Lancaster University Management School

# Essays on Models with Mixed Frequency Data and Time Varying Parameters

**Aya Ghalayini, BSc, MFin**

Supervisors: Prof. Marwan Izzeldin & Prof. Mike Tsionas

Department of Economics

Lancaster University Management School

A thesis submitted for the degree of

*Doctor of Philosophy*

July, 2022

I would like to dedicate this thesis to my family.

# Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my own work. The material has not been submitted, either in whole or in part, for a degree at this, or any other university.

Aya Ghalayini

**Essays on Models with Mixed Frequency Data and Time Varying Parameters**

Aya Ghalayini, BSc, MFin.

Department of Economics, Lancaster University Management School

A thesis submitted for the degree of *Doctor of Philosophy*. July, 2022

# Abstract

This thesis focuses on two statistical challenges in time-series modelling. The first is when variables' observations are available at different frequencies. The second is when the coefficients of a model are time-varying with stochastic volatility. The impact of these challenges and the value of the suggested remedies are assessed in empirical financial-economic applications.

In addressing the first statistical challenge, disaggregation from the low- to the high-frequency domain is one of the methods that has long been used in several pieces of literature. The first chapter evaluates the existing disaggregation methods with thorough comparisons to provide comprehensive guidance for an empirical user. The second chapter builds on these results to examine the value-added in forecasting the volatility of financial stock prices by incorporating information from variables with mixed frequencies such as market sentiment indicators, economic variables, and activity measures. A representative factor(s) of all potential predictors from both frequencies results in significant forecast gains in predicting long-term financial volatility even during the 2007-08 financial crisis.

The third chapter proposes a state-space model to incorporate features of time-varying coefficients to reflect the dynamic relationship between the dependent and the explanatory variables. Mainly, the model consists of two hierarchical states. First, the time-varying coefficients follow an autoregressive (AR) process with heteroskedastic innovations. Second, the log-transformation of the conditional variance of these innovations is also modelled as an AR process. In an empirical study, we utilize the proposed methodology to forecast the volatility of financial stock prices. We find that the proposed features consistently and significantly enhance the forecasting accuracy compared to a benchmark model and its existing variants.

# Acknowledgements

I am very grateful for all the blessings that Allah endowed me with throughout my learning journey. First and foremost, I want to thank my supervisors, Professor Marwan Izzeldin and Professor Mike Tsionas, for their advice, inspiration, patience and continued support. I especially want to thank Marwan for training me on high-frequency financial data and helping me grow as an academic in multiple dimensions. I want to thank Mike specifically for teaching me Bayesian Analysis and equipping me with the necessary skills.

I am grateful to Dr Adriana Cornea-Madeira, whose feedback and the conversations we had were very inspiring for me to start my PhD research. I am also very thankful to my high school and university teachers for their passionate dedication - particularly Dr Ibrahim Jamali and Dr Wassim Dbouk, for their support during my postgraduate studies, PhD application, and beyond.

Special praise goes to Dr Gerry Steele, whose editorial input and comprehensive feedback helped me improve my writing skills. I appreciate the fruitful discussions I had with Dr Vasilis Pappas, even beyond the scope of this work. I want to also thank my friend and academic mentor, Dr Layal al-Hakim, for all her support, especially last year.

I want to thank my fellow PhD colleagues for the exciting discussions and especially Sherry, Sara, and Afnan, for their friendship. I wish them all the best and hope to keep in touch. I am also thankful to the academic and professional staff of the Economics Department at Lancaster University. I particularly want to thank Caren, Craig, and Sarah for their positive spirit and administrative support.

My gratitude goes to Lancaster University Management School, the Economics Department, and the Gulf One Lab for Computational and Economic Research (GOLCER). The prestigious LUMS PhD studentship made this work possible. Their and GOLCER's financial support helped me attend workshops and present my work at conferences. Further, the Data used in this thesis was provided by GOLCER, and computations were performed using The High-End Computing facility at Lancaster University.

I am very appreciative to the dissertation committee members, Professor Lorenzo Trapani and Dr Stefano Soccorsi, for their time, reading, and valuable comments on this thesis.

Last but not least, I sincerely appreciate the unconditional support from my family and friends throughout the years. I am immensely grateful to my devoted husband, Rami Chehab, for his emotional and intellectual support; he believes in me, pushes me to dream big, and encourages me to pursue my dreams and the work I enjoy. I am indebted to my loving parents, Mokhtar Mohamad and Rima, for their endless sacrifices and how they raised me; they grew my curiosity and taught me the value of knowledge. I am thankful to my siblings, Amira, Mahmoud, and Majed, for their inspirational work, which motivates me to succeed. I am also grateful to my aunt Lamia and in-laws, Sami, Hilda, and Omar, for being very passionate and caring. Finally, I wish to thank my friends Aysegul, Laiyan, Zeena, and Aziz, who helped me grow spiritually and constantly reminded me to look after my well-being.

Thank you from the heart!

# Contents

# Chapter 1

# Evaluation of interpolation and extrapolation methods of low-frequency series to the high-frequency domain

## Abstract

We revisit the temporal disaggregation methods in the Chow-Lin (CL) framework. First, we extend the dynamic Chow-Lin model of Poissonnier (2018) to estimate high-frequency extrapolated and interpolated data points. Second, we show that CL methods can be expressed as a linear transformation function in a matrix form. Specifically, the estimated high-frequency series is the sum of two signals. The underlying assumptions drive the first signal. In contrast, the second signal combines the latter and the chosen indicator variable(s). Such decomposition reveals which factors influence the accuracy of both interpolated and extrapolated estimates. Lastly, we perform a Global Monte Carlo study to compare the models under different statistical scenarios. The study guides practitioners in choosing the most appropriate temporal disaggregation method for their application.

**Keywords:** Chow-Lin interpolation Extrapolation Mixed Frequency Temporal Disaggregation

## 1.1 Introduction

It is often the case that the financial and macroeconomic series are observed at different frequencies. Researchers use the mixed data sampling (MIDAS) model by Ghysels, Santa-Clara, and Valkanov (2004) to incorporate high-frequency (HF) variables to estimate a low-frequency (LF) variable. In contrast, Ghysels (2016) and Foroni, Guérin, and Marcellino (2018) propose the mixed-frequency VAR and Reverse Unrestricted MIDAS model, respectively, to incorporate LF information to predict a HF variable.

This paper's interest lies in the frequency conversion from low to high, including monthly to daily, which permits the study of models at a HF level. It is particularly interesting to practitioners who believe that HF forecasting accuracy is enhanced when LF variables are available on a HF level. Interpolation refers to generating HF estimates of a specific LF variable between two of its LF observations. Whereas generating HF values beyond a LF observation is expressed as extrapolation (Chow and Lin, 1971). By definition, extrapolation estimates future HF values of a LF variable by assuming that its unobserved HF values will continue to follow a particular hypothesised model.

The most prominent interpolation-extrapolation method in the literature is the one proposed by Chow and Lin (1971). More recently, a number of researchers extended the Chow-Lin (CL) methodology to dynamic models. However, their primary interest lay in temporal disaggregation (interpolation) rather than inter-and extrapolation. For example, Poissonnier (2018) mention that extrapolation can be obtained in a similar approach to the interpolation method but do not illustrate the extrapolation adaption to their proposed models. Furthermore, CL methodology is popular in applications with constant frequency ratios such as quarterly to monthly or yearly to quarterly. Researchers suggest that it can also be implemented, in a similar fashion, for non-constant frequency ratios such as monthly to daily. However, such implementation in the case of non-constant and large frequency ratio was neither demonstrated nor assessed.

There are multiple methods introduced in the literature. While all have common grounds, the difference is in the proposed relationship between the unobserved HF values of the LF variable and the HF indicator variable(s). In particular, Chow and Lin (1971), Fernández (1981), and Litterman (1983) assume a static model with a first-order auto-regression, unit root, and ARIMA(1,1,0) stochastic disturbance terms, respectively. Poissonnier (2018) presents a dynamic model with a general data generating process (DGP), such that the stochastic disturbance terms may follow either an AR, unit root or ARIMA process.

According to the authors above, each of these models is deemed helpful depending on the setting of the application. For example, the traditional CL method is more suited when the data are generated from a static model. On the other hand, it is more sensible to apply dynamic models when it is assumed that the observations possess some stagnant characteristics. Although the intuition behind these arguments is sound, to the best of our knowledge, there is a lack of theoretical justification and thorough empirical comparison of these models in the literature. Hence, the purpose of this paper is to address the following points. i) The algebraic difference between the methods. ii) The consequences of using a misspecified model to obtain in-sample and out-of-sample estimates of unobserved HF values under the CL framework. iii) The recommendation for practitioners interested in estimating unobserved HF in-sample and out-of-sample values of a LF variable.

In Section 1.2, we summarize the main models and their estimation procedures within the CL framework. Then, we present the theoretical properties of the hypothetical estimates in Section 1.3. We show that all the HF estimates obtained by the temporal disaggregation methods in the CL framework can be decomposed into the sum of two series: 1) one that relies on the assumptions underlying the model and 2) another that depends on the assumptions as well as the observed HF indicator variable(s). Exploring the properties of each signal unveils the consequence of each input on the estimation. Section 1.4 illustrates the feasible estimates of the unobserved HF estimates for each method. Section 1.5 provides an extensive simulation study based on multiple pre-assigned DGP, then obtains the feasible HF estimates of each simulated series using all the mentioned models in the Chow-Lin framework. The estimated series are compared with the true simulated ones to evaluate the models under various specifications. Section 1.6 concludes.

## 1.2 Chow-Lin Framework: a Linear Transformation Function

To apply the timescale switching techniques from low to high, the practitioner needs to consider the following ingredients: the first, second, third, and fourth, respectively, as below.

**Ingredients for time scale switching techniques**

1. Observe the LF observations, $\underset{(N\times1)}{\mathbf{x}_{lf}}$, of the variable of interest.
2. Design the temporal aggregation method that defines the relationship between HF values, $\underset{(c\times1)}{\mathbf{x}}$, and its LF counterparts, $\underset{(N\times1)}{\mathbf{x}_{lf}}$. Where $c$ represents the number of HF data points in the sample period where the LF observations, $\mathbf{x}_{lf}$, occurred. Note that $c > N$ since there are more HF than the LF data points in a given period.
3. Specify the indicator variables, $\underset{(m\times k)}{\mathbf{S}}$, where $k$ is the number of indicator variables used, $m$ is the number of HF observations in the period of interest such that $m \geq c$ and $m - c$ is the number of extrapolated HF data points.
4. Specify a hypothesized relationship model between $\underset{(m\times1)}{\mathbf{z}}$ and $\underset{(m\times k)}{\mathbf{S}}$, where $\{\mathbf{z}_i\}_{i=1}^{c} := \mathbf{x}$ are the interpolated data points and $\{\mathbf{z}_i\}_{i=c+1}^{m}$ are the extrapolated data points.

The first and third ingredients are self-explanatory. The second ingredient is not usually specified for series when the frequency ratio of low- and high-frequency variables (periodicity of the flows) is not constant. There are various aggregation set-up schemes based on the assumed relationship between the HF and LF observations of the variable of interest. Let $\mathbf{C}$ be the aggregation matrix that transforms the unknown interpolated HF vector $\mathbf{x}$ to the LF observed vector $\mathbf{x}_{lf}$ such that $\mathbf{x}_{lf} := \mathbf{C}\mathbf{x}$. For example, if the observed LF vector, $\mathbf{x}_{lf}$, and the unknown interpolated HF vector, $\mathbf{x}$, are defined as:

$$\mathbf{x}_{lf} := (X_{f_0}, X_{f_1}, \cdots, X_{f_{N-1}})', \qquad \text{where} \qquad f_j = f_{j-1} + \kappa_j = 1 + \sum_{i=1}^{j} \kappa_i, \tag{1.1}$$

where $\{\kappa_i\}_{i=1}^{N}$ is the number of HF periods in each LF period $i$ (i.e. the number of days in a month[1]) and $f_0 := 1$.

$$\mathbf{x} := (X_1, X_2, \cdots, X_c)', \qquad \text{where} \qquad c := f_{N-1}$$

then, in this case, the aggregation matrix, $\mathbf{C_1}$, is:

$$\underset{(N\times c)}{\mathbf{C_1}} := \begin{bmatrix} 1^{(1)} & 0^{(2)} \cdots & 0 & \cdots & 0^{(f_{N-1})} \\ 0^{(1)} & \cdots 1^{(f_1)} \cdots & 0 & \cdots & 0 \\ 0^{(1)} & 0^{(2)} \cdots 1^{(f_2)} \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0^{(1)} & \cdots & 0 & \cdots & 1^{(f_{N-1})} \end{bmatrix} \tag{1.2}$$

Whereas, if the observed LF variable, $\mathbf{x}_{lf}$, is assumed as follows:

$$\{\mathbf{x}_{lf}\}_i := \frac{1}{\kappa_i} \sum_{j=f_{i-1}}^{f_i - 1} X_j\,, \qquad \text{where} \qquad i = 1, \cdots, N \tag{1.3}$$

[1]For example, if $i = 1$ represents January 1999 this means that there are $\kappa_1 = 20$ working days in this month.

then, the aggregation matrix, $\mathbf{C_2}$, is:

$$\{\mathbf{C_2}\}_{i,j} := \kappa_i^{-1}\mathbb{1}_{(f_{i-1}\leq j\leq f_i-1)}; \quad i=1,\cdots,N \qquad \text{and} \qquad j=1,\cdots,c. \tag{1.4}$$

where $\mathbb{1}_{(.)}$ is the indicator function.

The fourth ingredient depends on the practitioner's statistical understanding and assumptions of the series. CL presents the best linear unbiased estimate (BLUE) for interpolated as well as extrapolated HF observations of the LF variable based on the LF observations, the HF indicator variable(s), and the static relation model proposed (eq. 1.5). Poissonnier (2018) points out that high autocorrelation of the residuals limits the impact of the unexplained component on the HF profile. Therefore, they introduce a dynamic relationship model (eq. 1.6) in the CL framework and is assumed to be superior, in some circumstances[2], to the traditional static relationship model of Chow and Lin (1971).

The static model is defined as:

$$X_t = \sum_{j=1}^{k} \varphi_j s_{t,j} + e_t \tag{1.5}$$

The dynamic model is defined as:

$$X_t = \sum_{i=1}^{p} \rho_i X_{t-i} + \sum_{j=1}^{k} \varphi_j s_{t,j} + e_t \tag{1.6}$$

where $X_t$ is the HF observation at time $t$, $s_{t,j}$ is the indicator variable, $j$, at time $t$, and $k$ is the number of indicator variables. The stochastic disturbance terms in either

[2]For example, when the HF observations exhibit a temporal time dependency feature.

models[3] may follow:

- An AR(1) process $e_t = \mu_{CL} e_{t-1} + \omega_t$ (Chow and Lin, 1971).
- A Random walk (RW) process $e_t = e_{t-1} + \omega_t$ (Fernández, 1981).
- An ARIMA(1,1,0) process $\Delta e_t = \mu_{litt} \Delta e_{t-1} + \omega_t$ (Litterman, 1983).

where $\omega_t\}_{t=1}^{\infty}$ is a series of white noise with mean zero and variance $\sigma^2$.

Hence, we denote the six combination of models studied in this paper as follows:

1. CLCL model (Chow and Lin, 1971):

$$X_t = \sum_{j=1}^{k} \varphi_j s_{t,j} + e_t$$

$$e_t = \mu e_{t-1} + \omega_t \;;\;\; \omega_t \sim i.i.d.\ N(0, \sigma^2)$$

2. CLL model (Chow and Lin, 1971; Litterman, 1983)

$$X_t = \sum_{j=1}^{k} \varphi_j s_{t,j} + e_t$$

$$\Delta e_t = \mu \Delta e_{t-1} + \omega_t \;;\;\; \omega_t \sim i.i.d.\ N(0, \sigma^2)$$

3. CLF model (Chow and Lin, 1971; Fernández, 1981)

$$X_t = \sum_{j=1}^{k} \varphi_j s_{t,j} + e_t$$

[3] We intentionally do not refer to these terms as error terms of the model in order to distinguish between them and the error that we will define later in the sequel.

$$e_t = e_{t-1} + \omega_t \;;\quad \omega_t \sim i.i.d.\ N(0, \sigma^2)$$

4. PCL model (Chow and Lin, 1971; Poissonnier, 2018)

$$X_t = \rho X_{t-i} + \sum_{j=1}^{k} \varphi_j s_{t,j} + e_t$$

$$e_t = \mu e_{t-1} + \omega_t \;;\quad \omega_t \sim i.i.d.\ N(0, \sigma^2)$$

5. PL model (Litterman, 1983; Poissonnier, 2018)

$$X_t = \rho X_{t-i} + \sum_{j=1}^{k} \varphi_j s_{t,j} + e_t$$

$$\Delta e_t = \mu \Delta e_{t-1} + \omega_t \;;\quad \omega_t \sim i.i.d.\ N(0, \sigma^2)$$

6. PF model (Fernández, 1981; Poissonnier, 2018)

$$X_t = \rho X_{t-i} + \sum_{j=1}^{k} \varphi_j s_{t,j} + e_t$$

$$e_t = e_{t-1} + \omega_t \;;\quad \omega_t \sim i.i.d.\ N(0, \sigma^2)$$

where $|\mu| < 1$ and $|\rho| < 1$.

In what follows, we demonstrate how the HF estimates produced by the above models can be expressed as a linear transformation of the LF observations and the HF indicator series. In particular, Proposition 1.2.1 shows that the BLUE estimates of the unknown HF observations for models CLCL, CLL, and CF can be written in the form of a slope matrix pre-multiplied by the column vector of the observed LF data points.

**Proposition 1.2.1** *Given the LF vector* $\mathbf{x}_{lf}$*, the BLUE (interpolated and extrapolated) estimates of the unobserved HF vector,* $\underset{(m\times 1)}{\mathbf{z}}$*, of the static model (1.5) defined in matrix form,* $\mathbf{z} := \mathbf{S}\varphi + \mathbf{e}$*, can be found using the following linear transformation*

$$\hat{\mathbf{z}} = f(\mathbf{x}_{lf}) = \underset{(m\times N)}{\mathbf{A}}\underset{(N\times 1)}{\mathbf{x}_{lf}} \tag{1.7}$$

- *In the case of the CLCL model,* $\mathbf{A} := \mathbf{A}_{CL}$

$$\underset{(m\times N)}{\mathbf{A}_{CL}} = \underset{(m\times N)}{\mathbf{A}_{CL,1}} + \underset{(m\times N)}{\mathbf{A}_{CL,2}}$$

*where*

$$\mathbf{A}_{CL,1} := \mathbf{\Omega}\phi_{TCL}$$

$$\mathbf{A}_{CL,2} := \left(\mathbf{I}_m - \mathbf{A}_{CL,1}\mathbf{C}_z\right)\mathbf{S}\delta_{TCL}$$

$$\underset{(k\times N)}{\delta_{TCL}} := \left(\mathbf{S}^{'}\phi_{TCL}\mathbf{C}_z\mathbf{S}\right)^{-1}\mathbf{S}^{'}\phi_{TCL}$$

$$\underset{(m\times N)}{\phi_{TCL}} := \mathbf{C}_z^{'}\left(\mathbf{C}_z\mathbf{\Omega}\mathbf{C}_z^{'}\right)^{-1}$$

$$\underset{(N\times m)}{\mathbf{C}_z} := \underset{(N\times c)}{\mathbf{C}}\underset{(\mathbf{C}\times m)}{\mathbf{F}}$$

*Where* $\mathbf{F}$ *is an identity matrix,* $\mathbf{C}$ *is the aggregation matrix (see eq. 1.2 and 1.4), and* $\underset{(m\times m)}{\mathbf{\Omega}}$ *is the variance-covariance matrix of the stochastic disturbance* *terms* $\underset{(m\times 1)}{\mathbf{e}}$:

$$\underset{(m\times m)}{\boldsymbol{\Omega}} := \frac{\sigma^2}{1-\mu^2}\begin{pmatrix} 1 & \mu & \cdots & \mu^{m-1} \\ \mu & 1 & \cdots & \mu^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu^{m-1} & \mu^{m-2} & \cdots & 1 \end{pmatrix}$$

- *In the case of the CLL model,* $\mathbf{A} := \mathbf{A}_{litt}$

$$\underset{(m\times N)}{\mathbf{A}_{litt}} = \underset{(m\times N)}{\mathbf{A}_{litt,1}} + \underset{(m\times N)}{\mathbf{A}_{litt,2}}$$

*where*

$$\mathbf{A}_{litt,1} := \mathbf{D}^{-1}\boldsymbol{\Omega}\phi_{litt}$$

$$\mathbf{A}_{litt,2} := \left(\mathbf{I}_m - \mathbf{A}_{litt,1}\mathbf{C}_z\right)\mathbf{S}\delta_{litt}$$

$$\underset{(m\times N)}{\phi_{litt}} := \mathbf{D}^{'-1}\mathbf{C}_z^{'}\left(\mathbf{C}_z\left(\mathbf{D}^{'}\boldsymbol{\Omega}^{-1}\mathbf{D}\right)^{-1}\mathbf{C}_z^{'}\right)^{-1}$$

$$\underset{(k\times N)}{\delta_{litt}} := \left(\mathbf{S}^{'}\mathbf{D}^{'}\phi_{litt}\mathbf{C}_z\mathbf{S}\right)^{-1}\mathbf{S}^{'}\mathbf{D}^{'}\phi_{litt}$$

*Where* $\mathbf{C}$ *is the aggregation matrix (see eq. 1.2 and 1.4),* $\boldsymbol{\Omega}$ *is the variance-covariance matrix of* $\Delta e_t$ *defined as above, and* $\mathbf{D}$ *is difference matrix:*

$$\underset{(m\times m)}{\mathbf{D}} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \ddots & 0 \\ 0 & \cdots & \ddots & 1 \end{pmatrix}$$

- *In the case of the CLF model:*

$\mathbf{A} := \mathbf{A}_{Fer}$ *which is a special case of* $\mathbf{A}_{litt}$ *where* $\mathbf{\Omega} = \sigma^2 \mathbf{I}_m$.

Observe that the slope matrix, $\mathbf{A}$, defined in equation (1.7) for each of the proposed methods, is invariant to the variance of stochastic disturbance terms, $\sigma^2$. Furthermore, equation (1.7) presents a direct way of computing $\hat{\mathbf{z}}$ using the CL framework.

For the dynamic models, Poissonnier (2018) use partial recursive substitutions to obtain the last recorded LF observation and then, under some assumptions, formulate a constrained optimization problem to find these estimates in a closed-form expression. In particular, as illustrated in Lemma 1.2.1.1 below, the slope matrix, $\mathbf{A}_{pois}$, and the transition vector, $\mathbf{b}_{pois}$, of this linear temporal disaggregation defined in equation (1.8) are a function of Model (1.6) parameters, the underlying structure of the stochastic disturbance terms, and the HF indicator series.

