount to $D^\top = 0$ and $X = 0$, respectively. This shows that $A = \left(\begin{array}{c|c} K & C \\ \hline 0 & E \end{array}\right)$ and $A^{-1} = \left(\begin{array}{c|c} P & W \\ \hline 0 & U \end{array}\right)$. We now impose the condition **(ii)**, that is, $AT = TA$:

$$\left(\begin{array}{c|c} K & C \\ \hline 0 & E \end{array}\right)\left(\begin{array}{c|c} 0 & 0 \\ \hline M & 0 \end{array}\right) = \left(\begin{array}{c|c} 0 & 0 \\ \hline M & 0 \end{array}\right)\left(\begin{array}{c|c} K & C \\ \hline 0 & E \end{array}\right) \tag{A.13}$$

with $M := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. This relation implies that

$$CM = 0, \tag{A.14}$$

$$EM = MK, \tag{A.15}$$

$$0 = MC. \tag{A.16}$$

The expressions (A.14) and (A.16) imply that $C = 0$, which yields that

$$A = \left(\begin{array}{c|c} K & 0 \\ \hline 0 & E \end{array}\right). \tag{A.17}$$

As $A$ is by definition invertible, in view of (A.17) so are the submatrices $K$ and $E$, and hence in the block structure of $A^{-1}$ we can set $W = 0$, $P = K^{-1}$, and $U = E^{-1}$, respectively, that is,

$$A^{-1} = \left(\begin{array}{c|c} K^{-1} & 0 \\ \hline 0 & E^{-1} \end{array}\right). \tag{A.18}$$

At the same time it is easy to verify that the relation (A.15) implies that

$$K = \left(\begin{array}{c|cc} k_{11} & k_{12} & k_{13} \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$E$}} \\ 0 & & \end{array}\right). \tag{A.19}$$

Let now denote by $k^*_{ij}$ and by $e^*_{ij}$ with $i, j \in \{1, 2, 3\}$ the generic entries of the matrices $K^{-1}$ and $E^{-1}$, respectively. We may hence write by (A.19) that

$$K^{-1} = \left(\begin{array}{c|cc} \frac{1}{k_{11}} & k^*_{12} & k^*_{13} \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$E^{-1}$}} \\ 0 & & \end{array}\right) = \begin{pmatrix} \frac{1}{k_{11}} & k^*_{12} & k^*_{13} \\ 0 & e^*_{11} & e^*_{12} \\ 0 & e^*_{21} & e^*_{22} \end{pmatrix}. \tag{A.20}$$

We now use the fact that condition **(i)** requires that the matrix $BA^{-1}$ has the same structure as $B$, that is, there exist some $\bar{\beta}_1, \bar{\beta}_2, \bar{\beta}_3 \in \mathbb{R}$ such that

$$BA^{-1} = \begin{pmatrix} \bar{\beta}_1 & 0 & 0 & \bar{\beta}_1 & \bar{\beta}_1 \\ \bar{\beta}_2 & \bar{\beta}_2 & 0 & 0 & \bar{\beta}_2 \\ \bar{\beta}_3 & \bar{\beta}_3 & \bar{\beta}_3 & 0 & 0 \end{pmatrix}. \tag{A.21}$$

We first partition the matrix $B$ and write it as $B := (B_1|B_2)$, with

$$B_1 = \begin{pmatrix} \beta_1 & 0 & 0 \\ \beta_2 & \beta_2 & 0 \\ \beta_3 & \beta_3 & \beta_3 \end{pmatrix}, \quad B_2 = \begin{pmatrix} \beta_1 & \beta_1 \\ 0 & \beta_2 \\ 0 & 0 \end{pmatrix}. \tag{A.22}$$

We now use (A.18), (A.22) and write

$$BA^{-1} = (B_1|B_2) \cdot \left(\begin{array}{c|c} K^{-1} & 0 \\ \hline 0 & E^{-1} \end{array}\right) = (B_1K^{-1}|B_2E^{-1})$$

which by (A.21) requires both

$$\begin{pmatrix} \beta_1 & 0 & 0 \\ \beta_2 & \beta_2 & 0 \\ \beta_3 & \beta_3 & \beta_3 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{k_{11}} & k^*_{12} & k^*_{13} \\ 0 & e^*_{11} & e^*_{12} \\ 0 & e^*_{21} & e^*_{22} \end{pmatrix} = \begin{pmatrix} \bar{\beta}_1 & 0 & 0 \\ \bar{\beta}_2 & \bar{\beta}_2 & 0 \\ \bar{\beta}_3 & \bar{\beta}_3 & \bar{\beta}_3 \end{pmatrix} \tag{A.23}$$

and

$$\begin{pmatrix} \beta_1 & \beta_1 \\ 0 & \beta_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} e^*_{11} & e^*_{12} \\ e^*_{21} & e^*_{22} \end{pmatrix} = \begin{pmatrix} \bar{\beta}_1 & \bar{\beta}_1 \\ 0 & \bar{\beta}_2 \\ 0 & 0 \end{pmatrix}. \tag{A.24}$$