**Lemma 1.2.1.1** *Given the LF vector,* $\mathbf{x}_{lf}$*, to generate interpolated estimates of the unobserved HF vector,* $\underset{(\mathbf{C}\times 1)}{\mathbf{x}}$*, from Model (1.6), Poissonnier (2018) suggests the following linear transformation*

$$\hat{\mathbf{x}} = f(\mathbf{x}_{lf}) = \mathbf{A}_{pois}\mathbf{x}_{lf} + \mathbf{b}_{pois} \tag{1.8}$$

- *In the case of the PCL model:*

$$\underset{(\mathbf{C}\times N)}{\mathbf{A}_{pois,1}} = \mathbf{\Omega}_{\mathbf{x}}\mathbf{C}^{'}(\mathbf{C}\mathbf{\Omega}_{\mathbf{x}}\mathbf{C}^{'})^{-1}$$

$$\underset{(\mathbf{C}\times 1)}{\mathbf{b}_{pois,1}} = (\mathbf{I}_c - \mathbf{A}_{pois}\mathbf{C})(\mathbf{I}_c - \mathbf{M}_1)^{-1}(\mathbf{M}_2\mathbf{x}^{init} + \mathbf{S}_{inter}\varphi)$$

*where*

$$\mathbf{\Omega_x} := (\mathbf{I}_c - \mathbf{M}_1)^{-1}\mathbf{F}\mathbf{\Omega}\mathbf{F}^{'}(\mathbf{I}_c - \mathbf{M}_1^{'})^{-1}$$

$$\underset{(\mathbf{C}\times c)}{\mathbf{M}_1} := \begin{bmatrix} 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \rho_1 & 0 & \ddots & \ddots & \ddots & \vdots \\ \rho_2 & \rho_1 & \ddots & \ddots & \ddots & \vdots \\ \rho_3 & \rho_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \rho_p & \cdots & \rho_1 & 0 \end{bmatrix} \quad \text{and} \quad \underset{(\mathbf{C}\times p)}{\mathbf{M}_2} := \begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_p \\ 0 & \rho_1 & \cdots & \rho_{p-1} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_1 \\ \vdots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

$\underset{(m\times m)}{\mathbf{\Omega}}$ *is the variance-covariance matrix of* $\underset{(m\times 1)}{\mathbf{e}}$ *as defined in Proposition 1.2.1,* $\mathbf{F}$ *is an identity matrix,* $\mathbf{C}$ *is the aggregation matrix (see (1.2)),* $\mathbf{S}_{inter} := \{\mathbf{S}_i\}_{i=1}^{c}$ *where inter stands for interpolation, and* $\mathbf{x}^{init}$ *is the initial value of* $\mathbf{x}$

- *In the case of the PL model:*

$$\mathbf{A}_{pois,2} := \mathbf{\Omega_{x,2}}\mathbf{C}^{'}\mathbf{\Omega_{x,1}}^{-1} \qquad \text{and} \qquad \mathbf{b}_{pois,2} = \lambda_{pois}\left(\mathbf{M}_2\mathbf{x}^{init} + \mathbf{F}\mathbf{S}_{inter}\varphi\right)$$

*where*

$$\lambda_{pois} := (\mathbf{I}_c - \mathbf{M}_1)^{-1} - \mathbf{\Omega_{x,2}}\mathbf{C}^{'}\mathbf{\Omega}_{x,1}^{-1}\mathbf{C}(\mathbf{I}_c - \mathbf{M}_1)^{-1}$$

$$\underset{(c\times c)}{\mathbf{\Omega_{x,1}}} = \mathbf{C}\mathbf{\Omega_{x,2}}\mathbf{C}^{'} \;, \quad \underset{(c\times c)}{\mathbf{\Omega_{x,2}}} := (\mathbf{I}_c - \mathbf{M}_1)^{-1}\left(\mathbf{D}_1^{'}\left(\mathbf{F}\mathbf{\Omega}\mathbf{F}^{'}\right)^{-1}\mathbf{D}_1\right)^{-1}(\mathbf{I}_c - \mathbf{M}_1^{'})^{-1}$$

$\mathbf{D}_1$ *is the* $(c\times c)$ *difference matrix and* $\mathbf{\Omega}$ *is the variance covariance matrix* $\Delta e$ *as defined in Proposition 1.2.1,.* $\mathbf{M}_1$ *and* $\mathbf{M}_2$ *are defined as in the PCL case.*

- *In the case of the PF model:*

*The $(c \times N)$ matrix $\mathbf{A}_{pois,3}$ and the $c$ vector $\mathbf{b}_{pois,3}$ are exactly as $\mathbf{A}_{pois,2}$ and $\mathbf{b}_{pois,2}$, respectively. The only difference is that, in this case, $\boldsymbol{\Omega} = \sigma^2 \mathbf{I}_m$.*

Unlike the static model, the dynamic models were proposed and demonstrated for interpolation only and not extrapolation. Hence, the application of the linear transformation stated in Lemma 1.2.1.1 generates only interpolated HF data estimates of the LF variable. Proposition 1.2.2 generalizes this Lemma to produce extrapolated and interpolated HF data estimates of the LF variable.

**Proposition 1.2.2** *Given the LF vector, $\mathbf{x}_{lf}$, to generate the best linear unbiased interpolated as well as extrapolated HF estimates from Model (1.6), the linear transformation becomes:*

$$\hat{\mathbf{z}} = f(\mathbf{x}_{lf}) = \mathbf{A}_{epois}\mathbf{x}_{lf} + \mathbf{b}_{epois} \tag{1.9}$$

*where the subscript epois stands for extended Poissonnier (2018).*

- *In the case of the PCL model:*

$$\underset{(m \times N)}{\mathbf{A}_{epois,1}} = \boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}} \mathbf{C}_z' {\boldsymbol{\Omega}_{\mathbf{x},\mathbf{1}}}^{-1}$$

$$\underset{(m \times 1)}{\mathbf{b}_{epois,1}} = (\mathbf{I}_m - \mathbf{A}_{epois,1}\mathbf{C}_z)\,(\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}(\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi)$$

*where*

$$\boldsymbol{\Omega}_{\mathbf{x},\mathbf{1}} = \mathbf{C}_z \boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}} \mathbf{C}_z' \ , \quad \boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}} := (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}\boldsymbol{\Omega}(\mathbf{I}_m - \mathbf{M}_{1,z}')^{-1}$$

*$\mathbf{C}_z$ is the $N \times m$ matrix, defined exactly in the same manner as the $N \times c$ matrix*

$\mathbf{C}$ *defined in either set-ups.* $\underset{(m\times m)}{\mathbf{\Omega}}$ *is the variance-covariance matrix of* $\underset{(m\times 1)}{\mathbf{e}}$.

$\underset{(m\times m)}{\mathbf{M}_{1,z}}$ *and* $\underset{(m\times p)}{\mathbf{M}_{2,z}}$ *are defined in a similar structure to* $\mathbf{M}_1$ *and* $\mathbf{M}_2$ *respectively.*

- *In the case of the PL model:*

  $\mathbf{A}_{epois,2}$ *and* $\mathbf{b}_{epois,2}$ *have the same structure as* $\mathbf{A}_{epois,1}$ *and* $\mathbf{b}_{epois,1}$, *respectively, with variance-covariance matrix* $\mathbf{D}^{-1}\mathbf{\Omega}\mathbf{D}'^{-1}$ *instead of* $\mathbf{\Omega}$ *in* $\mathbf{\Omega}_{\mathbf{x},\mathbf{2}}$.

- *In the case of the PF model:*

  $\mathbf{A}_{epois,3}$ *and* $\mathbf{b}_{epois,3}$ *are exactly as* $\mathbf{A}_{epois,2}$ *and* $\mathbf{b}_{epois,2}$ *respectively, with* $\mathbf{\Omega} = \sigma^2\mathbf{I}_m$.

## 1.3 Theoretical Properties of the Linear Transformations

Firstly, observe that both linear transformations depicted in equations (1.7) and (1.9) generate interpolated as well as extrapolated estimates, $\hat{\mathbf{z}}$, of the unobserved HF variable, $\mathbf{z}$. Note that the slope coefficient or/and the transition vector in these transformations depend on unknown coefficients to be estimated. For example, $\mathbf{A}_{CL}$ is a function of the unknown autoregressive parameter $\mu$. We denote by $\widetilde{\mathbf{z}}$ the feasible estimate of the unobserved HF vector $\mathbf{z}$. One needs to distinguish between hypothetical errors, feasible errors, and residuals. In this framework, a hypothetical error term is a difference between the unobserved HF variable $\mathbf{z}$ and its theoretical estimate $\hat{\mathbf{z}}$, i.e. $u_z := \mathbf{z} - \hat{\mathbf{z}}$. A feasible error term is the difference between the unobserved HF variable $\mathbf{z}$ and its feasible estimate $\widetilde{\mathbf{z}}$ i.e. $u_{\widetilde{z}} := \mathbf{z} - \widetilde{\mathbf{z}}$. Residuals, by definition, are the difference between an actual and estimated value. In this context, actual values (observed values) only exist at a LF (i.e. $\mathbf{x}_{lf}$). Hence, to avoid confusion, we restrict the definition of residuals, in this framework, to the LF timescale.[4]

Secondly, the slope matrix $\mathbf{A}$ and $\mathbf{A}_{epois}$ in equations (1.7) and (1.9) are independent of the variance scale of the stochastic disturbance terms, $\sigma^2$. Although this seems like an encouraging feature in an econometric method – i.e. no need to estimate the variance scale parameter–, this has a detrimental effect on the mean squared hypothetical error. The theoretical estimate $\hat{\mathbf{z}}$ may not capture (and hence mimic) the variation of the actual unobserved variable $\mathbf{z}$. For a given two sets of HF data points, $\mathbf{z}$, generated from the same pre-determined model with different variance scales each, the theoretical estimates of the data with a larger variance scale have a higher average squared hypothetical error. Hence, the CL method is less accurate for data with high variances, such as disaggregating monthly historical stock returns to a higher frequency.

Thirdly, the estimated HF data points using the models discussed earlier are the sum of two signals. Precisely, the estimation represented by equation (1.7) of the static models, can be rewritten as $\hat{\mathbf{z}} = \mathbf{A}_{\pi,1}\mathbf{x}_{lf} + \mathbf{A}_{\pi,2}\mathbf{x}_{lf}$ where $\pi$ refers to the disaggregation method: CLCL, CLL, or CLF. $\mathbf{A}_{\pi,1}\mathbf{x}_{lf}$ and $\mathbf{A}_{\pi,2}\mathbf{x}_{lf}$ represent the trajectory and perturbation signal, respectively. Model parameters and the LF observations determine the trajectory signal. It passes through the LF data points similar to the naive interpolation of connecting LF observations. Because of the model parameters involved, these segments are bent. In the out-of-sample period, where there are no observed LF data points, the path of the trajectory signal will rely solely on the model's parameters. Hence, the estimation accuracy of the in-

[4] In Poissonnier (2018)'s paper, the residual is similar to the feasible error term in this paper.

sample dataset outperforms the relative out-of-sample estimation accuracy. The perturbation signal, independent of the trajectory, represents the fluctuations driven by the model's parameters and the observed series of the LF variable and the HF indicator variable(s). Therefore, if the indicator variable(s) variance is high, the perturbation signal fluctuations tend to be substantial.

The above discussions provide an insightful understanding of the disaggregation characteristics. However, in reality, a practitioner is often unaware of the defined model and parameters of the data generating process (DGP). Therefore, it is reasonable to ask whether assuming a different model from the DGP might lead to a substantial difference and if there is a particular model that out-stands its peers regardless of the DGP. To answer these questions, we first outline the estimation process in each of the mentioned models. Second, we generate artificial series from each of the discussed models with various parameters' values as DGPs to evaluate the in- and out-of-sample mean squared feasible errors (MSFE) in each case for each model.

## 1.4 Estimation

Once the DGP model is hypothesised, unknown parameters need to be estimated to find a feasible estimate of the HF points. Hence, one needs to distinguish between parameters: Parameters Estimated using data available (PEDA) and Parameters that require estimation (PRE). For example, using the CLCL model, Chow and Lin (1971) considered both $\varphi_j$'s in equation (1.5) and the AR(1) coefficient, $\mu$, to be unknown parameters that need to be estimated. However, based on Proposition 1.2.1, only

$\mu$ needs to be estimated using a likelihood function. In such a setting, the $\varphi_j$'s are PEDA, and $\mu$ is a PRE.

In the six mentioned models, most of the literature assumes that the error term, $\omega$, in the model of the stochastic disturbance term, $e_{hf}$ or $\Delta e_{hf}$, is white noise. However, many authors assume normality with zero mean and unknown variance, $\sigma^2$, to formulate a likelihood function. In particular, $\mathbf{e}_{hf} \sim N(0, \mathbf{\Omega})$ where $\mathbf{e}_{hf}$ represents the vector of HF error terms in the CLCL or PCL models, or $\mathbf{De}_{hf} \sim N(0, \mathbf{\Omega})$ in the other models. Therefore, the LF vector of error terms, $\mathbf{e}_{lf}$, has a similar structure. Having said this, Bournay and Laroque (1979) and Sax and Steiner (2013), defined the log-likelihood function as:

$$\mathscr{L}(\text{PRE}) := -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\mathbf{C_z}\mathbf{\Omega}^*\mathbf{C}_{\mathbf{z}}^{'}\right| - \frac{1}{2}\hat{\mathbf{e}}_{lf}^{'}(\mathbf{C}_z\mathbf{\Omega}^*\mathbf{C}_z^{'})^{-1}\hat{\mathbf{e}}_{lf} \tag{1.10}$$

where $\mathbf{\Omega}^* := \mathbf{\Omega}$ in the case of CLCL and PCL, or $\mathbf{\Omega}^* := \left(\mathbf{D}^{'}(\mathbf{\Omega})^{-1}\mathbf{D}\right)^{-1}$ in the other cases.

However, Poissonnier (2018), instead, maximizes the log-likelihood function of the HF error terms directly for their proposed dynamic model:

$$\mathscr{L}(\text{PRE}) := -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Omega}^*| - \frac{1}{2}\hat{\mathbf{e}}_{hf}^{'}(\mathbf{\Omega}^*)^{-1}\hat{\mathbf{e}}_{hf} \tag{1.11}$$

This likelihood function, in fact, can be used for all the models in this paper including the static ones and it can be rewritten as:

$$\mathscr{L}(\text{PRE}) := -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Omega}^*| - \frac{1}{2}\hat{\mathbf{e}}_{lf}^{'}(\mathbf{C}_z\mathbf{\Omega}^*\mathbf{C}_z^{'})^{-1}\hat{\mathbf{e}}_{lf} \tag{1.12}$$

Hint: $\hat{\mathbf{e}}_{hf} = \mathbf{\Omega}^{*}\mathbf{C}_{z}^{'}(\mathbf{C}_{z}\mathbf{\Omega}^{*}\mathbf{C}_{z}^{'})^{-1}\hat{\mathbf{e}}_{lf}$ and equivalently $\hat{\mathbf{e}}_{lf} = \mathbf{C}\hat{\mathbf{e}}_{hf}$.

Note that $\mathbf{\Omega}$ and $\hat{\mathbf{e}}_{hf}$ are a function of the unknown parameters. In particular, for the static models, $\hat{\mathbf{e}}_{hf} = (\mathbf{A} - \mathbf{S}\delta)x_{lf}$ where the expressions of $\mathbf{A}$ and $\delta$ depend on the model as outlined in Proposition 1.2.1 (see Appendix 1.A.1). Similarly, for the dynamic models, $\hat{\mathbf{e}}_{hf} = (I_m - \mathbf{M}_{1,z})(\mathbf{A}\mathbf{x}_{lf} + \mathbf{b}) - \mathbf{M}_{2,z}\mathbf{z}^{init} - \mathbf{S}\varphi$ where $\mathbf{M}_{1,z}$ and $\mathbf{M}_{2,z}$ are defined in Lemma 1.2.1.1, while the expressions of $\mathbf{A}$ and $\mathbf{b}$ depend on the specified model as outlined in Proposition 1.2.2 (see Appendix 1.A.3).

The log-likelihood can be used to estimate the PREs in the hypothesised model using maximum likelihood estimation. However, not all the PREs are required to find feasible HF estimates, $\widetilde{\mathbf{z}}$. For example, in the CLCL model, there are two PREs, the autoregressive parameter, $\mu$, and the variance $\sigma^2$. The only parameter that needs to be estimated using the likelihood function is $\mu$ since the matrix $\mathbf{A}_{CL}$ in equation (1.5) is a function of it, $\mathbf{A}_{CL} = \mathbf{A}_{CL}(\mu)$. Thus, the feasible HF estimate $\widetilde{\mathbf{z}}$ is computed by pre-multiplying the LF vector $\mathbf{x}_{lf}$ by the estimated transformation matrix $\hat{\mathbf{A}}_{CL} = \mathbf{A}_{CL}(\hat{\mu})$ as suggested by Proposition 1.2.1. However, in the case of the CLF model, the slope transformation matrix does not depend on the PREs, so there is no need to estimate the likelihood function.

## 1.5 Simulation Study

We implement a Monte Carlo simulation to examine the performance of the models covered in this paper. The simulation compares the models' estimation of the "unobserved" HF daily series using the observed LF monthly series. To this end, firstly, we choose two consecutive calendar years to determine the number of working days in each calendar month; this leads to around 500 HF observations. Secondly, in every simulation, we use the same simulated daily indicator series from an AR(1) process for all the DGPs, for comparison purposes. Thirdly, we generate the error term, $\omega_t$, from a standard normal distribution with zero mean and variance of $\sigma^2$. Fourthly, the stochastic disturbance term, $e_t$, and the HF series, $\mathbf{x}$, are generated according to the specified model and parameters ($\mu$, $\rho$, $\varphi$) as illustrated in Table (1.1). Fifthly, the LF series, $\mathbf{x}_{lf} := \mathbf{C}\mathbf{x}$, are computed using the specified aggregation matrix (eq. 1.2 or 1.4). Finally, using the LF series and the indicator series, we estimate the HF series using each of the six models as illustrated in sections 1.2 and 1.4. We repeat the process 1000 times to obtain 1000 artificial samples for each DGP. We use the six models for each simulated sample to estimate the HF series given the simulated LF series and the HF indicator. We estimate the models with the following constraints on the parameters: $0 < \mu < 1$, $0 < \rho < 1$, and $0 < \varphi < 1$.

Table 1.1: DGP Parameters

| | $\sigma^2$ | $\mu$ | $\varphi$ | $\rho$ | **Total # DGPs** |
|---|---|---|---|---|---|
| **CLCL** | 0.1, 1, 5, or 10 | 0.2 or 0.8 | 0.2 or 0.8 | | 16 |
| **CLL** | 0.1, 1, 5, or 10 | 0.2 or 0.8 | 0.2 or 0.8 | | 16 |
| **CLF** | 0.1, 1, 5, or 10 | | 0.2 or 0.8 | | 8 |
| **PCL** | 0.1, 1, 5, or 10 | 0.2 or 0.8 | 0.2 or 0.8 | 0.2 or 0.8 | 32 |
| **PL** | 0.1, 1, 5, or 10 | 0.2 or 0.8 | 0.2 or 0.8 | 0.2 or 0.8 | 32 |
| **PF** | 0.1, 1, 5, or 10 | | 0.2 or 0.8 | 0.2 or 0.8 | 16 |

Many studies are focused on evaluating and comparing models in estimation and forecasting; see, for example, Poon and Granger (2003) for an extensive review. Most apply a loss function, where model-based predictions of the series of interest are compared to its estimates. We compare the estimated HF series with the true simulated one to evaluate the models. We use the Mean Squared Error (MSE) as a model evaluation method (Andersen, Bollerslev, and Lange, 1999):

$$MSE = \frac{1}{N}\Sigma_{t=1}^{N}[Xt - \hat{X}_t]^2$$

where $N$ is the total number of observations in a given series.

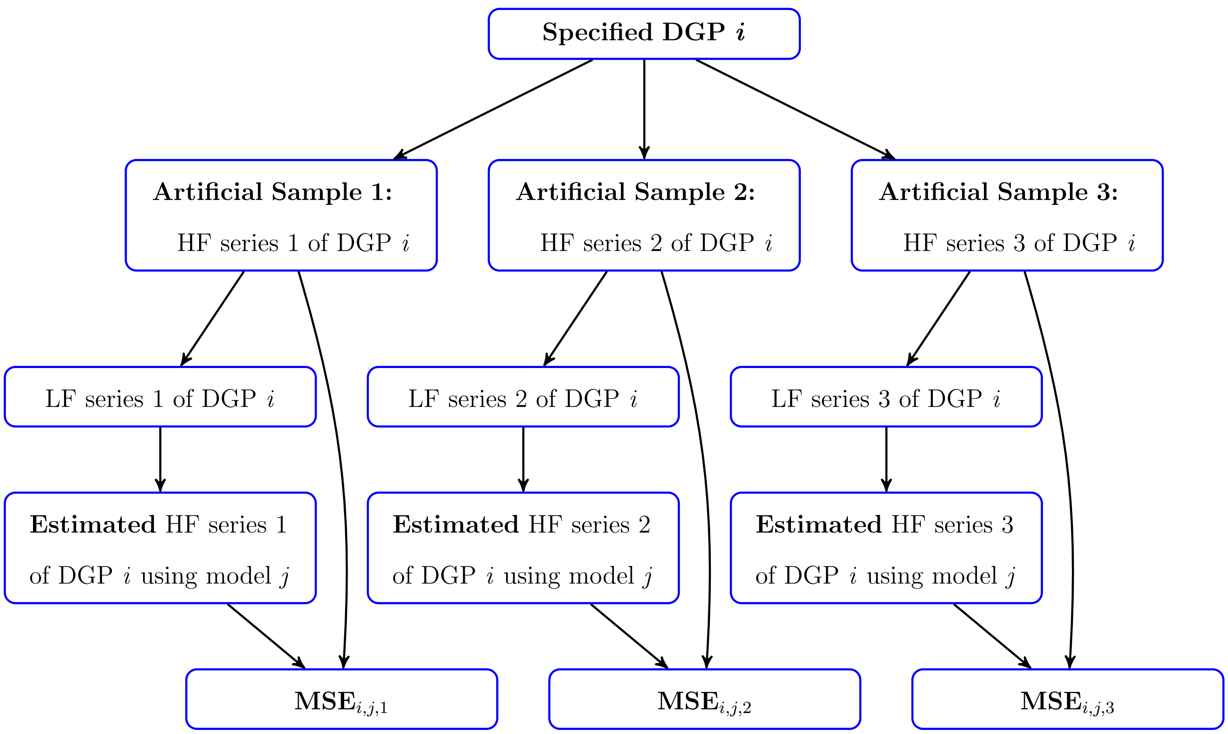
After computing the MSE for each simulation, we report for every model of estimation the average of the MSEs using the 1000 simulations of each DGP. As illustrated in figure 1.1, we compute the $MSE_{i,j,sim}$ for simulations $sim = 1, .., 1000$, then we get the average mean squared error (AMSE) over the simulations:

$$AMSE_{i,j} = \frac{1}{1000}\Sigma_{sim=1}^{1000} MSE_{i,j,sim}$$

where $i$ and $j$ refer to the DGP and model of estimation respectively. We also consider the mean absolute error as a robustness check and found no significant differences in the findings. Hence, we only report the results using the MSE as a loss function for conciseness.

Table 1.1 summarises the DGPs (models and parameters) analysed in the simulation study. We include a range of magnitudes for each parameter to unveil their consequences on the estimated data points, ceteris paribus. In particular, it helps the practitioners understand what features they need to check before choosing the "right" model for a particular exercise. For the AR parameters, $\mu$ and $\rho$, we consider low (0.2) and high (0.8) values. The range of $\rho$ allows us to study whether the dynamic models are better than their static counterparts if an autoregressive relationship exists in the HF series. At the same time, the range of $\mu$ helps us examine the influence of its magnitude on the overall AMSE and the choice of the "best" model. We also consider a high correlation, $\varphi = 0.8$, as well as a low correlation, $\varphi = 0.2$, between the "unobserved" HF series and the indicator variable. We can thus examine the importance of choosing the "right" indicator variable. Finally, we consider different magnitudes of the variance, $\sigma^2$, to determine if different models are preferred based on the magnitude of variation of the series of interest. Finally, we repeat the study twice, once using the aggregation matrix $\mathbf{C}_1$ (eq. 1.2) and another using $\mathbf{C}_2$ (eq. 1.4).

Figure 1.1: Descriptive Diagram of the Monte Carlo Study

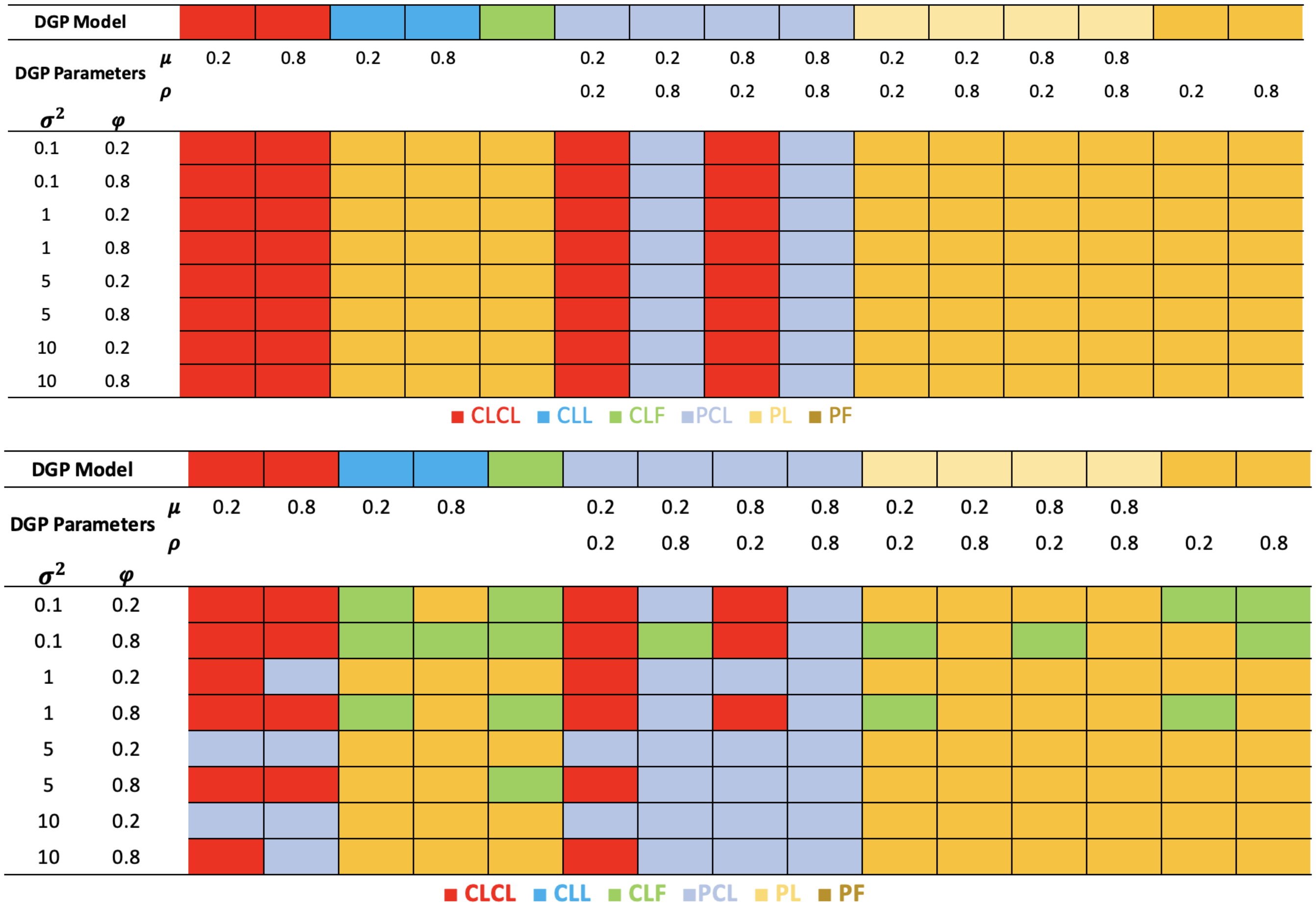



For each DGP, $i$, we generate 1000 artificial samples. Each artificial sample consists of a HF and its associated LF series. We use the LF series along with the indicator variable to estimate the HF series using each of the six models.

The Monte Carlo study reveals essential features in the performance of the models. We divide the analysis of the findings into two. In the first part of the analysis, we comment on the influence of the chosen DGP model and parameters on the AMSE. The second part examines the best-performing model under various DGPs.

We report the AMSE and its standard deviation using the 1000 simulations from each model, $j$, in each DGP, $i$, in Appendix 1.B for the interpolation as well as extrapolation sample. Our findings can be summarized as follows. Firstly, while the AMSE of extrapolation is higher than that of interpolation, there is no significant difference in the ranking of models between interpolation and extrapolation. Secondly, we note that the performance of the models is sensitive to some parameters of the underlying DGP and the relationship between the HF observations and its corresponding LF series (i.e. the aggregation matrix). Thirdly, the AMSE is overall lower when the DGP is CLCL or PCL. In other words, temporal disaggregation is generally more accurate for series where the $e_t$ is an AR(1) rather than ARIMA(1,1,0) or RW. Fourthly, better estimates can be found for series with a defined aggregation matrix as in eq. (1.4) (i.e. when the LF observation represents the average of the HF values within the period) than eq. (1.2) (i.e. when the LF observation is one particular observed datapoint of the HF values). Fifthly, temporal disaggregation is more accurate when the defined DGP is static compared to dynamic. Finally, we do not find a significant impact of the relevancy of the indicator variable on the AMSE.

Figure 1.2: Best Estimating Model in each DGP of the Global Monte Carlo Study




The tables represent the best estimating model in each DGP of the GMC. The upper table represents the results for aggregation matrix one, and the lower represents the results for aggregation matrix two. The colour of the cells in the True model row represents the model used to generate the true HF series. The model and parameters determine the DGP *i*s. Hence, each cell in the inner triangle represents one case of the 120 DGPs considered in the MC. The colour of the cell represents the estimating model with the lowest average MSE over the 1000 simulations for a particular DGP.

Figure 1.2 presents the results of the best model with the lowest AMSE for each DGP considered. In the case of the first aggregation matrix (eq. 1.2), the findings are consistent regardless of the magnitudes of $\sigma^2$ and $\varphi$. In particular, the CLCL model is always dominant when the DGP is CLCL or PCL with low $\rho$. In contrast, the PCL model is only dominant (by a small margin) when the DGP is PCL with high $\rho$. In other words, the CLCL model is preferred when the stochastic disturbance terms, $e_t$, follow an AR(1) process, except when the model is dynamic with a high autoregressive parameter, $\rho$. The PF model is dominant in all cases where the stochastic disturbance terms, $e_t$, follow an ARIMA(1,1,0) or RW. In particular, the RW specification for the $e_t$ is superior to the ARIMA(1,1,0) in static and dynamic estimated models under all DGPs.

There are a couple of instances where the results are different when the aggregation matrix is $\mathbf{C}_2$ instead of $\mathbf{C}_1$. For example, the PCL model moderately dominates, with few exceptions, when the DGP is CLCL or PCL with high variance, $\sigma^2$. It means that under the second aggregation matrix, the PCL model performs better when the stochastic disturbance terms follow an AR process with high variance. Further, the CLF ranks moderately better in a few instances, especially when the DGP is CLL, CLF, or PL (with low $\rho$) with high $\varphi$ and low variance, $\sigma^2$. That is to say, using the second aggregation matrix, the CLF may outperform under three conditions: a relevant indicator is added, the stochastic disturbance terms follow an ARIMA(1,1,0) or RW process with low variance, and the autoregressive parameter is low under the dynamic model.

Therefore, based on the findings of the Monte Carlo, our recommendation is as follows. When the stochastic disturbance terms are assumed to follow an AR(1) process, the static model (CLCL) is the best model to use unless the AR parameter of the unobserved HF values is assumed to be high. If the AR parameter is low, the relationship seems to be captured by the AR(1) process of the stochastic disturbance terms. On the other hand, when the stochastic disturbance terms follow an ARIMA(1,1,0) or RW, the dynamic model by Poissonnier (2018) would be more suitable since it depicts the autocorrelation component in the HF series. Specifically, the RW assumption is recommended in this case because although it yields only moderate improvement over its counterparts, it also has fewer parameters to estimate. Further, as noted by Poissonnier (2018), using the RW specification minimizes the residual variations from period to period and hence reduces the impact of the unexplained component of the model on the HF estimates.

## 1.6 Conclusion

This essay revisits temporal disaggregation models under the Chow-Lin framework: the static (Chow and Lin, 1971) and dynamic (Poissonnier, 2018) models, each with three different variants of the stochastic disturbance terms process. We show that the HF estimates found by these models are a linear transformation of their corresponding LF series. We also illustrate the necessary modification for a non-constant frequency ratio such as monthly to daily series. Further, we present the extended model of Poissonnier (2018) to estimate extrapolated and interpolated HF series.

We discuss the theoretical features of the temporal disaggregation methods. Specifically, we show that the HF estimates consist of a trajectory and a perturbation signal. We examine the assumptions and parameters that influence each signal. It gives a closer look to the practitioner on how and which assumptions and parameters drive the estimated values. We also illustrate the log-likelihood estimation and clarify which parameters need to be estimated.

Finally, we perform a global Monte Carlo study whereby we analyse the performance of the models under various combinations of different representative magnitudes of the parameters. The study examines which model provides the highest accuracy in the HF estimates under various data generating processes. We find that the preferred model generally depends on the process of the stochastic disturbance terms in the DGP rather than whether the true model is static or dynamic. The traditional static model by Chow and Lin (1971) dominates when the DGP error term follows an AR(1) process. In contrast, the dynamic model of Poissonnier (2018) with RW error term process (Fernández, 1981) is preferred when the DGP error terms follow an ARIMA(1,1,0) or RW process.

# Appendices

## 1.A Proofs

### 1.A.1 Proof of Proposition 1.2.1

Observe that, in matrix form, the static model is defined by: $\mathbf{x}_{lf}= \mathbf{C}\mathbf{x}$, and the true high-frequency observation $\mathbf{z}$ is given by the model: $\mathbf{z} = \mathbf{S}\varphi + \mathbf{e}$. $\mathbf{S}$ is the $(m \times k)$ indicator matrix that includes both interpolated data-points and extrapolated data-points (if $m > c$). $\mathbf{C}$ is the $(N \times c)$ aggregation matrix defined in either (1.2) or (1.4). $\mathbf{e}$ is a $m \times 1$ vector of stochastic disturbances terms (both interpolated and extrapolated) and is defined by $\mathbf{e} := \begin{pmatrix} e_1 & \cdots & e_m \end{pmatrix}$.

**Note**: $\mathbf{z} = \mathbf{x}$ if $m = c$, hence, when $m \neq c$, $\mathbf{x}_{lf} = \mathbf{Cx} = \mathbf{CFz}$ where $\mathbf{F}$ is a $c \times m$ non-square identity matrix defined as $\mathbf{F} = \mathbf{I}_{c\times m}$.

The proof consists of three parts for: CLCL, CLL, and CLF.

- Let the estimated high-frequency observation of $\mathbf{z}$ be denoted by $\hat{\mathbf{z}}$ where $\hat{\mathbf{z}} = \mathbf{A}\mathbf{x}_{lf} = \mathbf{ACFx} = \mathbf{ACFS}\varphi + \mathbf{ACFe}$. Define the error of the high-frequency estimation by

$$\begin{aligned} \mathbf{u}_z &= \mathbf{z} - \hat{\mathbf{z}} \\ &= \mathbf{S}\varphi - \mathbf{ACFS}\varphi + \mathbf{e} - \mathbf{ACFe} \end{aligned}$$

For $\hat{\mathbf{z}}$ to be an unbiased estimator of $\mathbf{z}$ in the case of Chow and Lin (1971), $\mathbb{E}(\mathbf{u}_z) = 0$. Hence, $\mathbf{S} = \mathbf{ACFS}$ and $var(\mathbf{u}_z) = (\mathbf{I}_m - \mathbf{ACF})\,\mathbf{\Omega}\,(\mathbf{I}_m - \mathbf{ACF})'$ where $\mathbf{\Omega}^e$ is the $m \times m$ variance covariance matrix of $\mathbf{e}$.

Our objective is to minimize the $var(\mathbf{u}_z)$ s.t. $\mathbf{S} = \mathbf{ACFS}$ hence, the Lagrangian expression is given by

$$\mathscr{L} = \frac{1}{2} tr(var(\mathbf{u}_z)) - tr(M'(\mathbf{ACFS} - \mathbf{S}))$$

where $M$ is a $m \times k$ matrix of Lagrange multipliers.

$$\frac{\partial \mathscr{L}}{\partial \mathbf{A}} = (\mathbf{ACF} - \mathbf{I}_m)\mathbf{\Omega}\mathbf{F}'\mathbf{C}' - \mathbf{M}\mathbf{S}'\mathbf{F}'\mathbf{C}' = 0$$

Hence,

$$\mathbf{A} = \mathbf{\Omega F}^{'}\mathbf{C}^{'}(\mathbf{CF\Omega F}^{'}\mathbf{C}^{'})^{-1} + \mathbf{MS}^{'}\mathbf{F}^{'}\mathbf{C}^{'}(\mathbf{CF\Omega F}^{'}\mathbf{C}^{'})^{-1} \tag{1.13}$$

since $\mathbf{S} = \mathbf{ACFS}$, then

$$\mathbf{M} = \left(\mathbf{S} - \mathbf{\Omega F}^{'}\mathbf{C}^{'}\left(\mathbf{CF\Omega F}^{'}\mathbf{C}^{'}\right)^{-1}\mathbf{CFS}\right)\left(\mathbf{S}^{'}\mathbf{F}^{'}\mathbf{C}^{'}\left(\mathbf{CF\Omega F}^{'}\mathbf{C}^{'}\right)^{-1}\mathbf{CFS}\right)^{-1}$$

Replace the matrix $M$ in equation (1.13), then denote $\phi_{TCL} = \mathbf{C}^{'}\left(\mathbf{CF\Omega F}^{'}\mathbf{C}^{'}\right)^{-1}$ and $\delta_{TCL} := \left(\mathbf{S}^{'}\mathbf{F}^{'}\phi_{TCL}\mathbf{CFS}\right)^{-1}\mathbf{S}^{'}\mathbf{F}^{'}\phi_{TCL}$ then one can show that

$$\begin{aligned}\mathbf{A} &= \mathbf{\Omega F}^{'}\phi_{TCL} + (\mathbf{I}_m - \mathbf{\Omega F}^{'}\phi_{TCL}\mathbf{CF})\mathbf{S}\delta_{TCL} \\ &= \mathbf{\Omega F}^{'}\phi_{TCL}(\mathbf{I}_N - \mathbf{CFS}\delta_{TCL}) + \mathbf{S}\delta_{TCL}\end{aligned}$$

Note that the GLS estimator of $\varphi$:

$$\begin{aligned}\hat{\varphi} &= (\mathbf{S}'\mathbf{C}'(\mathbf{C\Omega C}')^{-1}\mathbf{CS})^{-1}\mathbf{S}'\mathbf{C}(\mathbf{C\Omega C}')^{-1}\mathbf{Cx} \\ &= (\mathbf{S}'\phi_{TCL}\mathbf{CS})^{-1}\mathbf{S}'\phi_{TCL}\mathbf{x_{lf}} \\ &= \delta_{TCL}\mathbf{x_{lf}}\end{aligned}$$

Hint: $\mathbf{Cx} = (\mathbf{CS})\varphi + \mathbf{Ce}$.

Hence,

$$\hat{\mathbf{e}}_{lf} = \mathbf{x_{lf}} - \mathbf{CS}\delta_{TCL}\mathbf{x_{lf}}$$

Hint: $\hat{\mathbf{e}}_{hf} = \hat{\mathbf{z}} - \mathbf{S}\hat{\varphi}$

- Using similar notation, it is assume that the stochastic disturbances terms follow an AR(1) process or an ARIMA(1,1,0) process. Let $\mathbf{\Delta}$ be the $N \times N$ difference matrix, defined by

$$\mathbf{\Delta} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \ddots & 0 \\ 0 & \cdots & \ddots & 1 \end{pmatrix} \tag{1.14}$$

Furthermore, let $\mathbf{D}$ be also a $m \times m$ difference matrix similar in structure to $\mathbf{\Delta}$. Let $\mathbf{Q}$ be $N \times c$ matrix such that $\mathbf{QFD} := \mathbf{\Delta CF}$, then

$$\mathbf{\Delta x}_{lf} = \mathbf{\Delta Cx} = \mathbf{\Delta CFz} = \mathbf{QFDz}$$

The difference series $\mathbf{D\hat{z}}$ and $\mathbf{Dz}$ can be presented by

$$\mathbf{D\hat{z}} = \mathbf{R\Delta x}_{lf} = \mathbf{RQFDz} = \mathbf{RQF}(\mathbf{DS}\varphi + \mathbf{De}) \qquad \text{and} \qquad \mathbf{Dz} = \mathbf{DS}\varphi + \mathbf{De}$$

the error in estimation is given by

$$u_z = \mathbf{Dz} - \mathbf{D\hat{z}} = (\mathbf{I}_m - \mathbf{RQF})\mathbf{DS}\varphi + (\mathbf{I}_m - \mathbf{RQF})\mathbf{De}$$

imposing $E(u_z) = 0$ for unbiasedness of estimator, we get $\mathbf{RQFDS} = \mathbf{DS}$. Furthermore,

$$var(u_z) = (\mathbf{I}_m - \mathbf{RQF})\,\mathbf{\Omega}\left(\mathbf{I}_m - \mathbf{F}'\mathbf{Q}'\mathbf{R}'\right)$$

where here $\mathbf{\Omega} := \mathbb{E}\left(\mathbf{Dee}'\mathbf{D}\right)$, hence almost the variance covariance matrix of the $(m-1) \times 1$ vector of white noise $\mathbf{\Omega}$. However, this is not the case, this is due to the fact that $\mathbf{D}$ is square matrix with an element 1 in the upper left corner.

It is assumed that $e_0 = 0$, if the sample size is not too small, this assumption should not constitute a serious problem. Hence, in case of Fernández (1981), $\mathbf{\Omega} = \sigma^2 I_m$. In case of Litterman (1983),

$$\mathbf{\Omega} := \frac{\sigma^2}{1-\mu^2}\begin{pmatrix} 1 & \mu & \cdots & \mu^{m-1} \\ \mu & 1 & \cdots & \mu^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \mu^{m-1} & \mu^{m-2} & \cdots & 1 \end{pmatrix}$$

Our objective is to minimize the $var\,(u_z)$ s.t. $\mathbf{RQFDS} = \mathbf{DS}$ hence, the Lagrangian expression is given by

$$\mathscr{L} = \frac{1}{2} tr(var(u_z)) - tr(M^{'}(\mathbf{RQFDS} - \mathbf{DS}))$$

where $M$ is a $m \times k$ matrix of Lagrange multipliers

$$\frac{\partial \mathscr{L}}{\partial R} = (\mathbf{RQF} - \mathbf{I}_m)\mathbf{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'} - \mathbf{M}\mathbf{S}^{'}\mathbf{D}^{'}\mathbf{F}^{'}\mathbf{Q}^{'} = 0$$

Hence,

$$\mathbf{R} = \mathbf{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'}(\mathbf{QF\Omega F}^{'}\mathbf{Q}^{'})^{-1} + \mathbf{M}\mathbf{S}^{'}\mathbf{D}^{'}\mathbf{F}^{'}\mathbf{Q}^{'}(\mathbf{QF\Omega F}^{'}\mathbf{Q}^{'})^{-1} \tag{1.15}$$

replacing $\mathbf{R}$ in $\mathbf{RQFDS} = \mathbf{DS}$, we get the Lagrange matrix as

$$\mathbf{M} = (I_m - \mathbf{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'}(\mathbf{QF\Omega F}^{'}\mathbf{Q}^{'})^{-1}\mathbf{QF})\mathbf{DS}(\mathbf{S}^{'}\mathbf{D}^{'}\mathbf{F}^{'}\mathbf{Q}^{'}(\mathbf{QF\Omega F}^{'}\mathbf{Q}^{'})^{-1}\mathbf{QFDS})^{-1}$$

Replacing $M$ in equation (1.15), we get

$$\mathbf{R} = \boldsymbol{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'}\left(\mathbf{Q}\mathbf{F}\boldsymbol{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'}\right)^{-1} + \left(\mathbf{I}_m - \boldsymbol{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'}\left(\mathbf{Q}\mathbf{F}\boldsymbol{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'}\right)^{-1}\mathbf{Q}\mathbf{F}\right)\mathbf{D}\mathbf{S}$$
$$\left(\mathbf{S}^{'}\mathbf{D}^{'}\mathbf{F}^{'}\mathbf{Q}^{'}\left(\mathbf{Q}\mathbf{F}\boldsymbol{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'}\right)^{-1}\mathbf{Q}\mathbf{F}\mathbf{D}\mathbf{S}\right)^{-1}\mathbf{S}^{'}\mathbf{D}^{'}\mathbf{F}^{'}\mathbf{Q}^{'}\left(\mathbf{Q}\mathbf{F}\boldsymbol{\Omega}\mathbf{F}^{'}\mathbf{Q}^{'}\right)^{-1}$$

Let

$$\begin{aligned}\phi_1 &:= \mathbf{F}^{'}\mathbf{Q}^{'}(\mathbf{Q}\mathbf{F}\boldsymbol{\Omega}^{\omega}\mathbf{F}^{'}\mathbf{Q}^{'})^{-1}\\ &= \mathbf{D}^{'-1}\mathbf{F}^{'}\mathbf{C}^{'}\boldsymbol{\Delta}^{'}(\boldsymbol{\Delta}\mathbf{C}\mathbf{F}\mathbf{D}^{-1}\boldsymbol{\Omega}^{\omega}\mathbf{D}^{'-1}\mathbf{F}^{\prime}\mathbf{C}^{'}\boldsymbol{\Delta}^{\prime})^{-1}\\ &= \mathbf{D}^{'-1}\mathbf{F}^{\prime}\mathbf{C}^{\prime}(\mathbf{C}\mathbf{F}(\mathbf{D}^{\prime}(\boldsymbol{\Omega}^{\omega})^{-1}\mathbf{D})^{-1}\mathbf{F}^{\prime}\mathbf{C}^{\prime})^{-1}\boldsymbol{\Delta}^{-1}\end{aligned}$$

Observe the fact that $\mathbf{Q}\mathbf{F} := \boldsymbol{\Delta}\mathbf{C}\mathbf{F}\mathbf{D}^{-1}$ and define the $m \times N$ matrix

$$\phi := \mathbf{D}^{'-1}\mathbf{F}^{'}\mathbf{C}^{'}\left(\mathbf{C}\mathbf{F}\left(\mathbf{D}^{'}\left(\boldsymbol{\Omega}^{\omega}\right)^{-1}\mathbf{D}\right)^{-1}\mathbf{F}^{'}\mathbf{C}^{'}\right)^{-1}$$

Observe that $\phi_1\mathbf{Q}\mathbf{F} = \phi\boldsymbol{\Delta}^{-1}\boldsymbol{\Delta}\mathbf{C}\mathbf{F}\mathbf{D}^{-1} = \phi\mathbf{C}\mathbf{F}\mathbf{D}^{-1}$. Furthermore, define

$$\begin{aligned}\delta_1 &:= (\mathbf{S}^{'}\mathbf{D}^{'}\phi_1\mathbf{Q}\mathbf{F}\mathbf{D}\mathbf{S})^{-1}\mathbf{S}^{'}\mathbf{D}^{'}\phi_1\\ &= (\mathbf{S}^{'}\mathbf{D}^{'}\phi\mathbf{C}\mathbf{F}\mathbf{D}^{-1}\mathbf{D}\mathbf{S})^{-1}\mathbf{S}^{'}\mathbf{D}^{'}\phi\boldsymbol{\Delta}^{-1}\end{aligned}$$

Let $\delta := \left(\mathbf{S}^{'}\mathbf{D}^{'}\phi\mathbf{C}\mathbf{F}\mathbf{D}^{-1}\mathbf{D}\mathbf{S}\right)^{-1}\mathbf{S}^{'}\mathbf{D}^{'}\phi$. Then,

$$\begin{aligned}\mathbf{R} &= \boldsymbol{\Omega}\phi_1 + (\mathbf{I}_m - \boldsymbol{\Omega}\phi_1\mathbf{Q}\mathbf{F})\mathbf{D}\mathbf{S}(\mathbf{S}^{'}\mathbf{D}^{'}\phi_1\mathbf{Q}\mathbf{F}\mathbf{D}\mathbf{S})^{-1}\mathbf{S}^{'}\mathbf{D}^{'}\phi_1\\ &= \boldsymbol{\Omega}\phi_1 + (\mathbf{I}_m - \boldsymbol{\Omega}\phi_1\mathbf{Q}\mathbf{F})\mathbf{D}\mathbf{S}\delta_1\\ &= \boldsymbol{\Omega}\phi + (\mathbf{I}_m - \boldsymbol{\Omega}\phi\mathbf{C}\mathbf{F}\mathbf{D}^{-1})\mathbf{D}\mathbf{S}\delta\boldsymbol{\Delta}^{-1}\end{aligned}$$

Recall that $\mathbf{D}\hat{\mathbf{z}} = \mathbf{R}\Delta\mathbf{x}_{lf}$, hence

$$\begin{aligned}\hat{\mathbf{z}} &= \mathbf{D}^{-1}\mathbf{R}\Delta\mathbf{x}_{lf} \\ &= \mathbf{A}_{litt}\mathbf{x}_{lf}\end{aligned}$$

where

$$\begin{aligned}\mathbf{A}_{litt} &:= \mathbf{D}^{-1}\mathbf{R}\Delta \\ &= \mathbf{D}^{-1}(\boldsymbol{\Omega}^{\omega}\phi + (\mathbf{I}_m - \boldsymbol{\Omega}^{\omega}\phi\mathbf{CFD}^{-1})\mathbf{DS}\delta) \\ &= \mathbf{D}^{-1}(\boldsymbol{\Omega}^{\omega}\phi(\mathbf{I}_N - \mathbf{CFS}\delta) + \mathbf{DS}\delta) \\ &= \mathbf{D}^{-1}\boldsymbol{\Omega}^{\omega}\phi(\mathbf{I}_N - \mathbf{CFS}\delta) + \mathbf{S}\delta\end{aligned}$$

Also, $\hat{\mathbf{e}}_{lf} = \mathbf{x}_{lf} - \mathbf{CS}'\delta\mathbf{x}_{lf}$

### 1.A.2 Proof of Lemma 1.2.1.1

The proof consists of three parts for: PCL, PL, and PF. The proof provided here is different than the one stated in Poissonnier (2018)—however, the same result.