The relations implied by the matrix equation (A.23) for the entries (1,2), (1,3), and (2,3) yield that $k^*_{12} = 0$, $k^*_{13} = 0$, and $e^*_{12} = 0$, respectively. At the same time, the relation for the component (2,1) of the matrix equation (A.24) yields that $e^*_{21} = 0$. Consequently, both $K^{-1}$ and $E^{-1}$ are diagonal matrices. Finally, the relation (A.23) computed for the corresponding diagonal elements of $K^{-1}$ implies that that $e^*_{11} = \frac{1}{k_{11}}$, $e^*_{22} = \frac{1}{k_{11}}$ and we can hence write that

$$K = \lambda \mathbb{I}_3, \quad \lambda \in \mathbb{R}.$$

Consequently, by (A.19)

$$E = \lambda \mathbb{I}_2,$$

which automatically guarantees by (A.17) that $A = \lambda \mathbb{I}_5$, necessarily. This implies that the matrix $B$ in the model (2.5a)-(2.5b) is defined up to multiplication by a homothety. It is hence sufficient to impose one of the following two constraints so that the model is well identified: (i) one of the elements $\beta_i$, $i \in \{1, 2, 3\}$, is set equal to some constant number, (ii) one of the variances $\sigma^2_{v,i}$, $i \in \{1, 2, 3\}$ is set equal to an arbitrary positive constant number. ■

*Appendix A.5. Proof of Proposition 3.1*

**Proof of part (i).** The recursion that defines the tvGARCH(1,1) model in (3.11) implies that (see for example formula (2.2) in Ambroževičit and Klivečka (2008)):

$$\sigma^2_{t_i} = \sum_{j=0}^{\infty} a_{t_i - j} \prod_{k=1}^{j} \left( \gamma_{t_i-k+1} v^2_{t_i-k} + \delta_{t_i-k+1} \right) = \sum_{j=0}^{\infty} a_{t_i-j} \left[ b_{t_i} b_{t_i-1} \cdots b_{t_i-j+1} \right], \tag{A.25}$$

where $b_{t_i-k+1} := \left( \gamma_{t_i-k+1} v^2_{t_i-k} + \delta_{t_i-k+1} \right)$. We notice that the process $\{b_{t_i}\}$ is made of positive independent random variables. Moreover, by the Cauchy rule for series with non-negative terms, expression (A.25) converges if

$$\lambda := \lim_{j \to \infty} \left[ b_{t_i} b_{t_i-1} \cdots b_{t_i-j+1} \right]^{1/j} < 1.$$

We therefore compute:

$$\begin{aligned} \lim_{j\to\infty} \left[ b_{t_i} b_{t_i-1} \cdots b_{t_i-j+1} \right]^{1/j} &= \lim_{j\to\infty} \exp \left[ \frac{1}{j} \sum_{k=1}^{j} \log \left( b_{t_i-k+1} \right) \right] = \exp \lim_{j\to\infty} \left[ \frac{1}{j} \sum_{k=1}^{j} \log \left( b_{t_i-k+1} \right) \right] \\ &= \exp \frac{1}{3} \sum_{l=1}^{3} \mathrm{E} \left[ \log \left( \gamma_l v_t^2 + \delta_l \right) \right] \leq \exp \frac{1}{3} \sum_{l=1}^{3} \log \left( \mathrm{E} \left[ \left( \gamma_l v_t^2 + \delta_l \right) \right] \right) &\text{(A.26)} \\ &= \left[ (\delta_1 + \gamma_1)(\delta_2 + \gamma_2)(\delta_3 + \gamma_3) \right]^{1/3}, &\text{(A.27)} \end{aligned}$$

where the first equality in (A.26) follows from the strong law of large numbers and the relation that follows it is a consequence of Jensen's inequality. The inequality in the statement implies hence by (A.27) that $\lambda := \lim_{j\to\infty} \left[ b_{t_i} b_{t_i-1} \cdots b_{t_i-j+1} \right]^{1/j} < 1$. A strategy mimicking, for example, the proof of Theorem 2.1 in Francq and Zakoian (2010), shows that in that situation model 1 has a unique stationary solution.