- The matrix representation of the proposed equation (1.6) is given by

$$\mathbf{x} = \mathbf{M}_1\mathbf{x} + \mathbf{M}_2\mathbf{x}^{init} + \mathbf{FS}\varphi + \mathbf{Fe}$$

  where the $\mathbf{C} \times c$ and the $\mathbf{C} \times p$ matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ are defined in the lemma. With some matrix algebra, one can reach

$$\mathbf{x} = (\mathbf{I}_c - \mathbf{M}_1)^{-1}(\mathbf{M}_2\mathbf{x}^{init} + \mathbf{FS}\varphi + \mathbf{Fe}) \tag{1.16}$$

Observe that its interpolated estimate of $\mathbf{x}$, $\hat{\mathbf{x}}$, is given by the suggested linear transformation[5]

$$\hat{\mathbf{x}} = f(\mathbf{x}_{lf}) = \mathbf{A}\mathbf{x}_{lf} + \mathbf{b} = \mathbf{AC}x + \mathbf{b}$$

the residual is given by

$$\hat{\mathbf{u}}_x = \mathbf{x} - \hat{\mathbf{x}} = (\mathbf{I}_c - \mathbf{AC})\mathbf{x} - \mathbf{b}$$

Observe that the matrix $\boldsymbol{A}$ and the vector $\boldsymbol{b}$ should satisfy the following three properties: 1) When you pre-multiply the interpolated estimate vector $\hat{\boldsymbol{x}}$, by the conversion matrix $\boldsymbol{C}$, you get the recorded observation vector $\boldsymbol{x}_{lf}$, that is: $\mathbf{C}\hat{\boldsymbol{x}} = \boldsymbol{x}_{lf}$, 2) The estimate, $\hat{\boldsymbol{x}}$, needs to be unbiased $\mathbb{E}\left(\hat{\mathbf{u}}_x\right) = 0$. 3) The estimate, $\hat{\boldsymbol{x}}$, needs to be efficient subject to the other initial two constraints i.e.

$$\min_{\mathbf{A},\mathbf{b}} cov(\hat{\mathbf{u}}_x) \quad \text{s.t.} \quad \mathbf{C}\hat{\mathbf{x}} = \mathbf{x}_{lf} \qquad \text{and} \qquad \mathbb{E}(\hat{\mathbf{u}}_x) = 0$$

1. Solving the first constraint, we get

$$\mathbf{C}\hat{\mathbf{x}}_{inter} = \mathbf{CA}\mathbf{x}_{\mathbf{lf}} + \mathbf{Cb} = \mathbf{x}_{lf}$$

   hence

$$\mathbf{CA} = \mathbf{I}_N \qquad \text{and} \qquad \mathbf{Cb} = \mathbf{0}_N$$

[5] We chose to write it here as $\hat{\mathbf{x}}$ rather than $\hat{\mathbf{z}}$ purposely to distinguish between interpolated and extrapolated estimates.

2. Solving the second constraint, we get

$$\mathbf{b} = (\mathbf{I}_c - \mathbf{AC})\mathbb{E}(\mathbf{x})$$

3. Define the variance covariance matrix of the vector $\mathbf{x}$, by the $(m \times m)$ matrix $\boldsymbol{\Omega}_x := Var(\mathbf{x}) = (\mathbf{I}_c - \mathbf{M}_1)^{-1}\mathbf{F}\boldsymbol{\Omega}^e\mathbf{F}'\left(\mathbf{I}_c - \mathbf{M}_1'\right)^{-1}$. Furthermore, Observe that

$$\begin{aligned} cov(\hat{\mathbf{u}}_x) &= (\mathbf{I}_c - \mathbf{AC})cov(x)(\mathbf{I}_c - \mathbf{AC})' \\ &= (\mathbf{I}_c - \mathbf{AC})(\mathbf{I}_c - \mathbf{M}_1)^{-1}\mathbf{F}\boldsymbol{\Omega}^e\mathbf{F}'(\mathbf{I}_c - \mathbf{M}_1')^{-1}(\mathbf{I}_c - \mathbf{AC})' \end{aligned}$$

We wish to minimize a scalar function $f_1(A)$

$$\begin{aligned} f_1(A) :&= \frac{1}{2}tr(cov(\hat{\mathbf{u}}_x)) \\ &= \frac{1}{2}tr(\boldsymbol{\Omega}_x - \mathbf{AC}\boldsymbol{\Omega}_x - \boldsymbol{\Omega}_x\mathbf{C}'\mathbf{A}' + \mathbf{AC}\boldsymbol{\Omega}_x\mathbf{C}'\mathbf{A}') \end{aligned}$$

hence

$$\frac{\partial f_1}{\partial \mathbf{A}} = (\mathbf{AC} - \mathbf{I}_m)\boldsymbol{\Omega}_x\mathbf{C}' = 0$$

This results

$$\mathbf{A} = \boldsymbol{\Omega}_x\mathbf{C}'(\mathbf{C}\boldsymbol{\Omega}_x\mathbf{C}')^{-1}$$

It is trivial to prove that the matrix $\mathbf{A}$ satisfies the other two constraints.

- Using the matrix representation of equation (1.6), i.e. (1.16). Let $\mathbf{R}$ and $\mathbf{b}_1$ be the $\mathbf{C} \times N$ matrix and the $\mathbf{C} \times 1$ vector to be determined respectively, such that the following equation holds:

$$\mathbf{D}_1\hat{\mathbf{x}} = \mathbf{R}\boldsymbol{\Delta}\mathbf{x}_{lf}+\mathbf{b}_1 = \mathbf{R}\boldsymbol{\Delta}\mathbf{C}\mathbf{x}+\mathbf{b}_1 = \mathbf{R}\boldsymbol{\Delta}\mathbf{C}(I_c-\mathbf{M}_1)^{-1}(\mathbf{M}_2\mathbf{x}^{init}+\mathbf{F}\mathbf{S}\varphi+\mathbf{F}\mathbf{e})+\mathbf{b}_1$$

where $\boldsymbol{\Delta}$ is $N \times N$ matrix defined in 1.14, $\mathbf{D}$ is the $m \times m$ matrix defined in 1.2.1. $\mathbf{D}_1$ is the $\mathbf{C} \times c$ difference matrix of similar structure to $\mathbf{D}$ and $\boldsymbol{\Delta}$ such that $\mathbf{F}\mathbf{D} = \mathbf{D}_1\mathbf{F}$ where $\mathbf{F}$ is the $\mathbf{C} \times m$ identity matrix.

$$\mathbf{D}_1\mathbf{x} =(\mathbf{I}_c - \mathbf{M}_1)^{-1}\Lambda + (\mathbf{I}_c - \mathbf{M}_1)^{-1}\mathbf{F}\mathbf{D}\mathbf{e}$$

where $\Lambda$ is a $\mathbf{C} \times 1$ vector defined as $\Lambda := \mathbf{D}_1\mathbf{M}_2\mathbf{x}^{init} + \mathbf{F}\mathbf{D}\mathbf{S}\varphi$. Let $\mathbf{Q}$ be an $N \times c$ matrix defined by $\mathbf{Q}\mathbf{D}_1 := \boldsymbol{\Delta}\mathbf{C}(\mathbf{I}_c - \mathbf{M}_1)^{-1}$, hence

$$\begin{aligned}\mathbf{D}_1\hat{\mathbf{x}} =&\mathbf{R}\mathbf{Q}\mathbf{D}_1(\mathbf{M}_2\mathbf{x}^{init} + \mathbf{F}\mathbf{S}\varphi + \mathbf{F}\mathbf{e}) + \mathbf{b}_1\\ &= \mathbf{R}\mathbf{Q}(\mathbf{D}_1\mathbf{M}_2\mathbf{x}^{init} + \mathbf{D}_1\mathbf{F}\mathbf{S}\varphi) + \mathbf{R}\mathbf{Q}\mathbf{D}_1\mathbf{F}\mathbf{e} + \mathbf{b}_1\\ &= \mathbf{R}\mathbf{Q}\Lambda + \mathbf{R}\mathbf{Q}\mathbf{F}\mathbf{D}\mathbf{e} + \mathbf{b}_1\end{aligned}$$

The residual is given by

$$\begin{aligned}\hat{\mathbf{u}}_x &= \mathbf{D}_1\mathbf{x} - \mathbf{D}_1\hat{\mathbf{x}}\\ &= \lambda_1(\Lambda + \mathbf{F}\mathbf{D}\mathbf{e}) - \mathbf{b}_1\end{aligned}$$

where $\lambda_1 := (\mathbf{I}_c - \mathbf{M}_1)^{-1} - \mathbf{R}\mathbf{Q}$. Observe that 1) When you pre-multiply the interpolated estimate vector $\hat{\boldsymbol{x}}$, by the conversion matrix $\boldsymbol{C}$, you get the recorded observation vector $\boldsymbol{x}_{lf}$, that is: $\mathbf{C}\hat{\boldsymbol{x}} = \boldsymbol{x}_{lf}$, hence

$$\mathbf{C}\hat{\mathbf{x}} = \mathbf{C}\mathbf{D}_{\mathbf{1}}^{-\mathbf{1}}\mathbf{R}\boldsymbol{\Delta}\mathbf{x}_{lf} + \mathbf{C}\mathbf{D}_1^{-1}\mathbf{b}_1 = \mathbf{x}_{lf}$$

hence $\mathbf{C}\mathbf{D}_{\mathbf{1}}^{-\mathbf{1}}\mathbf{R}\boldsymbol{\Delta} = \mathbf{I}_N$ and $\mathbf{C}\mathbf{D}_1^{-1}\mathbf{b}_1 = \mathbf{0}_{N\times 1}$. Furthermore, 2) The estimate, $\mathbf{D}_1\hat{\boldsymbol{x}}$, needs to be unbiased $\mathbb{E}\left(\hat{\mathbf{u}}_x\right) = 0$. Hence, $\mathbf{b}_1 = \lambda_1\Lambda$. Using simple algebra

$$\mathbf{C}\mathbf{D}_1^{-1}\mathbf{R}\mathbf{Q}\Lambda = \mathbf{C}\mathbf{D}_1^{-1}(\mathbf{I}_c - \mathbf{M}_1)^{-1}\Lambda$$

the variance covariance matrix of $\hat{\mathbf{u}}_x$ is

$$\begin{aligned}\mathbb{E}(\hat{\mathbf{u}}_x\hat{\mathbf{u}}_x^{'}) &= \lambda_1\mathbf{F}\mathbb{E}(\mathbf{D}\mathbf{e}\mathbf{e}^{'}\mathbf{D}^{'})\mathbf{F}^{'}\lambda_1^{'} \\ &= \lambda_1\mathbf{F}\boldsymbol{\Omega}^{\omega}\mathbf{F}^{'}\lambda_1^{'}\end{aligned}$$

where $\boldsymbol{\Omega}^{\omega}$ is the $m \times m$ variance matrix of the error terms in the stochastic disturbance equation. Our objective is to minimize the variance matrix of $\hat{\mathbf{u}}_x$ such that $\mathbf{C}\mathbf{D}_{\mathbf{1}}^{-\mathbf{1}}\mathbf{R}\boldsymbol{\Delta} = \mathbf{I}_N$ holds

$$\mathscr{L} = \frac{1}{2}tr(\mathbb{E}(\hat{\mathbf{u}}_x\hat{\mathbf{u}}_x^{'})) - tr(\wp^{'}(\mathbf{C}\mathbf{D}_{\mathbf{1}}^{-\mathbf{1}}\mathbf{R}\boldsymbol{\Delta} - \mathbf{I}_N))$$

$$\frac{\partial\mathscr{L}}{\partial\mathbf{R}} = -\lambda_1\mathbf{F}\boldsymbol{\Omega}^{\omega}\mathbf{F}^{'}\mathbf{Q}^{'} - \mathbf{D}_1^{'-1}\mathbf{C}^{'}\wp\boldsymbol{\Delta}^{'} = 0$$

since $\mathbf{Q}^{'} = \mathbf{D}_1^{'-1}(I_c - \mathbf{M}_1^{'})^{-1}\mathbf{C}^{'}\boldsymbol{\Delta}^{'}$ and $\left(\mathbf{Q}\mathbf{F}\boldsymbol{\Omega}^{\omega}\mathbf{F}^{'}\mathbf{Q}^{'}\right)^{-1} = \boldsymbol{\Delta}^{'-1}\boldsymbol{\Omega}_{x,1}^{-1}\boldsymbol{\Delta}^{-1}$ where $\boldsymbol{\Omega}_{x,1} = \mathbf{C}\boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}}\mathbf{C}^{'}$ and $\boldsymbol{\Omega}_{x,2} := (\mathbf{I}_c - \mathbf{M}_1)^{-1}\left(\mathbf{D}_{\mathbf{1}}^{'}\left(\mathbf{F}\boldsymbol{\Omega}^{\omega}\mathbf{F}^{'}\right)^{-1}\mathbf{D}_1\right)^{-1}(\mathbf{I}_c - \mathbf{M}_1)^{-1}$. One can show that,

$$\mathbf{R} = \mathbf{D}_1^{'-1}\mathbf{C}^{'}\wp\boldsymbol{\Omega}_{x,1}^{-1}\boldsymbol{\Delta}^{-1} + (\mathbf{I}_c - \mathbf{M}_1)^{-1}\mathbf{F}\boldsymbol{\Omega}^{\omega}\mathbf{F}^{'}\mathbf{D}_1^{'-1}(I_c - \mathbf{M}_1^{'})^{-1}\mathbf{C}^{'}\boldsymbol{\Omega}_{x,1}^{-1}\boldsymbol{\Delta}^{-1}$$

since $\mathbf{C}\mathbf{D}_\mathbf{1}^{-\mathbf{1}}\mathbf{R}\boldsymbol{\Delta} = \mathbf{I}_N$

$$\mathbf{C}\mathbf{D}_\mathbf{1}^{-\mathbf{1}}\mathbf{D}_1^{'-1}\mathbf{C}'\wp = \boldsymbol{\Omega}_{x,1} - \mathbf{C}(\mathbf{I}_c - \mathbf{M}_1)^{-1}\mathbf{D}_\mathbf{1}^{-\mathbf{1}}\mathbf{F}\boldsymbol{\Omega}^\omega\mathbf{F}'\mathbf{D}_1^{'-1}(I_c - \mathbf{M}_1')^{-1}\mathbf{C}'$$
$$= 0$$

Hence,

$$\mathbf{R} = \mathbf{D_1}\boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}}\mathbf{C}'\boldsymbol{\Omega}_{x,1}^{-1}\boldsymbol{\Delta}^{-1}$$

and

$$\lambda_1 = (\mathbf{I}_c - \mathbf{M}_1)^{-1} - \mathbf{D_1}\boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}}\mathbf{C}'\boldsymbol{\Omega}_{x,1}^{-1}\mathbf{C}(I_c - \mathbf{M}_1)^{-1}\mathbf{D}_1^{-1}$$

Observe that $\mathbf{D}_\mathbf{1}^{-\mathbf{1}}\lambda_1 = \lambda_{pois,1}\mathbf{D}_1^{-1}$, where $\lambda_{pois,1} := (\mathbf{I}_c - \mathbf{M}_1)^{-1} - \boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}}\mathbf{C}'\boldsymbol{\Omega}_{x,1}^{-1}\mathbf{C}(I_c - \mathbf{M}_1)^{-1}$. Hence,

$$\hat{\mathbf{x}} = \mathbf{D}_\mathbf{1}^{-\mathbf{1}}\mathbf{R}\boldsymbol{\Delta}\mathbf{x}_{lf} + \mathbf{D}_\mathbf{1}^{-\mathbf{1}}\lambda_1\Lambda$$

$$\begin{aligned}\mathbf{b} &= \mathbf{D}_\mathbf{1}^{-\mathbf{1}}\lambda_1\Lambda \\ &= \lambda_{pois,1}\mathbf{D}_1^{-1}\Lambda \\ &= \lambda_{pois,1}(\mathbf{M}_2\mathbf{x}^{init} + \mathbf{F}\mathbf{S}\varphi)\end{aligned} \quad \text{and} \quad \begin{aligned}\mathbf{A}_{pois,2} &:= \mathbf{D}_\mathbf{1}^{-\mathbf{1}}\mathbf{R}\boldsymbol{\Delta} \\ &= \boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}}\mathbf{C}'\boldsymbol{\Omega}_{x,1}^{-1}\end{aligned}$$

### 1.A.3 Proof of Proposition 1.2.2

The proof consists of three parts for: PCL, PL, and PF.

1. The matrix representation of the proposed equation (1.6) is given by

$$\mathbf{z} := \mathbf{M}_{1,z}\mathbf{z} + \mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi + \mathbf{e}$$

where the $m \times m$ and the $m \times p$ matrices $\mathbf{M}_{1,z}$ and $\mathbf{M}_{2,z}$ are defined of similar structure to the $\mathbf{M}_1$ and $\mathbf{M}_2$ in the Lemma. In particular, $\mathbf{F}\mathbf{M}_{2,z} = \mathbf{M}_2$ and

$\mathbf{F}\mathbf{M}_{\mathbf{1},\mathbf{z}}\mathbf{F}^{'} = \mathbf{M}_1$. With some matrix algebra, one can reach

$$\mathbf{z} = (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}(\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi + \mathbf{e})$$

Observe that $\mathbf{x} = \mathbf{F}\mathbf{z} = \mathbf{F}\mathbf{M}_{1,z}\mathbf{z} + \mathbf{F}\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{F}\mathbf{S}\varphi + \mathbf{F}\mathbf{e} = \mathbf{M}_{\mathbf{1}}\mathbf{F}\mathbf{z} + \mathbf{M}_2\mathbf{z}^{init} + \mathbf{F}\mathbf{S}\varphi + \mathbf{F}\mathbf{e}$. Hence, $\mathbf{x} = (\mathbf{I}_c - \mathbf{M}_1)^{-1}(\mathbf{M}_2\mathbf{z}^{init} + \mathbf{F}\mathbf{S}\varphi + \mathbf{F}\mathbf{e})$. Furthermore, its estimate, $\hat{\mathbf{z}}$, is given by the suggested linear transformation

$$\hat{\mathbf{z}}= \mathbf{A}\mathbf{x}_{lf} + \mathbf{b}$$

where $\mathbf{A}$ is a $m \times N$ matrix and $\mathbf{b}$ is a $m \times 1$ vector to be determined. Observe that

$$\mathbf{z} = (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}(\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi + \mathbf{e}) = \Lambda_2 + (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}\mathbf{e}$$

Where $\Lambda$ is a $\mathbf{C} \times 1$ vector defined as $\Lambda := (\mathbf{I}_c - \mathbf{M}_1)^{-1}(\mathbf{M}_2\mathbf{x}^{init} + \mathbf{F}\mathbf{S}\varphi) = \mathbf{F}\Lambda_2$, $\Lambda_2 := (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}(\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi)$.

$$\hat{\mathbf{z}} = \mathbf{A}\mathbf{x}_{lf} + \mathbf{b} = \mathbf{A}\mathbf{C}\mathbf{F}\boldsymbol{\Lambda}_{\mathbf{2}} + \mathbf{A}\mathbf{C}\mathbf{F}(\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}\mathbf{e}+b$$

The residuals is defined by

$$\hat{\mathbf{u}}_z = \mathbf{z} - \hat{\mathbf{z}} = (\mathbf{I}_m - \mathbf{A}\mathbf{C}\mathbf{F})\Lambda_2 + (\mathbf{I}_m - \mathbf{A}\mathbf{C}\mathbf{F})(\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}e - b$$

Let $\mathbf{C}_z$ be the $N \times m$ matrix, defined exactly in the same manner as the $N \times c$ matrix $\mathbf{C}$ defined in either set-ups. In particular, when post-multiplied by the vector $\hat{\mathbf{z}}$ one should get the vector $\mathbf{x}_{lf}$, that is: $\mathbf{C}_z\hat{\mathbf{z}} = \boldsymbol{x}_{lf}$. Hence, $\mathbf{C_z}\hat{\mathbf{z}} = \mathbf{C_z}\mathbf{A}\mathbf{x_{lf}} + \mathbf{C_z}\mathbf{b} = \mathbf{x}_{lf}$. This implies that $\mathbf{C_z}\mathbf{A} = \mathbf{I}_N$ and $\mathbf{C_z}\mathbf{b} = 0_N$. Furthermore, 2) The estimate, $\hat{\mathbf{z}}$, needs to be unbiased estimate of $\mathbf{z}$, $\mathbb{E}\left(\hat{\mathbf{u}}_z\right) = 0$. Hence, $\mathbf{b} = \left(\mathbf{I}_m - \mathbf{A}\mathbf{C_z}\right)\Lambda_2$.

Observe that if $\mathbf{C_z}\mathbf{A} = \mathbf{I}_N$, then $\mathbf{b}$ satisfies $\mathbf{C_z}\mathbf{b} = 0_N$. So the only constraint that needs to be guaranteed is $\mathbf{C_z}\mathbf{A} = \mathbf{I}_N$. Let $\lambda_1 := \left(\mathbf{I}_m - \mathbf{A}\mathbf{C_z}\right)\left(\mathbf{I_m} - \mathbf{M_{1,z}}\right)^{-\mathbf{1}}$, then the variance covariance matrix of $\hat{\mathbf{u}}_x$ is

$$\mathbb{E}(\hat{\mathbf{u}}_x\hat{\mathbf{u}}_x^{'}) = \lambda_1\mathbb{E}(\mathbf{e}\mathbf{e}^{'})\lambda_1^{'} = \lambda_1\boldsymbol{\Omega}\lambda_1^{'}$$

where $\boldsymbol{\Omega}^e$ is the $m \times m$ variance matrix of the error terms in the stochastic disturbance equation. Our target is to minimize this matrix such that the $\mathbf{C_z}\mathbf{A} = \mathbf{I}_N$ holds. The Lagrangian expression is

$$\mathscr{L} = \frac{1}{2}tr(\mathbb{E}(\hat{\mathbf{u}}_x\hat{\mathbf{u}}_x^{'}))$$

The partial derivative is given by

$$\frac{\partial\mathscr{L}}{\partial\mathbf{A}} = -\lambda_1\boldsymbol{\Omega}^e(\mathbf{I}_m - \mathbf{M}_{1,z}^{'})^{-1}\mathbf{C}_z^{\prime} = 0$$

Let $\boldsymbol{\Omega}_{x,1} = \mathbf{C_z}\boldsymbol{\Omega}_{\mathbf{x},\mathbf{2}}\mathbf{C}_z^{\prime}$ and $\boldsymbol{\Omega}_{x,2} := (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}\boldsymbol{\Omega}^e\left(\mathbf{I}_m - \mathbf{M}_{1,z}^{'}\right)^{-1}$, hence solving for $\mathbf{A}$, we get

$$\mathbf{A} = \boldsymbol{\Omega}_{x,2}\mathbf{C}_z^{\prime}\boldsymbol{\Omega}_{x,1}^{-1} \qquad \text{and} \qquad \mathbf{b} = \left(\mathbf{I}_m - \mathbf{A}\mathbf{C_z}\right)\left(\mathbf{I}_m - \mathbf{M}_{1,z}\right)^{-1}(\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi)$$

It is easy to see that the constraint $\mathbf{C_z A} = \mathbf{I}_N$ holds for this choice of $\mathbf{A}$.

2. Under the assumptions of ARIMA(1,1,0) stochastic disturbance terms, we will consider the following linear transformation equation

$$\mathbf{D\hat{z}} = \mathbf{R\Delta x}_{lf} + \mathbf{b}_1 = \mathbf{R\Delta Cx} + \mathbf{b}_1 = \mathbf{R\Delta C_z z} + \mathbf{b}_1$$

where $\mathbf{D}$ and $\Delta$ is the $m \times m$ and the $N \times N$ difference matrix. $\mathbf{R}$ is a $m \times N$ matrix and $\mathbf{b}_1$ is a $m \times 1$ vector to be determined. Observe that

$$\mathbf{Dz} = \mathbf{D}(I_m - \mathbf{M}_{1,z})^{-1}(\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi + \mathbf{e}) = \Lambda_2 + (I_m - \mathbf{M}_{1,z})^{-1}\mathbf{De}$$

where $\Lambda_2 := \mathbf{D}(I_m - \mathbf{M}_{1,z})^{-1}(\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi)$.

Hence, $\mathbf{D\hat{z}}= \mathbf{R\Delta C_z D}^{-1}\Lambda_2 + \mathbf{R\Delta C_z (I_m - M_{1,z})^{-1}e} + \mathbf{b}_1$.

The differenced residuals is given by

$$\mathbf{\hat{u}}_z = \mathbf{Dz} - \mathbf{D\hat{z}} = (I_m - \mathbf{R\Delta C_z D}^{-1})\Lambda_2 + \lambda_1 \mathbf{De} - \mathbf{b}_1$$

where $\lambda_1 := (\mathbf{I_m} - \mathbf{M_{1,z}})^{-1} - \mathbf{R\Delta C_z} \left(\mathbf{I_m} - \mathbf{M_{1,z}}\right)^{-1}\mathbf{D}^{-1}$. When pre-multiplying the vector $\mathbf{\hat{z}}$ by $\mathbf{C}_z$ one should get the vector $\mathbf{x}_{lf}$. In particular,

$$\mathbf{C_z \hat{z}} = \mathbf{C_z D}^{-1}\mathbf{R\Delta x_{lf}} + \mathbf{C_z D}^{-1}\mathbf{b}_1 = \mathbf{x_{lf}}$$

Hence, $\mathbf{C_z D}^{-1}\mathbf{R\Delta} = \mathbf{I}_N$ and $\mathbf{C_z D}^{-1}\mathbf{b}_1 = 0_N$ should be satisfied. Furthermore, the differenced estimate, $\mathbf{D\hat{z}}$, needs to be unbiased estimate of $\mathbf{Dz}$, $\mathbb{E}\left(\mathbf{\hat{u}}_z\right) = 0$. Hence, $\mathbf{b}_1 = (I_m - \mathbf{R\Delta C_z D}^{-1})\Lambda_2$. Observe that if $\mathbf{C_z D}^{-1}\mathbf{R\Delta} = \mathbf{I}_N$, then $\mathbf{b}_1$

satisfies $\mathbf{C_z D^{-1} b}_1 = 0$. So the only constraint that need to be guaranteed is $\mathbf{C_z D^{-1} R \Delta} = \mathbf{I}_N$. The variance covariance matrix of $\hat{\mathbf{u}}_x$ is

$$\mathbb{E}(\hat{\mathbf{u}}_x \hat{\mathbf{u}}_x^{'}) = \lambda_1 \mathbb{E}(\mathbf{D e e}^{'} \mathbf{D}^{'}) \lambda_1^{'} = \lambda_1 \mathbf{\Omega}^{\omega} \lambda_1^{'}$$

where $\mathbf{\Omega}^{\omega}$ is the $m \times m$ variance matrix of the error terms in the stochastic disturbance equation. Our target is to minimize this matrix such that the constraint holds. The Lagrangian expression is

$$\mathscr{L} = \frac{1}{2} tr(E(\hat{u}_x \hat{u}_x^{'}))$$

The partial derivative is given by

$$\begin{aligned}\frac{\partial \mathscr{L}}{\partial \mathbf{R}} &= -\lambda_1 \mathbf{\Omega}^{\omega} \mathbf{D}^{'-1} (\mathbf{I}_m - \mathbf{M}_{1,z}^{\prime})^{-\mathbf{1}} \mathbf{F}^{\prime} \mathbf{C}^{'} \mathbf{\Delta}^{'} \\ &= (\mathbf{R \Delta C_z} (\mathbf{I}_m - \mathbf{M}_{1,z})^{-\mathbf{1}} \mathbf{D}^{-1} - (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}) \mathbf{\Omega}^{\omega} \mathbf{D}^{'-1} (\mathbf{I}_m - \mathbf{M}_{1,z}^{'})^{-\mathbf{1}} \mathbf{F}^{\prime} \mathbf{C}^{\prime} \mathbf{\Delta}^{\prime} \\ &= 0\end{aligned}$$

Let $\mathbf{\Omega}_{x,1} = \mathbf{C_z \Omega_{x,2} C}_z^{\prime}$ and $\mathbf{\Omega}_{x,2} := (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1} \mathbf{D}^{-1} \mathbf{\Omega}^{\omega} \mathbf{D}^{\prime -1} \left(\mathbf{I}_m - \mathbf{M}_{1,z}^{'}\right)^{-1}$, hence solving for $\mathbf{R}$, we get

$$\mathbf{R} = (\mathbf{I}_m - \mathbf{M}_{1,z})^{-1} \mathbf{\Omega}^{\omega} \mathbf{D}^{'-1} (\mathbf{I}_m - \mathbf{M}_{1,z}^{'})^{-\mathbf{1}} \mathbf{C}_z^{\prime} (\mathbf{\Omega}_{x,1})^{-1} \mathbf{\Delta}^{-1}$$

It is easy to see that the constraint $\mathbf{C_z D^{-1} R \Delta} = \mathbf{I}_N$ holds for this choice of $\mathbf{R}$. Recall that

$$\hat{\mathbf{z}} = \mathbf{D}^{-1} \mathbf{R \Delta x}_{lf} + \mathbf{D}^{-1} \mathbf{b}_1$$

Hence,

$$\mathbf{A}_{pois,2} = \mathbf{D}^{-1}\mathbf{R}\boldsymbol{\Delta} = \boldsymbol{\Omega}_{x,2}\mathbf{C}^{'}(\boldsymbol{\Omega}_{x,1})^{-1}$$

and

$$\mathbf{b}_{pois,2} := \mathbf{D}^{-1}\mathbf{b}_1 = (\mathbf{I}_m - \mathbf{A}_{pois,2}\mathbf{C}_{\mathbf{z}})(\mathbf{I}_m - \mathbf{M}_{1,z})^{-1}(\mathbf{M}_{2,z}\mathbf{z}^{init} + \mathbf{S}\varphi)$$

## 1.B Monte Carlo Study - Tables

Table 1.2: DGP Reference Key

| **DGP reference #** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **CLCL** | **CLCL** | **CLL** | **CLL** | **CLF** | **PCL** | **PCL** | **PCL** | **PCL** | **PL** | **PL** | **PL** | **PL** | **PF** | **PF** |
| $\mu$ | 0.2 | 0.8 | 0.2 | 0.8 | | 0.2 | 0.2 | 0.8 | 0.8 | 0.2 | 0.2 | 0.8 | 0.8 | | |
| $\rho$ | | | | | | 0.2 | 0.8 | 0.2 | 0.8 | 0.2 | 0.8 | 0.2 | 0.8 | 0.2 | 0.8 |

Table 1.3: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 0.1$ with aggregation matrix $\mathbf{C}_1$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **2.024** | **2.694** | 11.288 | 42.992 | 9.319 | **2.282** | 8.971 | **3.718** | 26.608 | 17.858 | 262.318 | 68.127 | 1028.365 | 14.232 | 14.106 |
| | (0.111) | (0.193) | (5.619) | (22.671) | (4.611) | (0.134) | (1.041) | (0.296) | (3.538) | (9.136) | (143.080) | (36.514) | (536.710) | (7.044) | (7.083) |
| CLL | 2.529 | 3.435 | 5.155 | 14.279 | 4.313 | 2.886 | 11.785 | 4.823 | 33.178 | 7.663 | 71.198 | 22.210 | 233.326 | 6.266 | 7.052 |
| | (0.362) | (0.589) | (1.120) | (4.466) | (0.961) | (0.427) | (2.653) | (0.824) | (8.755) | (1.857) | (25.460) | (7.082) | (95.650) | (1.497) | (1.997) |
| CLF | 2.307 | 3.016 | 4.205 | 11.016 | 3.528 | 2.607 | 9.475 | 4.134 | 26.034 | 6.127 | 52.620 | 16.956 | 174.798 | 5.061 | 5.464 |
| | (0.207) | (0.301) | (0.500) | (1.853) | (0.419) | (0.244) | (1.274) | (0.426) | (3.853) | (0.808) | (9.775) | (2.890) | (36.496) | (0.660) | (0.877) |
| PCL | 2.257 | 2.949 | 4.509 | 12.190 | 3.778 | 2.560 | **8.893** | 4.034 | **23.403** | 6.638 | 60.666 | 18.814 | 213.441 | 5.452 | 5.517 |
| | (0.164) | (0.237) | (0.710) | (2.868) | (0.601) | (0.202) | (0.984) | (0.348) | (2.956) | (1.139) | (18.943) | (4.671) | (74.874) | (0.931) | (0.914) |
| PL | 2.342 | 3.048 | 4.101 | 9.887 | 3.447 | 2.644 | 9.380 | 4.182 | 24.930 | 5.927 | 42.875 | 15.008 | 130.640 | 4.902 | 4.973 |
| | (0.172) | (0.258) | (0.424) | (1.296) | (0.359) | (0.217) | (1.096) | (0.381) | (3.330) | (0.665) | (7.152) | (2.048) | (23.620) | (0.524) | (0.538) |
| PF | 2.340 | 3.043 | **4.057** | **9.650** | **3.418** | 2.642 | 9.217 | 4.169 | 24.208 | **5.832** | **41.941** | **14.641** | **128.912** | **4.837** | **4.910** |
| | (0.172) | (0.256) | (0.409) | (1.242) | (0.348) | (0.216) | (1.059) | (0.378) | (3.122) | (0.644) | (6.737) | (1.955) | (23.237) | (0.506) | (0.520) |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **2.022** | **2.709** | 11.400 | 43.769 | 9.228 | **2.287** | 9.108 | **3.739** | 26.532 | 17.710 | 270.441 | 67.247 | 1041.455 | 19.097 | 19.624 |
| | (0.114) | (0.196) | (5.811) | (23.902) | (4.593) | (0.136) | (1.000) | (0.302) | (3.689) | (9.082) | (142.413) | (36.870) | (588.701) | (9.302) | (9.657) |
| CLL | 3.507 | 4.287 | 5.793 | 14.700 | 5.006 | 4.140 | 13.091 | 5.779 | 34.092 | 8.342 | 71.785 | 22.797 | 234.955 | 8.838 | 9.273 |
| | (0.708) | (1.015) | (1.596) | (4.519) | (1.318) | (0.849) | (3.468) | (1.380) | (9.352) | (2.346) | (26.265) | (7.094) | (91.480) | (2.047) | (2.308) |
| CLF | 2.905 | 3.508 | 4.553 | 11.210 | 3.922 | 3.333 | 10.227 | 4.663 | 26.327 | 6.496 | 53.200 | 17.193 | 173.333 | 7.161 | 7.380 |
| | (0.370) | (0.505) | (0.713) | (1.843) | (0.590) | (0.438) | (1.624) | (0.671) | (4.182) | (1.042) | (10.171) | (2.880) | (34.899) | (0.900) | (1.024) |
| PCL | 2.396 | 3.072 | 4.617 | 12.312 | 3.865 | 2.703 | **8.922** | 4.151 | **23.370** | 6.660 | 62.317 | 18.736 | 214.648 | 7.602 | 7.696 |
| | (0.179) | (0.241) | (0.748) | (2.998) | (0.578) | (0.201) | (0.954) | (0.353) | (3.097) | (1.132) | (19.290) | (4.774) | (81.083) | (1.261) | (1.210) |
| PL | 2.475 | 3.170 | 4.195 | 9.998 | 3.529 | 2.781 | 9.400 | 4.302 | 25.007 | 5.985 | 43.245 | 15.085 | 129.433 | 6.952 | 7.017 |
| | (0.188) | (0.264) | (0.433) | (1.342) | (0.334) | (0.215) | (1.062) | (0.387) | (3.334) | (0.645) | (7.088) | (2.059) | (23.916) | (0.766) | (0.725) |
| PF | 2.474 | 3.165 | **4.155** | **9.765** | **3.501** | 2.779 | 9.246 | 4.288 | 24.242 | **5.890** | **42.310** | **14.701** | **127.417** | **6.835** | **6.896** |
| | (0.187) | (0.262) | (0.421) | (1.265) | (0.326) | (0.214) | (1.021) | (0.383) | (3.194) | (0.621) | (6.693) | (1.970) | (23.437) | (0.735) | (0.690) |

The upper and lower tables present the results for DGPs with $\varphi = 0.2$ and 0.8, respectively. Each column represents a DGP according to table 1.2. The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.4: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 1$ with aggregation matrix $\mathbf{C}_1$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.870** | **3.821** | 16.362 | 63.419 | 12.808 | **3.225** | 12.606 | **5.257** | 37.596 | 24.771 | 370.621 | 95.229 | 1522.353 | 19.097 | 296.528 |
| | (0.162) | (0.274) | (8.435) | (33.590) | (6.253) | (0.184) | (1.388) | (0.440) | (5.012) | (12.297) | (206.809) | (49.425) | (851.675) | (9.302) | (154.906) |
| **CLL** | 3.621 | 4.922 | 7.416 | 20.537 | 6.174 | 4.092 | 16.599 | 6.824 | 47.161 | 10.892 | 100.482 | 30.959 | 333.804 | 8.838 | 77.948 |
| | (0.522) | (0.822) | (1.714) | (6.168) | (1.337) | (0.548) | (3.837) | (1.179) | (12.145) | (2.631) | (36.510) | (9.628) | (132.133) | (2.047) | (27.929) |
| **CLF** | 3.246 | 4.255 | 5.991 | 15.677 | 5.000 | 3.651 | 13.347 | 5.820 | 36.819 | 8.682 | 74.243 | 23.826 | 247.499 | 7.161 | 59.456 |
| | (0.284) | (0.431) | (0.743) | (2.497) | (0.579) | (0.304) | (1.744) | (0.612) | (5.584) | (1.112) | (13.982) | (3.773) | (48.612) | (0.900) | (10.813) |
| **PCL** | 3.220 | 4.176 | 6.420 | 17.562 | 5.329 | 3.628 | **12.496** | 5.703 | **32.947** | 9.335 | 86.076 | 26.419 | 309.639 | 7.602 | 69.630 |
| | (0.235) | (0.342) | (1.053) | (4.284) | (0.826) | (0.259) | (1.330) | (0.503) | (4.286) | (1.551) | (27.639) | (6.566) | (117.513) | (1.261) | (20.836) |
| **PL** | 3.313 | 4.320 | 5.851 | 14.024 | 4.896 | 3.738 | 13.266 | 5.921 | 35.144 | 8.400 | 60.315 | 21.208 | 183.631 | 6.952 | 48.890 |
| | (0.255) | (0.378) | (0.640) | (1.860) | (0.495) | (0.283) | (1.524) | (0.552) | (4.770) | (0.904) | (9.555) | (2.858) | (33.494) | (0.766) | (7.939) |
| **PF** | 3.309 | 4.307 | **5.762** | **13.667** | **4.837** | 3.732 | 12.951 | 5.889 | 34.092 | **8.229** | **59.064** | **20.704** | **181.261** | **6.835** | **47.884** |
| | (0.254) | (0.374) | (0.609) | (1.776) | (0.478) | (0.281) | (1.426) | (0.544) | (4.508) | (0.871) | (9.108) | (2.734) | (32.427) | (0.735) | (7.642) |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.865** | **3.814** | 16.091 | 59.530 | 12.765 | **3.222** | 12.741 | **5.243** | 37.336 | 24.469 | 375.206 | 94.898 | 1482.696 | 19.624 | 304.011 |
| | (0.154) | (0.285) | (7.839) | (31.016) | (6.228) | (0.198) | (1.362) | (0.438) | (5.097) | (12.186) | (210.825) | (51.295) | (781.425) | (9.657) | (162.312) |
| **CLL** | 4.378 | 5.461 | 7.809 | 20.595 | 6.585 | 5.060 | 17.562 | 7.422 | 47.176 | 11.299 | 100.570 | 31.620 | 335.536 | 9.273 | 81.110 |
| | (0.837) | (1.105) | (2.079) | (6.431) | (1.705) | (1.028) | (4.323) | (1.581) | (12.823) | (2.889) | (35.550) | (10.019) | (130.849) | (2.308) | (28.582) |
| **CLF** | 3.686 | 4.559 | 6.172 | 15.765 | 5.264 | 4.183 | 13.876 | 6.137 | 36.795 | 8.906 | 74.440 | 23.934 | 247.097 | 7.380 | 59.996 |
| | (0.429) | (0.558) | (0.933) | (2.564) | (0.772) | (0.518) | (2.022) | (0.771) | (5.850) | (1.247) | (13.577) | (3.961) | (49.183) | (1.024) | (11.196) |
| **PCL** | 3.317 | 4.245 | 6.408 | 17.119 | 5.392 | 3.719 | **12.534** | 5.728 | **32.882** | 9.331 | 87.071 | 26.487 | 306.880 | 7.696 | 69.908 |
| | (0.227) | (0.346) | (1.051) | (3.960) | (0.794) | (0.281) | (1.331) | (0.515) | (4.196) | (1.572) | (28.228) | (6.791) | (109.301) | (1.210) | (21.598) |
| **PL** | 3.411 | 4.393 | 5.825 | 14.027 | 4.961 | 3.824 | 13.299 | 5.946 | 35.082 | 8.428 | 60.479 | 20.906 | 183.361 | 7.017 | 48.860 |
| | (0.248) | (0.380) | (0.603) | (1.823) | (0.481) | (0.305) | (1.525) | (0.561) | (4.777) | (0.897) | (10.069) | (2.832) | (32.677) | (0.725) | (7.994) |
| **PF** | 3.407 | 4.380 | **5.742** | **13.661** | **4.904** | 3.818 | 12.989 | 5.913 | 34.031 | **8.266** | **59.268** | **20.409** | **180.951** | **6.896** | **47.747** |
| | (0.246) | (0.376) | (0.584) | (1.729) | (0.465) | (0.303) | (1.440) | (0.553) | (4.463) | (0.853) | (9.709) | (2.674) | (31.611) | (0.690) | (7.743) |

The upper and lower tables present the results for DGPs with $\varphi = 0.2$ and 0.8, respectively. Each column represents a DGP according to table 1.2. The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.5: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 5$ with aggregation matrix $\mathbf{C}_1$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.024** | **2.694** | 11.288 | 42.992 | 9.319 | **2.282** | 8.971 | **3.718** | 26.608 | 17.858 | 262.318 | 68.127 | 1028.365 | 14.232 | 14.106 |
| | (0.111) | (0.193) | (5.619) | (22.671) | (4.611) | (0.134) | (1.041) | (0.296) | (3.538) | (9.136) | (143.080) | (36.514) | (536.710) | (7.044) | (7.083) |
| **CLL** | 2.529 | 3.435 | 5.155 | 14.279 | 4.313 | 2.886 | 11.785 | 4.823 | 33.178 | 7.663 | 71.198 | 22.210 | 233.326 | 6.266 | 7.052 |
| | (0.362) | (0.589) | (1.120) | (4.466) | (0.961) | (0.427) | (2.653) | (0.824) | (8.755) | (1.857) | (25.460) | (7.082) | (95.650) | (1.497) | (1.997) |
| **CLF** | 2.307 | 3.016 | 4.205 | 11.016 | 3.528 | 2.607 | 9.475 | 4.134 | 26.034 | 6.127 | 52.620 | 16.956 | 174.798 | 5.061 | 5.464 |
| | (0.207) | (0.301) | (0.500) | (1.853) | (0.419) | (0.244) | (1.274) | (0.426) | (3.853) | (0.808) | (9.775) | (2.890) | (36.496) | (0.660) | (0.877) |
| **PCL** | 2.257 | 2.949 | 4.509 | 12.190 | 3.778 | 2.560 | **8.893** | 4.034 | **23.403** | 6.638 | 60.666 | 18.814 | 213.441 | 5.452 | 5.517 |
| | (0.164) | (0.237) | (0.710) | (2.868) | (0.601) | (0.202) | (0.984) | (0.348) | (2.956) | (1.139) | (18.943) | (4.671) | (74.874) | (0.931) | (0.914) |
| **PL** | 2.342 | 3.048 | 4.101 | 9.887 | 3.447 | 2.644 | 9.380 | 4.182 | 24.930 | 5.927 | 42.875 | 15.008 | 130.640 | 4.902 | 4.973 |
| | (0.172) | (0.258) | (0.424) | (1.296) | (0.359) | (0.217) | (1.096) | (0.381) | (3.330) | (0.665) | (7.152) | (2.048) | (23.620) | (0.524) | (0.538) |
| **PF** | 2.340 | 3.043 | **4.057** | **9.650** | **3.418** | 2.642 | 9.217 | 4.169 | 24.208 | **5.832** | **41.941** | **14.641** | **128.912** | **4.837** | **4.910** |
| | (0.172) | (0.256) | (0.409) | (1.242) | (0.348) | (0.216) | (1.059) | (0.378) | (3.122) | (0.644) | (6.737) | (1.955) | (23.237) | (0.506) | (0.520) |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.022** | **2.709** | 11.400 | 43.769 | 9.228 | **2.287** | 9.108 | **3.739** | 26.532 | 17.710 | 270.441 | 67.247 | 1041.455 | 19.097 | 19.624 |
| | (0.114) | (0.196) | (5.811) | (23.902) | (4.593) | (0.136) | (1.000) | (0.302) | (3.689) | (9.082) | (142.413) | (36.870) | (588.701) | (9.302) | (9.657) |
| **CLL** | 3.507 | 4.287 | 5.793 | 14.700 | 5.006 | 4.140 | 13.091 | 5.779 | 34.092 | 8.342 | 71.785 | 22.797 | 234.955 | 8.838 | 9.273 |
| | (0.708) | (1.015) | (1.596) | (4.519) | (1.318) | (0.849) | (3.468) | (1.380) | (9.352) | (2.346) | (26.265) | (7.094) | (91.480) | (2.047) | (2.308) |
| **CLF** | 2.905 | 3.508 | 4.553 | 11.210 | 3.922 | 3.333 | 10.227 | 4.663 | 26.327 | 6.496 | 53.200 | 17.193 | 173.333 | 7.161 | 7.380 |
| | (0.370) | (0.505) | (0.713) | (1.843) | (0.590) | (0.438) | (1.624) | (0.671) | (4.182) | (1.042) | (10.171) | (2.880) | (34.899) | (0.900) | (1.024) |
| **PCL** | 2.396 | 3.072 | 4.617 | 12.312 | 3.865 | 2.703 | **8.922** | 4.151 | **23.370** | 6.660 | 62.317 | 18.736 | 214.648 | 7.602 | 7.696 |
| | (0.179) | (0.241) | (0.748) | (2.998) | (0.578) | (0.201) | (0.954) | (0.353) | (3.097) | (1.132) | (19.290) | (4.774) | (81.083) | (1.261) | (1.210) |
| **PL** | 2.475 | 3.170 | 4.195 | 9.998 | 3.529 | 2.781 | 9.400 | 4.302 | 25.007 | 5.985 | 43.245 | 15.085 | 129.433 | 6.952 | 7.017 |
| | (0.188) | (0.264) | (0.433) | (1.342) | (0.334) | (0.215) | (1.062) | (0.387) | (3.334) | (0.645) | (7.088) | (2.059) | (23.916) | (0.766) | (0.725) |
| **PF** | 2.474 | 3.165 | **4.155** | **9.765** | **3.501** | 2.779 | 9.246 | 4.288 | 24.242 | **5.890** | **42.310** | **14.701** | **127.417** | **6.835** | **6.896** |
| | (0.187) | (0.262) | (0.421) | (1.265) | (0.326) | (0.214) | (1.021) | (0.383) | (3.194) | (0.621) | (6.693) | (1.970) | (23.437) | (0.735) | (0.690) |