**Proof of part (ii).** By Theorem 2.4 in Francq and Zakoian (2010), it suffices to show that the top Lyapunov exponent $\gamma$ of the sequence $\{A_{t_i}\}$ is smaller than zero. By Theorem 2.3 in Francq and Zakoian

(2010):

$$\gamma = \lim_{t_i \to \infty} \frac{1}{t_i} \mathrm{E}\left[\log \|A_{t_i} A_{t_i-1} \cdots A_1\|\right],$$

with $\|\cdot\|$ any matrix norm. We now use the norm $\|A\| = \sum_{i,j} |a_{ij}|$ and notice that if all the elements of $A$ are positive then

$$\mathrm{E}\left[\|A\|\right] = \|\mathrm{E}\left[A\right]\|. \tag{A.28}$$

Consequently,

$$\begin{aligned} \gamma &= \lim_{t_i \to \infty} \frac{1}{t_i} \mathrm{E}\left[\log \|A_{t_i} A_{t_i-1} \cdots A_1\|\right] \leq \lim_{t_i \to \infty} \frac{1}{t_i} \log\left(\mathrm{E}\left[\|A_{t_i} A_{t_i-1} \cdots A_1\|\right]\right) & \text{(A.29)} \\ &= \lim_{t_i \to \infty} \frac{1}{t_i} \log\left(\|\mathrm{E}\left[A_{t_i} A_{t_i-1} \cdots A_1\right]\|\right) = \lim_{t_i \to \infty} \frac{1}{t_i} \log\left(\|\mathrm{E}\left[A_{t_i}\right] \mathrm{E}\left[A_{t_i-1}\right] \cdots \mathrm{E}\left[A_1\right]\|\right) & \text{(A.30)} \\ &= \frac{1}{3} \lim_{t \to \infty} \frac{1}{t} \log\left(\|A_3 A_2 A_1\|^t\right) = \frac{1}{3} \log\left(\rho(A_3 A_2 A_1)\right) = \log\left(\rho(A_3 A_2 A_1)^{1/3}\right). & \text{(A.31)} \end{aligned}$$

The relation in (A.29) follows from Jensen's inequality, the first equality in (A.30) is a consequence of (A.28) and the second one of the independence of the elements in the process $\{A_{t_i}\}$. Finally, in (A.31) we use Gelfand's formula for the characterization of the spectral radius of a matrix. The inequality $\gamma \leq \log\left(\rho(A_3 A_2 A_1)^{1/3}\right)$ that we just proved guarantees that the condition in the statement ensures that $\log\left(\rho(A_3 A_2 A_1)^{1/3}\right) < 0$ and hence $\gamma < 0$, as required.

In both cases, the arguments that we provided show that the unconditional variance $\mathrm{E}\left[\sigma_{t_i}^2\right]$ depends only on $i$ and hence establishes the periodic stationarity claimed in the statement. ■

### *Appendix A.6. Empirical performance of the GST-based volatility forecasting schemes using a smaller estimation sample*

In this appendix we illustrate the robustness of the results obtained in the empirical study in Section 5 by performing a similar analysis based on the same dataset but using a shorter estimation period (January 5, 1996 – December 4, 2006) that does not contain the volatility events in the Fall 2008. The out-of-sample study in that case comprises the entire Great Recession (December 5, 2006 – April 1, 2015). This choice yields a dataset with a length of $T := 4581$ observations and for which $T_{\mathrm{est}} := 2600$ and $T_{\mathrm{out}} := 1981$.

The study follows the same scheme as the one in Section 5. In particular, we consider the same competing models and the same loss functions at the time of implementing the MCS strategy. In the case of the covariance loss functions (5.2), the results of the corresponding MCS comparison are contained in the Tables A.6 and A.7. These results correspond to the values of the covariance loss functions obtained in 185 different out-of-sample time intervals of the form $\{T_{\mathrm{est}}+1, \ldots, T_{\mathrm{est}}+141+10k\}$ with ,