The upper and lower tables present the results for DGPs with $\varphi = 0.2$ and 0.8, respectively. Each column represents a DGP according to table 1.2. The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.6: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 10$ with aggregation matrix $\mathbf{C}_1$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **2.870** | **3.821** | 16.362 | 63.419 | 12.808 | **3.225** | 12.606 | **5.257** | 37.596 | 24.771 | 370.621 | 95.229 | 1522.353 | 19.097 | 296.528 |
| | (0.162) | (0.274) | (8.435) | (33.590) | (6.253) | (0.184) | (1.388) | (0.440) | (5.012) | (12.297) | (206.809) | (49.425) | (851.675) | (9.302) | (154.906) |
| CLL | 3.621 | 4.922 | 7.416 | 20.537 | 6.174 | 4.092 | 16.599 | 6.824 | 47.161 | 10.892 | 100.482 | 30.959 | 333.804 | 8.838 | 77.948 |
| | (0.522) | (0.822) | (1.714) | (6.168) | (1.337) | (0.548) | (3.837) | (1.179) | (12.145) | (2.631) | (36.510) | (9.628) | (132.133) | (2.047) | (27.929) |
| CLF | 3.246 | 4.255 | 5.991 | 15.677 | 5.000 | 3.651 | 13.347 | 5.820 | 36.819 | 8.682 | 74.243 | 23.826 | 247.499 | 7.161 | 59.456 |
| | (0.284) | (0.431) | (0.743) | (2.497) | (0.579) | (0.304) | (1.744) | (0.612) | (5.584) | (1.112) | (13.982) | (3.773) | (48.612) | (0.900) | (10.813) |
| PCL | 3.220 | 4.176 | 6.420 | 17.562 | 5.329 | 3.628 | **12.496** | 5.703 | **32.947** | 9.335 | 86.076 | 26.419 | 309.639 | 7.602 | 69.630 |
| | (0.235) | (0.342) | (1.053) | (4.284) | (0.826) | (0.259) | (1.330) | (0.503) | (4.286) | (1.551) | (27.639) | (6.566) | (117.513) | (1.261) | (20.836) |
| PL | 3.313 | 4.320 | 5.851 | 14.024 | 4.896 | 3.738 | 13.266 | 5.921 | 35.144 | 8.400 | 60.315 | 21.208 | 183.631 | 6.952 | 48.890 |
| | (0.255) | (0.378) | (0.640) | (1.860) | (0.495) | (0.283) | (1.524) | (0.552) | (4.770) | (0.904) | (9.555) | (2.858) | (33.494) | (0.766) | (7.939) |
| PF | 3.309 | 4.307 | **5.762** | **13.667** | **4.837** | 3.732 | 12.951 | 5.889 | 34.092 | **8.229** | **59.064** | **20.704** | **181.261** | **6.835** | **47.884** |
| | (0.254) | (0.374) | (0.609) | (1.776) | (0.478) | (0.281) | (1.426) | (0.544) | (4.508) | (0.871) | (9.108) | (2.734) | (32.427) | (0.735) | (7.642) |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **2.865** | **3.814** | 16.091 | 59.530 | 12.765 | **3.222** | 12.741 | **5.243** | 37.336 | 24.469 | 375.206 | 94.898 | 1482.696 | 19.624 | 304.011 |
| | (0.154) | (0.285) | (7.839) | (31.016) | (6.228) | (0.198) | (1.362) | (0.438) | (5.097) | (12.186) | (210.825) | (51.295) | (781.425) | (9.657) | (162.312) |
| CLL | 4.378 | 5.461 | 7.809 | 20.595 | 6.585 | 5.060 | 17.562 | 7.422 | 47.176 | 11.299 | 100.570 | 31.620 | 335.536 | 9.273 | 81.110 |
| | (0.837) | (1.105) | (2.079) | (6.431) | (1.705) | (1.028) | (4.323) | (1.581) | (12.823) | (2.889) | (35.550) | (10.019) | (130.849) | (2.308) | (28.582) |
| CLF | 3.686 | 4.559 | 6.172 | 15.765 | 5.264 | 4.183 | 13.876 | 6.137 | 36.795 | 8.906 | 74.440 | 23.934 | 247.097 | 7.380 | 59.996 |
| | (0.429) | (0.558) | (0.933) | (2.564) | (0.772) | (0.518) | (2.022) | (0.771) | (5.850) | (1.247) | (13.577) | (3.961) | (49.183) | (1.024) | (11.196) |
| PCL | 3.317 | 4.245 | 6.408 | 17.119 | 5.392 | 3.719 | **12.534** | 5.728 | **32.882** | 9.331 | 87.071 | 26.487 | 306.880 | 7.696 | 69.908 |
| | (0.227) | (0.346) | (1.051) | (3.960) | (0.794) | (0.281) | (1.331) | (0.515) | (4.196) | (1.572) | (28.228) | (6.791) | (109.301) | (1.210) | (21.598) |
| PL | 3.411 | 4.393 | 5.825 | 14.027 | 4.961 | 3.824 | 13.299 | 5.946 | 35.082 | 8.428 | 60.479 | 20.906 | 183.361 | 7.017 | 48.860 |
| | (0.248) | (0.380) | (0.603) | (1.823) | (0.481) | (0.305) | (1.525) | (0.561) | (4.777) | (0.897) | (10.069) | (2.832) | (32.677) | (0.725) | (7.994) |
| PF | 3.407 | 4.380 | **5.742** | **13.661** | **4.904** | 3.818 | 12.989 | 5.913 | 34.031 | **8.266** | **59.268** | **20.409** | **180.951** | **6.896** | **47.747** |
| | (0.246) | (0.376) | (0.584) | (1.729) | (0.465) | (0.303) | (1.440) | (0.553) | (4.463) | (0.853) | (9.709) | (2.674) | (31.611) | (0.690) | (7.743) |

The upper and lower tables present the results for DGPs with $\varphi = 0.2$ and 0.8, respectively. Each column represents a DGP according to table 1.2. The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.7: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 0.1$ with aggregation matrix $\mathbf{C}_2$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **0.250** | **0.352** | 0.779 | 2.525 | 0.641 | **0.289** | 1.355 | **0.496** | 3.647 | 1.196 | 14.446 | 4.043 | 53.490 | 0.961 | 0.982 |
| | (0.009) | (0.028) | (0.235) | (0.915) | (0.186) | (0.013) | (0.243) | (0.048) | (0.781) | (0.359) | (5.968) | (1.558) | (24.565) | (0.295) | (0.292) |
| CLL | 0.273 | 0.364 | 0.520 | 1.456 | 0.437 | 0.309 | 1.244 | 0.499 | 3.449 | 0.770 | 6.958 | 2.235 | 22.434 | 0.630 | 0.779 |
| | (0.012) | (0.028) | (0.064) | (0.273) | (0.055) | (0.016) | (0.143) | (0.042) | (0.550) | (0.101) | (1.674) | (0.445) | (6.049) | (0.083) | (0.127) |
| CLF | 0.273 | 0.363 | **0.516** | 1.385 | **0.434** | 0.309 | 1.228 | 0.498 | 3.283 | 0.760 | 6.605 | 2.114 | 21.333 | **0.624** | **0.773** |
| | (0.012) | (0.028) | (0.062) | (0.229) | (0.054) | (0.016) | (0.137) | (0.042) | (0.475) | (0.096) | (1.393) | (0.367) | (4.972) | (0.079) | (0.123) |
| PCL | 0.315 | 0.393 | 0.693 | 1.926 | 0.583 | 0.360 | **1.202** | 0.522 | **3.085** | 1.016 | 8.936 | 2.828 | 30.924 | 0.833 | 1.259 |
| | (0.011) | (0.023) | (0.175) | (0.590) | (0.134) | (0.015) | (0.129) | (0.037) | (0.423) | (0.260) | (3.364) | (0.853) | (13.670) | (0.212) | (0.198) |
| PL | 0.318 | 0.400 | 0.530 | 1.319 | 0.457 | 0.363 | 1.248 | 0.535 | 3.210 | 0.759 | 5.989 | 1.979 | 18.888 | 0.636 | 1.078 |
| | (0.011) | (0.025) | (0.054) | (0.196) | (0.042) | (0.016) | (0.141) | (0.041) | (0.461) | (0.087) | (1.147) | (0.307) | (4.137) | (0.068) | (0.068) |
| PF | 0.319 | 0.400 | 0.530 | **1.319** | 0.457 | 0.364 | 1.248 | 0.535 | 3.210 | **0.759** | **5.975** | **1.978** | **18.599** | 0.636 | 1.078 |
| | (0.011) | (0.025) | (0.054) | (0.196) | (0.042) | (0.016) | (0.141) | (0.041) | (0.461) | (0.087) | (1.142) | (0.306) | (3.989) | (0.068) | (0.068) |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **0.251** | **0.352** | 0.773 | 2.583 | 0.629 | **0.330** | 2.552 | **0.523** | 4.215 | 1.215 | 15.256 | 4.045 | 55.606 | 3.059 | 3.040 |
| | (0.009) | (0.029) | (0.232) | (0.992) | (0.178) | (0.016) | (0.394) | (0.050) | (1.021) | (0.368) | (6.423) | (1.538) | (26.120) | (0.919) | (0.908) |
| CLL | 0.500 | 0.557 | 0.671 | 1.512 | 0.602 | 0.558 | 1.861 | 0.676 | 3.765 | 0.892 | 7.210 | 2.268 | 22.684 | 2.031 | 2.061 |
| | (0.018) | (0.047) | (0.108) | (0.332) | (0.092) | (0.027) | (0.158) | (0.066) | (0.570) | (0.151) | (1.699) | (0.477) | (6.172) | (0.289) | (0.312) |
| CLF | 0.497 | 0.553 | **0.666** | **1.450** | **0.596** | 0.554 | **1.831** | 0.671 | 3.575 | **0.886** | 6.807 | **2.168** | 21.652 | 1.954 | **2.002** |
| | (0.018) | (0.047) | (0.106) | (0.280) | (0.090) | (0.027) | (0.146) | (0.067) | (0.476) | (0.144) | (1.378) | (0.395) | (5.177) | (0.238) | (0.269) |
| PCL | 0.808 | 0.841 | 1.065 | 2.121 | 0.994 | 0.924 | 2.121 | 0.995 | **3.542** | 1.403 | 9.569 | 3.043 | 31.978 | 2.494 | 2.670 |
| | (0.014) | (0.027) | (0.165) | (0.601) | (0.128) | (0.020) | (0.153) | (0.041) | (0.435) | (0.258) | (3.441) | (0.855) | (13.961) | (0.610) | (0.602) |
| PL | 0.820 | 0.855 | 0.927 | 1.532 | 0.883 | 0.938 | 2.109 | 1.015 | 3.671 | 1.157 | 6.287 | 2.188 | 19.100 | 1.899 | 2.098 |
| | (0.015) | (0.030) | (0.054) | (0.192) | (0.045) | (0.021) | (0.154) | (0.046) | (0.474) | (0.086) | (1.127) | (0.307) | (4.073) | (0.212) | (0.218) |
| PF | 0.820 | 0.855 | 0.927 | 1.532 | 0.883 | 0.938 | 2.109 | 1.015 | 3.671 | 1.157 | **6.272** | 2.188 | **18.804** | **1.899** | 2.098 |
| | (0.015) | (0.030) | (0.054) | (0.192) | (0.045) | (0.021) | (0.154) | (0.046) | (0.474) | (0.086) | (1.121) | (0.307) | (3.935) | (0.212) | (0.218) |

The upper and lower tables present the results for DGPs with $\varphi = 0.2$ and 0.8, respectively. Each column represents a DGP according to table 1.2. The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.8: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 1$ with aggregation matrix $\mathbf{C}_2$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **0.794** | 1.112 | 2.481 | 8.157 | 2.031 | **0.908** | 3.913 | 1.559 | 11.366 | 3.760 | 47.377 | 12.584 | 172.990 | 3.059 | 37.805 |
| | (0.030) | (0.092) | (0.740) | (3.119) | (0.572) | (0.041) | (0.622) | (0.155) | (2.349) | (1.169) | (20.002) | (4.800) | (76.639) | (0.919) | (15.962) |
| **CLL** | 0.801 | 1.104 | 1.661 | 4.611 | 1.368 | 0.914 | 3.877 | 1.545 | 10.924 | 2.466 | 21.908 | 7.058 | 71.039 | 2.031 | 17.720 |
| | (0.030) | (0.076) | (0.223) | (0.894) | (0.176) | (0.039) | (0.498) | (0.129) | (1.823) | (0.346) | (5.184) | (1.353) | (19.098) | (0.289) | (4.246) |
| **CLF** | 0.800 | 1.100 | 1.605 | 4.362 | 1.333 | 0.913 | 3.712 | 1.535 | 10.385 | 2.368 | 20.838 | 6.681 | 67.981 | 1.954 | 16.829 |
| | (0.030) | (0.075) | (0.187) | (0.745) | (0.149) | (0.039) | (0.431) | (0.126) | (1.533) | (0.290) | (4.332) | (1.091) | (16.251) | (0.238) | (3.579) |
| **PCL** | 0.810 | **1.087** | 2.056 | 5.645 | 1.717 | 0.923 | **3.535** | **1.505** | **9.683** | 3.038 | 29.145 | 8.600 | 98.808 | 2.494 | 23.153 |
| | (0.030) | (0.067) | (0.504) | (1.772) | (0.412) | (0.039) | (0.380) | (0.116) | (1.323) | (0.762) | (11.442) | (2.722) | (41.937) | (0.610) | (8.695) |
| **PL** | 0.817 | 1.116 | 1.571 | 4.080 | 1.305 | 0.935 | 3.689 | 1.549 | 10.175 | 2.313 | 19.051 | 6.249 | 60.808 | 1.899 | 15.278 |
| | (0.030) | (0.075) | (0.172) | (0.614) | (0.134) | (0.041) | (0.431) | (0.127) | (1.501) | (0.270) | (3.820) | (0.949) | (13.538) | (0.212) | (2.944) |
| **PF** | 0.818 | 1.116 | **1.571** | **4.078** | **1.305** | 0.935 | 3.689 | 1.549 | 10.173 | **2.313** | **18.836** | **6.240** | **58.780** | **1.899** | **15.142** |
| | (0.030) | (0.075) | (0.172) | (0.613) | (0.134) | (0.041) | (0.431) | (0.127) | (1.500) | (0.270) | (3.740) | (0.945) | (12.559) | (0.212) | (2.890) |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **0.795** | **1.118** | 2.445 | 8.320 | 2.018 | **0.923** | 4.475 | **1.580** | 11.780 | 3.769 | 47.182 | 12.899 | 171.951 | 3.040 | 36.896 |
| | (0.028) | (0.090) | (0.743) | (3.158) | (0.563) | (0.042) | (0.858) | (0.155) | (2.682) | (1.140) | (19.593) | (5.017) | (76.603) | (0.908) | (14.953) |
| **CLL** | 0.905 | 1.185 | 1.691 | 4.639 | 1.418 | 1.024 | 4.116 | 1.613 | 11.118 | 2.516 | 22.259 | 7.155 | 71.403 | 2.061 | 17.990 |
| | (0.038) | (0.098) | (0.246) | (0.930) | (0.212) | (0.054) | (0.508) | (0.146) | (1.809) | (0.399) | (5.209) | (1.413) | (19.441) | (0.312) | (4.084) |
| **CLF** | 0.903 | 1.180 | **1.648** | 4.399 | **1.399** | 1.021 | 3.949 | 1.604 | 10.502 | **2.429** | 21.026 | 6.773 | 68.259 | **2.002** | 16.985 |
| | (0.038) | (0.097) | (0.209) | (0.751) | (0.188) | (0.054) | (0.436) | (0.142) | (1.535) | (0.330) | (4.357) | (1.151) | (16.641) | (0.269) | (3.371) |
| **PCL** | 1.103 | 1.323 | 2.194 | 5.806 | 1.895 | 1.255 | **3.943** | 1.731 | **9.770** | 3.164 | 28.594 | 8.705 | 98.241 | 2.670 | 23.062 |
| | (0.036) | (0.076) | (0.481) | (1.749) | (0.393) | (0.050) | (0.394) | (0.117) | (1.356) | (0.712) | (10.854) | (2.782) | (41.673) | (0.602) | (8.425) |
| **PL** | 1.116 | 1.353 | 1.735 | 4.176 | 1.519 | 1.274 | 4.061 | 1.776 | 10.205 | 2.479 | 19.134 | 6.373 | 60.974 | 2.098 | 15.509 |
| | (0.037) | (0.084) | (0.168) | (0.634) | (0.140) | (0.053) | (0.435) | (0.129) | (1.454) | (0.268) | (3.549) | (1.010) | (13.885) | (0.218) | (2.763) |
| **PF** | 1.116 | 1.353 | 1.735 | **4.173** | 1.519 | 1.274 | 4.061 | 1.776 | 10.203 | 2.479 | **18.934** | **6.363** | **59.007** | 2.098 | **15.382** |
| | (0.037) | (0.084) | (0.168) | (0.632) | (0.140) | (0.053) | (0.435) | (0.129) | (1.454) | (0.268) | (3.495) | (1.006) | (12.882) | (0.218) | (2.719) |

The upper and lower tables present the results for DGPs with $\varphi = 0.2$ and 0.8, respectively. Each column represents a DGP according to table 1.2. The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.9: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 5$ with aggregation matrix $\mathbf{C}_2$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | 1.772 | 2.499 | 5.449 | 18.449 | 4.525 | 2.032 | 8.704 | 3.516 | 25.825 | 8.299 | 104.273 | 28.350 | 383.173 | 6.769 | 6.795 |
| | (0.064) | (0.208) | (1.593) | (6.985) | (1.288) | (0.094) | (1.320) | (0.343) | (5.581) | (2.549) | (44.026) | (10.424) | (167.648) | (2.054) | (2.052) |
| CLL | 1.773 | 2.488 | 3.740 | 10.320 | 3.088 | 2.032 | 8.594 | 3.533 | 24.771 | 5.498 | 49.146 | 15.863 | 161.116 | 4.527 | 4.532 |
| | (0.063) | (0.180) | (0.500) | (2.039) | (0.402) | (0.089) | (1.091) | (0.304) | (3.951) | (0.812) | (11.585) | (3.185) | (44.060) | (0.629) | (0.656) |
| CLF | 1.773 | 2.458 | 3.596 | 9.766 | 2.981 | 2.029 | 8.246 | 3.446 | 23.419 | 5.279 | 46.432 | 14.989 | 152.345 | 4.347 | 4.367 |
| | (0.063) | (0.168) | (0.416) | (1.675) | (0.340) | (0.089) | (0.943) | (0.268) | (3.359) | (0.664) | (9.569) | (2.576) | (36.380) | (0.520) | (0.542) |
| PCL | **1.770** | **2.402** | 4.359 | 12.531 | 3.675 | **2.020** | **7.847** | **3.347** | **21.816** | 6.430 | 63.748 | 19.359 | 221.670 | 5.258 | 5.364 |
| | (0.062) | (0.151) | (0.945) | (4.006) | (0.819) | (0.086) | (0.862) | (0.240) | (3.006) | (1.505) | (24.598) | (6.083) | (93.908) | (1.206) | (1.225) |
| PL | 1.783 | 2.468 | 3.500 | 9.168 | 2.915 | 2.044 | 8.190 | 3.459 | 22.803 | 5.121 | 42.994 | 14.076 | 138.351 | 4.210 | 4.302 |
| | (0.064) | (0.170) | (0.388) | (1.437) | (0.314) | (0.090) | (0.959) | (0.272) | (3.337) | (0.621) | (8.397) | (2.292) | (31.881) | (0.465) | (0.470) |
| PF | 1.783 | 2.468 | **3.500** | **9.146** | **2.915** | 2.044 | 8.189 | 3.459 | 22.785 | **5.119** | **41.986** | **14.008** | **132.771** | **4.209** | **4.301** |
| | (0.064) | (0.170) | (0.387) | (1.429) | (0.313) | (0.090) | (0.959) | (0.272) | (3.331) | (0.620) | (7.997) | (2.263) | (28.832) | (0.464) | (0.470) |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **1.776** | **2.487** | 5.522 | 18.232 | 4.496 | **2.040** | 9.027 | 3.481 | 25.603 | 8.424 | 102.861 | 28.329 | 384.419 | 9.575 | 9.557 |
| | (0.062) | (0.197) | (1.705) | (6.883) | (1.235) | (0.095) | (1.563) | (0.341) | (5.547) | (2.654) | (45.064) | (10.586) | (175.072) | (2.851) | (2.806) |
| CLL | 1.828 | 2.514 | 3.721 | 10.304 | 3.092 | 2.087 | 8.730 | 3.514 | 24.713 | 5.569 | 48.838 | 15.979 | 160.465 | 6.434 | 6.399 |
| | (0.067) | (0.190) | (0.498) | (2.033) | (0.400) | (0.102) | (1.131) | (0.311) | (4.094) | (0.825) | (11.661) | (3.089) | (44.769) | (0.897) | (0.899) |
| CLF | 1.826 | 2.487 | 3.583 | 9.762 | **2.993** | 2.083 | 8.366 | 3.441 | 23.328 | 5.332 | 46.212 | 15.082 | 151.584 | 6.163 | 6.152 |
| | (0.067) | (0.178) | (0.416) | (1.658) | (0.338) | (0.102) | (0.971) | (0.280) | (3.441) | (0.662) | (9.638) | (2.498) | (36.610) | (0.730) | (0.760) |
| PCL | 1.927 | 2.513 | 4.467 | 12.520 | 3.746 | 2.195 | **8.033** | **3.420** | **21.732** | 6.516 | 63.392 | 19.297 | 220.483 | 7.409 | 7.464 |
| | (0.066) | (0.157) | (1.002) | (4.025) | (0.759) | (0.098) | (0.873) | (0.253) | (2.984) | (1.515) | (24.063) | (5.997) | (98.903) | (1.670) | (1.576) |
| PL | 1.944 | 2.581 | 3.548 | 9.230 | 2.996 | 2.223 | 8.352 | 3.525 | 22.792 | 5.215 | 42.959 | 14.179 | 137.539 | 5.974 | 6.025 |
| | (0.070) | (0.175) | (0.372) | (1.469) | (0.296) | (0.102) | (0.971) | (0.282) | (3.291) | (0.592) | (8.589) | (2.213) | (30.962) | (0.692) | (0.676) |
| PF | 1.944 | 2.581 | **3.547** | **9.207** | 2.996 | 2.223 | 8.352 | 3.525 | 22.774 | **5.213** | **41.965** | **14.117** | **131.469** | **5.971** | **6.023** |
| | (0.070) | (0.175) | (0.371) | (1.459) | (0.296) | (0.102) | (0.971) | (0.282) | (3.286) | (0.591) | (8.085) | (2.196) | (27.894) | (0.691) | (0.675) |

The upper and lower tables present the results for DGPs with $\varphi = 0.2$ and 0.8, respectively. Each column represents a DGP according to table 1.2. The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.10: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 10$ with aggregation matrix $\mathbf{C}_2$

| | 5 | 6 | 13 | 14 | 19 | 29 | 30 | 31 | 32 | 45 | 46 | 47 | 48 | 57 | 58 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | 2.507 | 3.539 | 7.824 | 25.542 | 6.338 | 2.882 | 12.307 | 4.924 | 36.608 | 11.841 | 144.492 | 39.876 | 535.446 | 9.575 | 116.858 |
| | (0.092) | (0.296) | (2.286) | (9.421) | (1.775) | (0.131) | (1.994) | (0.452) | (7.680) | (3.489) | (61.102) | (15.351) | (248.153) | (2.851) | (48.113) |
| **CLL** | 2.507 | 3.537 | 5.239 | 14.554 | 4.367 | 2.884 | 12.141 | 4.974 | 34.861 | 7.752 | 69.085 | 22.497 | 225.402 | 6.434 | 55.569 |
| | (0.091) | (0.263) | (0.698) | (2.850) | (0.543) | (0.126) | (1.531) | (0.428) | (5.702) | (1.122) | (15.859) | (4.563) | (58.453) | (0.897) | (13.358) |
| **CLF** | 2.505 | 3.475 | 5.053 | 13.805 | 4.206 | 2.877 | 11.649 | 4.851 | 33.069 | 7.432 | 65.384 | 21.183 | 214.439 | 6.163 | 52.655 |
| | (0.090) | (0.236) | (0.589) | (2.386) | (0.458) | (0.124) | (1.309) | (0.384) | (4.800) | (0.922) | (13.146) | (3.620) | (50.318) | (0.730) | (11.170) |
| **PCL** | **2.497** | **3.397** | 6.084 | 17.498 | 5.096 | **2.857** | **11.093** | **4.723** | **30.786** | 9.027 | 89.593 | 26.792 | 302.393 | 7.409 | 72.545 |
| | (0.089) | (0.214) | (1.330) | (5.353) | (1.102) | (0.120) | (1.176) | (0.352) | (4.325) | (2.111) | (33.468) | (8.648) | (133.069) | (1.670) | (26.654) |
| **PL** | 2.514 | 3.484 | 4.919 | 13.008 | 4.114 | 2.891 | 11.564 | 4.881 | 32.160 | 7.218 | 61.113 | 19.943 | 194.459 | 5.974 | 48.765 |
| | (0.091) | (0.240) | (0.563) | (2.029) | (0.427) | (0.127) | (1.295) | (0.395) | (4.823) | (0.875) | (11.786) | (3.138) | (42.523) | (0.692) | (10.173) |
| **PF** | 2.514 | 3.484 | **4.917** | **12.955** | **4.113** | 2.891 | 11.563 | 4.881 | 32.115 | **7.213** | **59.276** | **19.801** | **186.184** | **5.971** | **47.516** |
| | (0.091) | (0.240) | (0.563) | (2.009) | (0.426) | (0.127) | (1.294) | (0.395) | (4.804) | (0.873) | (11.059) | (3.065) | (39.389) | (0.691) | (9.584) |

| | 7 | 8 | 15 | 16 | 20 | 33 | 34 | 35 | 36 | 49 | 50 | 51 | 52 | 59 | 60 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.515** | 3.516 | 7.787 | 25.714 | 6.398 | **2.879** | 12.549 | 4.936 | 36.293 | 11.995 | 146.923 | 40.520 | 547.736 | 9.557 | 117.372 |
| | (0.089) | (0.282) | (2.321) | (9.640) | (1.877) | (0.132) | (1.969) | (0.481) | (8.057) | (3.862) | (63.184) | (14.891) | (246.160) | (2.806) | (47.245) |
| **CLL** | 2.551 | 3.551 | 5.281 | 14.766 | 4.408 | 2.911 | 12.353 | 4.978 | 34.633 | 7.899 | 69.750 | 22.365 | 225.422 | 6.399 | 55.847 |
| | (0.094) | (0.263) | (0.758) | (3.078) | (0.610) | (0.134) | (1.614) | (0.439) | (5.651) | (1.202) | (16.038) | (4.622) | (59.690) | (0.899) | (12.825) |
| **CLF** | 2.548 | 3.492 | 5.068 | 13.940 | 4.240 | 2.904 | 11.819 | 4.855 | 32.805 | 7.574 | 66.111 | 21.119 | 214.261 | 6.152 | 53.115 |
| | (0.094) | (0.236) | (0.616) | (2.483) | (0.494) | (0.132) | (1.393) | (0.398) | (4.782) | (0.990) | (13.531) | (3.676) | (50.795) | (0.760) | (10.746) |
| **PCL** | 2.614 | **3.469** | 6.165 | 17.680 | 5.146 | 2.973 | **11.279** | **4.780** | **30.460** | 9.168 | 90.537 | 27.176 | 314.050 | 7.464 | 72.625 |
| | (0.093) | (0.211) | (1.349) | (5.424) | (1.079) | (0.129) | (1.285) | (0.369) | (4.375) | (2.170) | (34.972) | (8.697) | (131.633) | (1.576) | (27.447) |
| **PL** | 2.635 | 3.564 | 4.982 | 13.130 | 4.176 | 3.009 | 11.740 | 4.927 | 31.860 | 7.376 | 61.281 | 19.961 | 196.006 | 6.025 | 48.781 |
| | (0.097) | (0.232) | (0.554) | (2.134) | (0.425) | (0.134) | (1.398) | (0.408) | (4.730) | (0.859) | (11.732) | (3.288) | (44.860) | (0.676) | (9.301) |
| **PF** | 2.635 | 3.564 | **4.981** | **13.076** | **4.175** | 3.009 | 11.739 | 4.927 | 31.817 | **7.371** | **59.529** | **19.814** | **187.120** | **6.023** | **47.656** |
| | (0.097) | (0.232) | (0.553) | (2.112) | (0.425) | (0.134) | (1.397) | (0.408) | (4.717) | (0.857) | (11.008) | (3.234) | (40.557) | (0.675) | (8.851) |

The upper and lower tables present the results for DGPs with $\varphi = 0.2$ and 0.8, respectively. Each column represents a DGP according to table 1.2. The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.11: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 1$ and $\varphi = 0.2$ with aggregation matrix $\mathbf{C}_1$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.871** | **3.813** | 15.351 | 59.098 | 12.013 | **3.225** | 12.506 | **5.237** | 37.121 | 23.164 | 345.239 | 89.005 | 1415.381 | 17.921 | 275.293 |
| | (0.163) | (0.278) | (7.787) | (31.036) | (5.774) | (0.189) | (1.416) | (0.444) | (5.053) | (11.378) | (191.233) | (45.782) | (789.588) | (8.590) | (144.075) |
| **CLL** | 3.592 | 4.863 | 7.254 | 19.964 | 6.058 | 4.054 | 16.266 | 6.723 | 46.035 | 10.658 | 97.441 | 30.070 | 321.774 | 8.650 | 75.220 |
| | (0.517) | (0.815) | (1.692) | (6.071) | (1.324) | (0.543) | (3.799) | (1.158) | (12.040) | (2.606) | (36.498) | (9.576) | (132.287) | (2.030) | (27.846) |
| **CLF** | 3.226 | 4.215 | 5.849 | 15.121 | 4.899 | 3.625 | 13.110 | 5.752 | 35.981 | 8.471 | 70.939 | 22.961 | 233.947 | 6.985 | 56.653 |
| | (0.280) | (0.430) | (0.728) | (2.402) | (0.574) | (0.303) | (1.744) | (0.601) | (5.537) | (1.094) | (13.767) | (3.695) | (46.761) | (0.889) | (10.516) |
| **PCL** | 3.220 | 4.165 | 5.680 | 14.254 | 4.761 | 3.627 | **12.354** | 5.678 | **32.348** | 8.148 | 66.320 | 21.604 | 225.117 | 6.740 | 53.232 |
| | (0.237) | (0.349) | (0.622) | (2.257) | (0.499) | (0.264) | (1.351) | (0.510) | (4.265) | (0.927) | (14.359) | (3.617) | (61.881) | (0.769) | (11.097) |
| **PL** | 3.287 | 4.268 | 5.671 | 13.329 | 4.774 | 3.705 | 12.882 | 5.836 | 33.675 | 8.127 | 56.609 | 20.183 | 169.271 | 6.731 | 45.609 |
| | (0.249) | (0.375) | (0.603) | (1.719) | (0.482) | (0.280) | (1.489) | (0.547) | (4.543) | (0.860) | (8.760) | (2.565) | (29.936) | (0.733) | (7.348) |
| **PF** | 3.284 | 4.257 | **5.598** | **13.052** | **4.722** | 3.700 | 12.639 | 5.806 | 32.950 | **7.987** | **55.641** | **19.792** | **167.448** | **6.636** | **44.907** |
| | (0.248) | (0.372) | (0.581) | (1.658) | (0.467) | (0.279) | (1.414) | (0.539) | (4.375) | (0.829) | (8.443) | (2.489) | (28.785) | (0.710) | (7.140) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.873** | **4.152** | 43.004 | 177.361 | 33.774 | **3.230** | **15.241** | **5.772** | 50.112 | 67.149 | 1039.918 | 259.334 | 4343.040 | 50.097 | 856.448 |
| | (0.627) | (1.411) | (32.948) | (132.081) | (25.033) | (0.803) | (7.412) | (2.114) | (29.164) | (48.820) | (798.639) | (196.385) | (3263.325) | (38.261) | (603.817) |
| **CLL** | 5.494 | 7.332 | 11.703 | 35.624 | 9.228 | 5.089 | 25.371 | 9.482 | 76.871 | 17.051 | 180.655 | 54.397 | 651.026 | 13.802 | 149.875 |
| | (2.544) | (3.789) | (6.916) | (22.937) | (5.267) | (2.433) | (14.801) | (5.181) | (48.446) | (9.886) | (117.088) | (34.777) | (432.520) | (8.214) | (100.089) |
| **CLF** | 4.359 | 5.837 | **9.733** | 30.359 | **7.664** | 4.327 | 19.584 | 7.616 | 58.934 | **14.259** | 161.351 | 46.632 | 604.840 | **11.813** | 133.352 |
| | (1.797) | (2.770) | (5.530) | (19.125) | (4.260) | (1.838) | (10.870) | (3.811) | (36.437) | (8.234) | (108.983) | (29.988) | (438.496) | (6.829) | (90.227) |
| **PCL** | 3.324 | 4.683 | 25.942 | 104.806 | 20.321 | 3.668 | 16.245 | 6.356 | **48.747** | 40.626 | 606.996 | 153.383 | 2538.333 | 30.341 | 502.018 |
| | (1.059) | (1.926) | (19.319) | (76.734) | (14.714) | (1.249) | (8.425) | (2.721) | (30.587) | (28.721) | (468.732) | (114.807) | (1912.935) | (22.632) | (355.710) |
| **PL** | 4.111 | 5.965 | 10.597 | 32.349 | 8.117 | 4.608 | 23.378 | 8.177 | 73.891 | 15.622 | 158.057 | 48.218 | 562.277 | 12.778 | 135.384 |
| | (1.892) | (3.154) | (6.380) | (21.710) | (4.786) | (2.187) | (14.347) | (4.369) | (51.431) | (9.505) | (111.802) | (32.991) | (410.817) | (7.803) | (95.265) |
| **PF** | 4.102 | 5.923 | 10.069 | **29.898** | 7.861 | 4.593 | 21.173 | 8.056 | 64.225 | 14.619 | **149.310** | **44.763** | **545.485** | 12.097 | **126.373** |
| | (1.884) | (3.116) | (5.932) | (19.414) | (4.558) | (2.172) | (12.466) | (4.267) | (43.417) | (8.707) | (105.094) | (30.210) | (403.948) | (7.196) | (87.215) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.12: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 1$ and $\varphi = 0.8$ with aggregation matrix $\mathbf{C}_1$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.865** | **3.802** | 15.031 | 55.591 | 11.973 | **3.222** | 12.619 | **5.222** | 36.893 | 22.964 | 349.056 | 88.084 | 1374.329 | 18.413 | 283.380 |
| | (0.157) | (0.289) | (7.222) | (28.800) | (5.745) | (0.202) | (1.388) | (0.442) | (5.167) | (11.247) | (194.668) | (47.089) | (720.737) | (8.974) | (150.564) |
| **CLL** | 4.336 | 5.390 | 7.655 | 20.083 | 6.466 | 5.007 | 17.192 | 7.320 | 46.098 | 11.067 | 97.550 | 30.750 | 323.553 | 9.078 | 78.484 |
| | (0.827) | (1.095) | (2.075) | (6.414) | (1.689) | (1.014) | (4.262) | (1.563) | (12.733) | (2.881) | (35.577) | (9.908) | (130.653) | (2.290) | (28.620) |
| **CLF** | 3.660 | 4.510 | 6.038 | 15.298 | 5.161 | 4.153 | 13.611 | 6.068 | 36.004 | 8.707 | 71.257 | 23.075 | 233.560 | 7.206 | 57.229 |
| | (0.427) | (0.553) | (0.930) | (2.536) | (0.766) | (0.513) | (2.004) | (0.765) | (5.842) | (1.234) | (13.389) | (3.834) | (47.173) | (1.015) | (10.875) |
| **PCL** | 3.317 | 4.228 | 5.645 | 14.118 | 4.820 | 3.720 | **12.373** | 5.701 | **32.303** | 8.218 | 66.639 | 21.343 | 222.136 | 6.800 | 53.690 |
| | (0.232) | (0.346) | (0.630) | (2.166) | (0.471) | (0.287) | (1.347) | (0.521) | (4.255) | (0.944) | (15.008) | (3.493) | (55.579) | (0.730) | (11.174) |
| **PL** | 3.384 | 4.334 | 5.659 | 13.386 | 4.833 | 3.800 | 12.908 | 5.857 | 33.675 | 8.170 | 56.803 | 19.854 | 169.276 | 6.798 | 45.700 |
| | (0.246) | (0.369) | (0.583) | (1.727) | (0.462) | (0.305) | (1.488) | (0.554) | (4.582) | (0.856) | (9.357) | (2.651) | (30.195) | (0.696) | (7.245) |
| **PF** | 3.381 | 4.322 | **5.587** | **13.118** | **4.784** | 3.795 | 12.661 | 5.827 | 32.928 | **8.037** | **55.835** | **19.474** | **167.314** | **6.699** | **44.843** |
| | (0.245) | (0.366) | (0.567) | (1.643) | (0.450) | (0.303) | (1.420) | (0.547) | (4.384) | (0.828) | (9.064) | (2.526) | (29.163) | (0.672) | (7.009) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.873** | **4.152** | 44.040 | 163.380 | 33.643 | **3.223** | **15.963** | **5.813** | 49.012 | 64.160 | 1064.734 | 274.578 | 4340.173 | 51.557 | 848.033 |
| | (0.627) | (1.411) | (31.894) | (123.237) | (25.297) | (0.764) | (7.724) | (2.306) | (28.825) | (48.986) | (826.948) | (206.547) | (3182.063) | (37.962) | (645.398) |
| **CLL** | 5.494 | 7.332 | 11.864 | 34.102 | 9.711 | 6.477 | 27.327 | 10.119 | 75.611 | 17.407 | 180.207 | 54.549 | 651.528 | 14.414 | 150.338 |
| | (2.544) | (3.789) | (6.821) | (21.350) | (5.848) | (2.949) | (15.762) | (5.411) | (47.960) | (10.115) | (115.821) | (35.259) | (421.007) | (8.622) | (98.363) |
| **CLF** | 4.359 | 5.837 | **9.699** | 28.091 | **7.991** | 4.985 | 20.866 | 7.933 | 57.657 | **14.147** | 158.379 | 46.597 | 604.041 | **11.978** | 132.953 |
| | (1.797) | (2.770) | (5.589) | (18.230) | (4.615) | (2.037) | (11.575) | (3.866) | (36.783) | (8.061) | (106.720) | (30.262) | (404.411) | (7.074) | (91.123) |
| **PCL** | 3.324 | 4.683 | 26.530 | 96.247 | 20.478 | 3.681 | 16.784 | 6.448 | **48.160** | 38.686 | 625.834 | 162.124 | 2541.434 | 31.332 | 497.564 |
| | (1.059) | (1.926) | (18.587) | (72.313) | (14.813) | (1.170) | (8.879) | (2.812) | (29.877) | (28.683) | (478.240) | (120.487) | (1860.550) | (22.390) | (378.037) |
| **PL** | 4.111 | 5.965 | 10.201 | 30.934 | 8.322 | 4.458 | 23.624 | 8.291 | 72.165 | 15.222 | 157.410 | 48.641 | 554.770 | 12.774 | 132.170 |
| | (1.892) | (3.154) | (6.131) | (20.224) | (4.872) | (1.923) | (14.639) | (4.386) | (49.940) | (9.481) | (109.686) | (32.428) | (385.011) | (7.946) | (92.187) |
| **PF** | 4.102 | 5.923 | 9.817 | **27.996** | 8.065 | 4.447 | 21.635 | 8.174 | 63.113 | 14.306 | **149.766** | **45.074** | **540.537** | 12.102 | **124.324** |
| | (1.884) | (3.116) | (5.803) | (18.350) | (4.656) | (1.914) | (12.876) | (4.272) | (42.316) | (8.662) | (101.894) | (29.774) | (365.648) | (7.366) | (87.941) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.13: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 5$ and $\varphi = 0.2$ with aggregation matrix $\mathbf{C}_1$