| | DCC models | | Linear state-space models | | | | | | | | Nonlinear state-space models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | |
| | | | Model 1 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | Model 2 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | |
| DCC models | S | H | S | H | S | H | S | H | S | H | S | H | S | H |
| $k$ | 11 | 21 | 30 | 50 | 30 | 50 | 42 | 62 | 36 | 56 | 25 | 45 | 31 | 51 |
| AIC | -48891.31$^{13}$ | -48871.32$^{14}$ | -54918.55$^{5}$ | -54878.41$^{7}$ | -54901.16$^{6}$ | -54860.66$^{8}$ | **-55320.41$^{1}$** | -55280.41$^{2}$ | -55094.39$^{3}$ | -55054.52$^{4}$ | -51076.04$^{9}$ | -51043.56$^{11}$ | -51064.04$^{10}$ | -51031.56$^{12}$ |
| BIC | -48826.82$^{13}$ | -48748.19$^{14}$ | -54742.66$^{4}$ | -54585.25$^{7}$ | -54725.26$^{6}$ | -54567.49$^{8}$ | **-55074.15$^{1}$** | -54916.88$^{2}$ | -54883.31$^{3}$ | -54726.17$^{5}$ | -50929.46$^{9}$ | -50779.72$^{11}$ | -50882.28$^{10}$ | -50732.54$^{12}$ |
| cAIC | -48891.21$^{13}$ | -48870.96$^{14}$ | -54917.83$^{5}$ | -54876.41$^{7}$ | -54900.44$^{6}$ | -54858.66$^{8}$ | **-55318.99$^{1}$** | -55277.33$^{2}$ | -55093.35$^{3}$ | -55052.01$^{4}$ | -51075.53$^{9}$ | -51041.94$^{11}$ | -51063.27$^{10}$ | -51029.48$^{12}$ |
| $\log L$ | 24456.66$^{14}$ | 24456.66$^{13}$ | 27489.28$^{5}$ | 27489.21$^{6}$ | 27480.58$^{7}$ | 27480.33$^{8}$ | **27702.20$^{1}$** | 27702.20$^{2}$ | 27583.20$^{4}$ | 27583.26$^{3}$ | 25563.02$^{11}$ | 25566.78$^{9}$ | 25563.02$^{12}$ | 25566.78$^{10}$ |

Table A.5: Total number of model parameters $k$, values the log-likelihood function $\log L$, and associated AIC, cAIC, and BIC statistics. The largest values of the log-likelihood function and the smallest values of the information criteria are displayed in red bold. Exponents of the values at each row indicate the rank of the model from 14 (the worse) to 1 (the best).

$k = \{0, 1, \ldots, 184\}$. The first 141 elements in the out-of-sample period are included in all these intervals in order to ensure that there are enough values available for the bootstrapping process that is necessary in the estimation of the distribution of the model equivalence test statistic. The date corresponding to the end of this offset interval is July 10, 2007.

| | DCC models | | Linear state-space models | | | | | | | | Nonlinear state-space models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | |
| | | | Model 1 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | Model 2 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | |
| DCC models | S | H | S | H | S | H | S | H | S | H | S | H | S | H |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{(t-1)_3}}$ | **180** | 68 | 74 | 73 | 51 | 51 | 72 | 8 | 68 | 67 | 2 | 2 | 2 | 2 |
| Sum $p$-vals | **160.508** | 19.651 | 44.471 | 23.938 | 16.650 | 15.687 | 21.398 | 7.357 | 21.305 | 20.129 | 6.028 | 6.028 | 6.028 | 6.111 |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{t_1}}$ | 4 | 3 | **133** | 4 | 4 | 4 | 56 | 4 | 3 | 3 | 1 | 1 | 1 | 1 |
| Sum $p$-vals | 6.942 | 6.863 | **132.809** | 7.138 | 6.981 | 6.981 | 59.329 | 6.981 | 6.949 | 6.902 | 6.413 | 6.413 | 6.413 | 6.413 |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{t_2}}$ | 5 | 4 | **178** | 5 | 5 | 5 | 11 | 5 | 6 | 5 | 2 | 2 | 2 | 2 |
| Sum $p$-vals | 6.891 | 6.784 | **176.400** | 7.143 | 6.923 | 6.923 | 13.111 | 6.923 | 9.661 | 7.029 | 6.351 | 6.351 | 6.351 | 6.351 |

Table A.6: Model confidence sets (MCS) constructed using the covariance based loss functions (5.2) for 185 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\mathrm{est}} + 141 + 10k$, $k \in \{0, 1, \ldots, 184\}$. The letters 'S' and 'H' stand for 'Scalar' and 'Hadamard', respectively. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at a 90% confidence level; the value underneath indicates the sum of all the MCS $p$-values obtained by a given model in the 185 tests. The best performing models (determined by the number of times included in the MCS) for the considered information set are marked in bold red.

The results corresponding to the KLIC loss functions (5.4) are contained in Table A.8 for the significance levels 10% and 25% (they are identical). These values are obtained out of 185 different out-of-sample time intervals of the form $\{T_{\mathrm{est}} + 1, \ldots, T_{\mathrm{est}} + 141 + 10k\}$ with $k = \{0, 1, \ldots, 184\}$. As the previous MCS experiment constructed using covariance based loss functions already showed, *the forecasting approaches based on the linear state-space Models 1 and 2 significantly outperform the standard DCC models in this context.* The results that we just obtained show the robustness of the study conducted in Section 5.2 with respect to the choice of estimation period and the number of the out-of-sample observations.

| | DCC models | | Linear state-space models | | | | | | | | Nonlinear state-space models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | |
| | | | Model 1 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | Model 2 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | |
| DCC models | S | H | S | H | S | H | S | H | S | H | S | H | S | H |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{(t-1)_3}}$ | **161** | 8 | 33 | 15 | 5 | 5 | 12 | 1 | 13 | 11 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | **160.519** | 19.676 | 44.477 | 23.940 | 16.679 | 15.717 | 21.411 | 7.385 | 21.336 | 20.142 | 6.045 | 6.045 | 6.045 | 6.128 |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{t_1}}$ | 0 | 0 | **131** | 0 | 0 | 0 | 54 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | 6.940 | 6.858 | **132.811** | 7.135 | 6.980 | 6.980 | 59.323 | 6.980 | 6.947 | 6.898 | 6.406 | 6.406 | 6.406 | 6.406 |
| MCS $d^{\mathrm{Cov}}_{\mathcal{F}^*_{t_2}}$ | 0 | 0 | **176** | 0 | 0 | 0 | 6 | 0 | 3 | 0 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | 6.890 | 6.784 | **176.399** | 7.143 | 6.922 | 6.922 | 13.112 | 6.922 | 9.659 | 7.026 | 6.353 | 6.353 | 6.353 | 6.353 |

Table A.7: Model confidence sets (MCS) constructed using the covariance based loss functions (5.2) for 185 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\mathrm{est}} + 141 + 10k$, $k \in \{0, 1, \ldots, 184\}$. The letters 'S' and 'H' stand for 'Scalar' and 'Hadamard', respectively. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at a 75% confidence level; the value underneath indicates the sum of all the MCS $p$-values obtained by a given model in the 185 tests. The best performing models (determined by the number of times included in the MCS) for the considered information set are marked in bold red.

| | DCC models | | Linear state-space models | | | | | | | | Nonlinear state-space models | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | | | | Model 1 for $\{\widehat{\varepsilon}_t\}$ | | Model 2 for $\{\widehat{\varepsilon}_t\}$ | |
| | | | Model 1 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | Model 2 for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | | GARCH for $\{\widehat{\mathbf{u}}_t\}$ | |
| DCC models | S | H | S | H | S | H | S | H | S | H | S | H | S | H |
| MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{(t-1)_3}}$ | 0 | 0 | 0 | **116** | 0 | 0 | 0 | 70 | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | 0.000 | 0.000 | 0.000 | **116.000** | 0.000 | 0.000 | 0.000 | 70.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_1}}$ | 0 | 0 | 0 | 27 | 0 | 0 | 0 | **158** | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | 0.000 | 0.000 | 0.000 | 27.000 | 0.000 | 0.000 | 0.000 | **158.000** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| MCS $d^{\mathrm{KLIC}}_{\mathcal{F}^*_{t_2}}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **185** | 0 | 0 | 0 | 0 | 0 | 0 |
| Sum $p$-vals | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | **185.000** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Table A.8: Model confidence sets (MCS) constructed using the KLIC loss functions (5.4), for 185 different out-of-sample lengths $l(k)$, namely for $l(k) = T_{\mathrm{est}} + 141 + 10k$, $k \in \{0, 1, \ldots, 184\}$. The letters 'S' and 'H' stand for 'Scalar' and 'Hadamard', respectively. For each model and information set under consideration the corresponding value indicates the number of times that model has been included in the MCS at the 90% and 75% confidence levels (they are identical); the value underneath indicates the sum of all the MCS $p$-values obtained by a given model in the 185 tests. The best performing models for the considered information set are marked in bold red.