| **INTERPOLATION** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| **CLCL** | **2.024** | **2.687** | 10.559 | 40.112 | 8.735 | **2.281** | 8.889 | **3.701** | 26.290 | 16.703 | 245.106 | 63.568 | 955.957 | 13.310 | 13.194 |
| | (0.113) | (0.197) | (5.196) | (21.094) | (4.258) | (0.136) | (1.053) | (0.299) | (3.577) | (8.435) | (133.125) | (33.754) | (497.320) | (6.513) | (6.521) |
| **CLL** | 2.508 | 3.393 | 5.055 | 13.892 | 4.226 | 2.859 | 11.551 | 4.756 | 32.414 | 7.485 | 69.204 | 21.631 | 224.445 | 6.135 | 6.909 |
| | (0.358) | (0.590) | (1.116) | (4.455) | (0.960) | (0.420) | (2.627) | (0.814) | (8.706) | (1.837) | (25.322) | (7.020) | (95.410) | (1.495) | (1.981) |
| **CLF** | 2.292 | 2.986 | 4.114 | 10.621 | 3.451 | 2.588 | 9.311 | 4.085 | 25.454 | 5.970 | 50.417 | 16.368 | 165.000 | 4.938 | 5.346 |
| | (0.206) | (0.306) | (0.491) | (1.809) | (0.419) | (0.241) | (1.267) | (0.419) | (3.865) | (0.793) | (9.654) | (2.806) | (35.410) | (0.652) | (0.870) |
| **PCL** | 2.258 | 2.940 | 3.980 | 9.972 | 3.353 | 2.558 | **8.790** | 4.013 | **22.984** | 5.792 | 47.086 | 15.298 | 156.277 | 4.775 | 4.858 |
| | (0.166) | (0.243) | (0.424) | (1.522) | (0.354) | (0.204) | (0.992) | (0.349) | (3.005) | (0.689) | (10.254) | (2.508) | (39.454) | (0.547) | (0.538) |
| **PL** | 2.321 | 3.009 | 3.988 | 9.396 | 3.350 | 2.620 | 9.130 | 4.118 | 23.908 | 5.745 | 40.359 | 14.292 | 120.119 | 4.753 | 4.828 |
| | (0.170) | (0.259) | (0.407) | (1.212) | (0.348) | (0.212) | (1.062) | (0.370) | (3.219) | (0.643) | (6.712) | (1.914) | (21.262) | (0.502) | (0.512) |
| **PF** | 2.320 | 3.005 | **3.949** | **9.215** | **3.324** | 2.618 | 8.997 | 4.106 | 23.406 | **5.664** | **39.667** | **13.999** | **118.871** | **4.697** | **4.774** |
| | (0.170) | (0.258) | (0.395) | (1.166) | (0.338) | (0.211) | (1.034) | (0.368) | (3.091) | (0.626) | (6.428) | (1.852) | (21.025) | (0.488) | (0.496) |
| **EXTRAPOLATION** | | | | | | | | | | | | | | | |
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| **CLCL** | **2.025** | **2.882** | 30.516 | 118.915 | 24.713 | **2.311** | **11.122** | **4.163** | 35.006 | 48.322 | 716.182 | 188.333 | 2937.658 | 38.524 | 38.161 |
| | (0.423) | (0.931) | (22.275) | (87.785) | (18.447) | (0.564) | (5.443) | (1.603) | (20.236) | (36.230) | (553.026) | (141.838) | (2162.243) | (28.311) | (28.539) |
| **CLL** | 3.066 | 4.531 | 7.775 | 24.487 | 6.632 | 3.582 | 17.945 | 6.609 | 53.311 | 12.365 | 123.772 | 37.469 | 467.518 | 9.716 | 10.803 |
| | (1.365) | (2.274) | (4.428) | (16.315) | (3.820) | (1.669) | (10.885) | (3.585) | (33.043) | (7.695) | (82.565) | (24.293) | (314.950) | (5.886) | (6.649) |
| **CLF** | 2.715 | 3.793 | **6.618** | 21.440 | **5.557** | 3.104 | 13.797 | 5.442 | 41.331 | **10.244** | 110.705 | 32.465 | 433.140 | **8.302** | 8.572 |
| | (1.040) | (1.749) | (3.749) | (14.009) | (3.103) | (1.267) | (8.071) | (2.703) | (23.820) | (6.052) | (74.589) | (21.373) | (304.558) | (5.046) | (5.208) |
| **PCL** | 2.253 | 3.200 | 18.467 | 70.660 | 14.973 | 2.595 | 11.622 | 4.578 | **34.450** | 28.965 | 418.764 | 111.536 | 1720.759 | 23.300 | 22.881 |
| | (0.642) | (1.280) | (12.951) | (51.647) | (10.730) | (0.841) | (6.264) | (2.036) | (19.939) | (20.999) | (319.821) | (83.314) | (1260.260) | (16.925) | (16.560) |
| **PL** | 2.877 | 4.068 | 7.094 | 22.831 | 6.001 | 3.274 | 15.980 | 5.874 | 51.871 | 10.732 | 109.209 | 33.888 | 408.065 | 8.811 | 8.794 |
| | (1.223) | (2.101) | (4.150) | (15.069) | (3.456) | (1.518) | (10.132) | (3.168) | (34.147) | (6.631) | (74.445) | (22.530) | (287.379) | (5.449) | (5.443) |
| **PF** | 2.874 | 4.054 | 6.920 | **21.101** | 5.886 | 3.269 | 15.005 | 5.832 | 45.368 | 10.267 | **101.898** | **31.571** | **393.679** | 8.538 | **8.490** |
| | (1.220) | (2.088) | (3.992) | (13.948) | (3.371) | (1.512) | (9.324) | (3.128) | (28.347) | (6.156) | (69.877) | (20.679) | (278.570) | (5.222) | (5.183) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.14: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 5$ and $\varphi = 0.8$ with aggregation matrix $\mathbf{C}_1$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.023** | **2.700** | 10.675 | 40.893 | 8.657 | **2.286** | 9.037 | **3.721** | 26.179 | 16.584 | 251.364 | 62.744 | 968.297 | 17.921 | 18.413 |
| | (0.116) | (0.198) | (5.371) | (22.242) | (4.250) | (0.138) | (1.007) | (0.309) | (3.771) | (8.403) | (131.466) | (33.946) | (546.187) | (8.590) | (8.974) |
| **CLL** | 3.476 | 4.239 | 5.683 | 14.301 | 4.921 | 4.097 | 12.823 | 5.696 | 33.253 | 8.179 | 69.476 | 22.155 | 227.207 | 8.650 | 9.078 |
| | (0.697) | (1.004) | (1.581) | (4.528) | (1.307) | (0.835) | (3.414) | (1.361) | (9.271) | (2.323) | (26.143) | (7.054) | (91.872) | (2.030) | (2.290) |
| **CLF** | 2.886 | 3.477 | 4.455 | 10.811 | 3.849 | 3.307 | 10.047 | 4.610 | 25.680 | 6.351 | 50.750 | 16.593 | 164.510 | 6.985 | 7.206 |
| | (0.366) | (0.501) | (0.706) | (1.831) | (0.592) | (0.433) | (1.601) | (0.666) | (4.182) | (1.032) | (9.963) | (2.820) | (33.718) | (0.889) | (1.015) |
| **PCL** | 2.398 | 3.063 | 4.094 | 10.089 | 3.455 | 2.703 | **8.820** | 4.130 | **22.908** | 5.837 | 47.505 | 15.270 | 156.970 | 6.740 | 6.800 |
| | (0.182) | (0.243) | (0.446) | (1.634) | (0.338) | (0.205) | (0.946) | (0.357) | (3.150) | (0.652) | (9.840) | (2.514) | (43.441) | (0.769) | (0.730) |
| **PL** | 2.460 | 3.134 | 4.086 | 9.521 | 3.450 | 2.763 | 9.140 | 4.237 | 23.929 | 5.812 | 40.450 | 14.347 | 119.650 | 6.731 | 6.798 |
| | (0.186) | (0.259) | (0.420) | (1.269) | (0.329) | (0.214) | (1.012) | (0.381) | (3.304) | (0.619) | (6.532) | (1.926) | (21.495) | (0.733) | (0.696) |
| **PF** | 2.458 | 3.130 | **4.050** | **9.328** | **3.425** | 2.760 | 9.016 | 4.225 | 23.381 | **5.730** | **39.743** | **14.055** | **118.183** | **6.636** | **6.699** |
| | (0.186) | (0.257) | (0.410) | (1.215) | (0.322) | (0.213) | (0.983) | (0.378) | (3.207) | (0.600) | (6.282) | (1.855) | (21.064) | (0.710) | (0.672) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.005** | **2.926** | 30.521 | 119.607 | 24.279 | **2.327** | **10.975** | **4.197** | 35.837 | 47.421 | 773.489 | 185.978 | 2970.516 | 50.097 | 51.557 |
| | (0.438) | (1.007) | (23.411) | (91.239) | (18.290) | (0.534) | (5.807) | (1.576) | (21.702) | (35.953) | (569.122) | (144.973) | (2236.619) | (38.261) | (37.962) |
| **CLL** | 4.337 | 5.549 | 8.682 | 25.221 | 7.243 | 5.294 | 20.159 | 7.970 | 56.215 | 12.635 | 132.675 | 39.719 | 439.259 | 13.802 | 14.414 |
| | (1.934) | (2.846) | (4.943) | (15.833) | (4.113) | (2.307) | (12.008) | (4.334) | (35.000) | (7.602) | (87.015) | (25.384) | (292.107) | (8.214) | (8.622) |
| **CLF** | 3.406 | 4.321 | 7.125 | 21.729 | 5.825 | 4.019 | 14.992 | 6.075 | 43.395 | 10.329 | 117.797 | 32.996 | 405.983 | **11.813** | **11.978** |
| | (1.334) | (1.983) | (3.997) | (13.930) | (3.194) | (1.576) | (8.558) | (3.006) | (26.288) | (6.089) | (76.995) | (22.582) | (287.697) | (6.829) | (7.074) |
| **PCL** | 2.341 | 3.292 | 18.421 | 70.912 | 14.691 | 2.699 | 11.623 | 4.703 | **35.563** | 28.374 | 452.904 | 110.130 | 1735.528 | 30.341 | 31.332 |
| | (0.716) | (1.296) | (13.777) | (53.460) | (10.697) | (0.826) | (6.507) | (2.003) | (22.222) | (21.025) | (335.187) | (84.803) | (1304.200) | (22.632) | (22.390) |
| **PL** | 2.890 | 4.120 | 7.066 | 22.595 | 5.614 | 3.270 | 16.254 | 5.996 | 53.442 | 10.541 | 116.967 | 34.552 | 387.402 | 12.778 | 12.774 |
| | (1.309) | (2.057) | (4.096) | (15.099) | (3.346) | (1.407) | (9.873) | (3.188) | (34.926) | (6.494) | (79.684) | (24.500) | (277.571) | (7.803) | (7.946) |
| **PF** | 2.887 | 4.106 | **6.918** | **21.307** | **5.522** | 3.266 | 15.297 | 5.952 | 46.965 | **10.088** | **110.008** | **31.726** | **370.893** | 12.097 | 12.102 |
| | (1.305) | (2.043) | (3.952) | (13.818) | (3.279) | (1.402) | (9.153) | (3.145) | (29.746) | (6.059) | (73.121) | (22.607) | (268.563) | (7.196) | (7.366) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.15: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 10$ and $\varphi = 0.2$ with aggregation matrix $\mathbf{C}_1$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.871** | **3.813** | 15.351 | 59.098 | 12.013 | **3.225** | 12.506 | **5.237** | 37.121 | 23.164 | 345.239 | 89.005 | 1415.381 | 17.921 | 275.293 |
| | (0.163) | (0.278) | (7.787) | (31.036) | (5.774) | (0.189) | (1.416) | (0.444) | (5.053) | (11.378) | (191.233) | (45.782) | (789.588) | (8.590) | (144.075) |
| **CLL** | 3.592 | 4.863 | 7.254 | 19.964 | 6.058 | 4.054 | 16.266 | 6.723 | 46.035 | 10.658 | 97.441 | 30.070 | 321.774 | 8.650 | 75.220 |
| | (0.517) | (0.815) | (1.692) | (6.071) | (1.324) | (0.543) | (3.799) | (1.158) | (12.040) | (2.606) | (36.498) | (9.576) | (132.287) | (2.030) | (27.846) |
| **CLF** | 3.226 | 4.215 | 5.849 | 15.121 | 4.899 | 3.625 | 13.110 | 5.752 | 35.981 | 8.471 | 70.939 | 22.961 | 233.947 | 6.985 | 56.653 |
| | (0.280) | (0.430) | (0.728) | (2.402) | (0.574) | (0.303) | (1.744) | (0.601) | (5.537) | (1.094) | (13.767) | (3.695) | (46.761) | (0.889) | (10.516) |
| **PCL** | 3.220 | 4.165 | 5.680 | 14.254 | 4.761 | 3.627 | **12.354** | 5.678 | **32.348** | 8.148 | 66.320 | 21.604 | 225.117 | 6.740 | 53.232 |
| | (0.237) | (0.349) | (0.622) | (2.257) | (0.499) | (0.264) | (1.351) | (0.510) | (4.265) | (0.927) | (14.359) | (3.617) | (61.881) | (0.769) | (11.097) |
| **PL** | 3.287 | 4.268 | 5.671 | 13.329 | 4.774 | 3.705 | 12.882 | 5.836 | 33.675 | 8.127 | 56.609 | 20.183 | 169.271 | 6.731 | 45.609 |
| | (0.249) | (0.375) | (0.603) | (1.719) | (0.482) | (0.280) | (1.489) | (0.547) | (4.543) | (0.860) | (8.760) | (2.565) | (29.936) | (0.733) | (7.348) |
| **PF** | 3.284 | 4.257 | **5.598** | **13.052** | **4.722** | 3.700 | 12.639 | 5.806 | 32.950 | **7.987** | **55.641** | **19.792** | **167.448** | **6.636** | **44.907** |
| | (0.248) | (0.372) | (0.581) | (1.658) | (0.467) | (0.279) | (1.414) | (0.539) | (4.375) | (0.829) | (8.443) | (2.489) | (28.785) | (0.710) | (7.140) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | 2.853 | 4.033 | 43.004 | 177.361 | 33.774 | **3.230** | **15.241** | **5.772** | 50.112 | 67.149 | 1039.918 | 259.334 | 4343.040 | 50.097 | 856.448 |
| | **0.629** | 1.295 | (32.948) | (132.081) | (25.033) | (0.803) | (7.412) | (2.114) | (29.164) | (48.820) | (798.639) | (196.385) | (3263.325) | (38.261) | (603.817) |
| **CLL** | 4.366 | 6.468 | 11.703 | 35.624 | 9.228 | 5.089 | 25.371 | 9.482 | 76.871 | 17.051 | 180.655 | 54.397 | 651.026 | 13.802 | 149.875 |
| | 2.058 | 3.296 | (6.916) | (22.937) | (5.267) | (2.433) | (14.801) | (5.181) | (48.446) | (9.886) | (117.088) | (34.777) | (432.520) | (8.214) | (100.089) |
| **CLF** | 3.775 | 5.300 | **9.733** | 30.359 | **7.664** | 4.327 | 19.584 | 7.616 | 58.934 | **14.259** | 161.351 | 46.632 | 604.840 | **11.813** | 133.352 |
| | 1.533 | 2.376 | (5.530) | (19.125) | (4.260) | (1.838) | (10.870) | (3.811) | (36.437) | (8.234) | (108.983) | (29.988) | (438.496) | (6.829) | (90.227) |
| **PCL** | 3.205 | 4.448 | 25.942 | 104.806 | 20.321 | 3.668 | 16.245 | 6.356 | **48.747** | 40.626 | 606.996 | 153.383 | 2538.333 | 30.341 | 502.018 |
| | 1.007 | 1.663 | (19.319) | (76.734) | (14.714) | (1.249) | (8.425) | (2.721) | (30.587) | (28.721) | (468.732) | (114.807) | (1912.935) | (22.632) | (355.710) |
| **PL** | 3.981 | 5.679 | 10.597 | 32.349 | 8.117 | 4.608 | 23.378 | 8.177 | 73.891 | 15.622 | 158.057 | 48.218 | 562.277 | 12.778 | 135.384 |
| | 1.801 | 2.775 | (6.380) | (21.710) | (4.786) | (2.187) | (14.347) | (4.369) | (51.431) | (9.505) | (111.802) | (32.991) | (410.817) | (7.803) | (95.265) |
| **PF** | 3.973 | 5.643 | 10.069 | **29.898** | 7.861 | 4.593 | 21.173 | 8.056 | 64.225 | 14.619 | **149.310** | **44.763** | **545.485** | 12.097 | **126.373** |
| | 1.792 | 2.734 | (5.932) | (19.414) | (4.558) | (2.172) | (12.466) | (4.267) | (43.417) | (8.707) | (105.094) | (30.210) | (403.948) | (7.196) | (87.215) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.16: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 10$ and $\varphi = 0.8$ with aggregation matrix $\mathbf{C}_1$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.865** | **3.802** | 15.031 | 55.591 | 11.973 | **3.222** | 12.619 | **5.222** | 36.893 | 22.964 | 349.056 | 88.084 | 1374.329 | 18.413 | 283.380 |
| | (0.157) | (0.289) | (7.222) | (28.800) | (5.745) | (0.202) | (1.388) | (0.442) | (5.167) | (11.247) | (194.668) | (47.089) | (720.737) | (8.974) | (150.564) |
| **CLL** | 4.336 | 5.390 | 7.655 | 20.083 | 6.466 | 5.007 | 17.192 | 7.320 | 46.098 | 11.067 | 97.550 | 30.750 | 323.553 | 9.078 | 78.484 |
| | (0.827) | (1.095) | (2.075) | (6.414) | (1.689) | (1.014) | (4.262) | (1.563) | (12.733) | (2.881) | (35.577) | (9.908) | (130.653) | (2.290) | (28.620) |
| **CLF** | 3.660 | 4.510 | 6.038 | 15.298 | 5.161 | 4.153 | 13.611 | 6.068 | 36.004 | 8.707 | 71.257 | 23.075 | 233.560 | 7.206 | 57.229 |
| | (0.427) | (0.553) | (0.930) | (2.536) | (0.766) | (0.513) | (2.004) | (0.765) | (5.842) | (1.234) | (13.389) | (3.834) | (47.173) | (1.015) | (10.875) |
| **PCL** | 3.317 | 4.228 | 5.645 | 14.118 | 4.820 | 3.720 | **12.373** | 5.701 | **32.303** | 8.218 | 66.639 | 21.343 | 222.136 | 6.800 | 53.690 |
| | (0.232) | (0.346) | (0.630) | (2.166) | (0.471) | (0.287) | (1.347) | (0.521) | (4.255) | (0.944) | (15.008) | (3.493) | (55.579) | (0.730) | (11.174) |
| **PL** | 3.384 | 4.334 | 5.659 | 13.386 | 4.833 | 3.800 | 12.908 | 5.857 | 33.675 | 8.170 | 56.803 | 19.854 | 169.276 | 6.798 | 45.700 |
| | (0.246) | (0.369) | (0.583) | (1.727) | (0.462) | (0.305) | (1.488) | (0.554) | (4.582) | (0.856) | (9.357) | (2.651) | (30.195) | (0.696) | (7.245) |
| **PF** | 3.381 | 4.322 | **5.587** | **13.118** | **4.784** | 3.795 | 12.661 | 5.827 | 32.928 | **8.037** | **55.835** | **19.474** | **167.314** | **6.699** | **44.843** |
| | (0.245) | (0.366) | (0.567) | (1.643) | (0.450) | (0.303) | (1.420) | (0.547) | (4.384) | (0.828) | (9.064) | (2.526) | (29.163) | (0.672) | (7.009) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.873** | **4.152** | 44.040 | 163.380 | 33.643 | **3.223** | **15.963** | **5.813** | 49.012 | 64.160 | 1064.734 | 274.578 | 4340.173 | 51.557 | 848.033 |
| | (0.627) | (1.411) | (31.894) | (123.237) | (25.297) | (0.764) | (7.724) | (2.306) | (28.825) | (48.986) | (826.948) | (206.547) | (3182.063) | (37.962) | (645.398) |
| **CLL** | 5.494 | 7.332 | 11.864 | 34.102 | 9.711 | 6.477 | 27.327 | 10.119 | 75.611 | 17.407 | 180.207 | 54.549 | 651.528 | 14.414 | 150.338 |
| | (2.544) | (3.789) | (6.821) | (21.350) | (5.848) | (2.949) | (15.762) | (5.411) | (47.960) | (10.115) | (115.821) | (35.259) | (421.007) | (8.622) | (98.363) |
| **CLF** | 4.359 | 5.837 | **9.699** | 28.091 | **7.991** | 4.985 | 20.866 | 7.933 | 57.657 | **14.147** | 158.379 | 46.597 | 604.041 | **11.978** | 132.953 |
| | (1.797) | (2.770) | (5.589) | (18.230) | (4.615) | (2.037) | (11.575) | (3.866) | (36.783) | (8.061) | (106.720) | (30.262) | (404.411) | (7.074) | (91.123) |
| **PCL** | 3.324 | 4.683 | 26.530 | 96.247 | 20.478 | 3.681 | 16.784 | 6.448 | **48.160** | 38.686 | 625.834 | 162.124 | 2541.434 | 31.332 | 497.564 |
| | (1.059) | (1.926) | (18.587) | (72.313) | (14.813) | (1.170) | (8.879) | (2.812) | (29.877) | (28.683) | (478.240) | (120.487) | (1860.550) | (22.390) | (378.037) |
| **PL** | 4.111 | 5.965 | 10.201 | 30.934 | 8.322 | 4.458 | 23.624 | 8.291 | 72.165 | 15.222 | 157.410 | 48.641 | 554.770 | 12.774 | 132.170 |
| | (1.892) | (3.154) | (6.131) | (20.224) | (4.872) | (1.923) | (14.639) | (4.386) | (49.940) | (9.481) | (109.686) | (32.428) | (385.011) | (7.946) | (92.187) |
| **PF** | 4.102 | 5.923 | 9.817 | **27.996** | 8.065 | 4.447 | 21.635 | 8.174 | 63.113 | 14.306 | **149.766** | **45.074** | **540.537** | 12.102 | **124.324** |
| | (1.884) | (3.116) | (5.803) | (18.350) | (4.656) | (1.914) | (12.876) | (4.272) | (42.316) | (8.662) | (101.894) | (29.774) | (365.648) | (7.366) | (87.941) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.17: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 0.1$ and $\varphi = 0.2$ with aggregation matrix $\mathbf{C}_2$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **0.250** | **0.350** | 0.604 | 1.828 | 0.504 | **0.288** | 1.334 | 0.491 | 3.545 | 0.919 | 10.256 | 2.989 | 36.466 | 0.748 | 0.772 |
| | (0.009) | (0.028) | (0.159) | (0.686) | (0.134) | (0.013) | (0.245) | (0.049) | (0.783) | (0.254) | (4.567) | (1.202) | (18.292) | (0.210) | (0.218) |
| CLL | 0.272 | 0.358 | 0.494 | 1.337 | 0.415 | 0.308 | 1.199 | 0.489 | 3.258 | 0.727 | 6.284 | 2.055 | 19.812 | 0.596 | 0.747 |
| | (0.012) | (0.027) | (0.060) | (0.257) | (0.051) | (0.016) | (0.136) | (0.042) | (0.524) | (0.091) | (1.546) | (0.414) | (5.649) | (0.076) | (0.124) |
| CLF | 0.272 | 0.358 | **0.490** | 1.268 | **0.413** | 0.308 | 1.186 | **0.489** | 3.119 | 0.717 | 5.888 | 1.933 | 18.575 | **0.589** | **0.742** |
| | (0.012) | (0.027) | (0.057) | (0.201) | (0.050) | (0.016) | (0.131) | (0.042) | (0.452) | (0.086) | (1.203) | (0.324) | (4.377) | (0.072) | (0.120) |
| PCL | 0.315 | 0.391 | 0.515 | 1.262 | 0.443 | 0.360 | **1.178** | 0.518 | **2.970** | 0.736 | 5.570 | 1.876 | 17.277 | 0.615 | 1.060 |
| | (0.011) | (0.023) | (0.049) | (0.179) | (0.039) | (0.016) | (0.129) | (0.037) | (0.413) | (0.081) | (1.029) | (0.270) | (3.890) | (0.064) | (0.063) |
| PL | 0.315 | 0.392 | 0.503 | 1.198 | 0.434 | 0.359 | 1.195 | 0.522 | 3.008 | 0.714 | 5.286 | 1.790 | 16.191 | 0.600 | 1.042 |
| | (0.011) | (0.023) | (0.046) | (0.167) | (0.036) | (0.015) | (0.133) | (0.038) | (0.421) | (0.074) | (0.913) | (0.247) | (3.407) | (0.059) | (0.056) |
| PF | 0.315 | 0.392 | 0.503 | **1.197** | 0.434 | 0.359 | 1.195 | 0.522 | 3.008 | **0.714** | **5.272** | **1.790** | **15.889** | 0.600 | 1.042 |
| | (0.011) | (0.023) | (0.046) | (0.166) | (0.036) | (0.015) | (0.133) | (0.038) | (0.421) | (0.074) | (0.906) | (0.247) | (3.223) | (0.059) | (0.056) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **0.260** | **0.427** | 5.405 | 20.907 | 4.257 | **0.305** | 1.913 | **0.625** | 6.327 | 8.485 | 124.914 | 31.829 | 502.386 | 6.594 | 6.515 |
| | (0.048) | (0.140) | (4.045) | (15.081) | (3.100) | (0.066) | (0.913) | (0.224) | (3.406) | (5.950) | (95.170) | (23.560) | (383.878) | (4.897) | (4.674) |
| CLL | 0.300 | 0.509 | 1.203 | 4.585 | 0.996 | 0.350 | 2.427 | 0.749 | 8.479 | 1.887 | 24.727 | 6.996 | 91.584 | **1.531** | 1.616 |
| | (0.061) | (0.201) | (0.756) | (2.968) | (0.594) | (0.090) | (1.442) | (0.357) | (5.312) | (1.216) | (17.341) | (4.727) | (63.519) | (0.913) | (0.984) |
| CLF | 0.300 | 0.509 | **1.201** | **4.467** | **0.995** | 0.350 | 2.339 | 0.746 | 7.598 | **1.879** | 25.528 | **6.888** | 94.064 | 1.532 | **1.603** |
| | (0.061) | (0.200) | (0.756) | (2.992) | (0.594) | (0.090) | (1.372) | (0.354) | (4.598) | (1.208) | (17.690) | (4.695) | (65.629) | (0.911) | (0.971) |
| PCL | 0.311 | 0.452 | 5.398 | 19.438 | 4.258 | 0.360 | **1.836** | 0.639 | **6.111** | 8.397 | 97.695 | 27.937 | 390.757 | 6.574 | 6.521 |
| | (0.056) | (0.143) | (4.030) | (13.242) | (3.094) | (0.078) | (0.908) | (0.229) | (3.401) | (5.850) | (73.974) | (19.109) | (300.747) | (4.859) | (4.614) |
| PL | 0.383 | 0.607 | 1.252 | 4.530 | 1.055 | 0.460 | 2.650 | 0.879 | 8.522 | 1.949 | 24.525 | 6.945 | **90.007** | 1.592 | 2.039 |
| | (0.095) | (0.270) | (0.797) | (2.978) | (0.627) | (0.144) | (1.587) | (0.459) | (5.476) | (1.268) | (17.447) | (4.750) | (62.471) | (0.972) | (1.152) |
| PF | 0.405 | 0.614 | 1.252 | 4.530 | 1.055 | 0.474 | 2.650 | 0.882 | 8.522 | 1.949 | **24.521** | 6.945 | 90.072 | 1.592 | 2.039 |
| | (0.102) | (0.274) | (0.797) | (2.978) | (0.627) | (0.148) | (1.587) | (0.459) | (5.476) | (1.268) | (17.446) | (4.750) | (62.477) | (0.972) | (1.152) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.18: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 0.1$ and $\varphi = 0.8$ with aggregation matrix $\mathbf{C}_2$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **0.250** | **0.349** | **0.611** | 1.895 | **0.495** | **0.329** | 2.540 | **0.518** | 4.129 | 0.940 | 10.711 | 2.978 | 37.608 | 2.351 | 2.358 |
| | (0.009) | (0.029) | (0.167) | (0.746) | (0.126) | (0.016) | (0.398) | (0.051) | (1.031) | (0.269) | (4.908) | (1.169) | (19.691) | (0.639) | (0.639) |
| **CLL** | 0.498 | 0.551 | 0.646 | 1.401 | 0.585 | 0.555 | 1.818 | 0.666 | 3.573 | 0.852 | 6.552 | 2.098 | 19.970 | 1.922 | 1.949 |
| | (0.019) | (0.047) | (0.105) | (0.314) | (0.091) | (0.028) | (0.150) | (0.067) | (0.536) | (0.146) | (1.602) | (0.453) | (5.721) | (0.269) | (0.292) |
| **CLF** | 0.495 | 0.548 | 0.641 | **1.340** | 0.579 | 0.551 | **1.788** | 0.662 | **3.411** | **0.848** | 6.128 | **1.995** | 18.742 | 1.845 | **1.895** |
| | (0.019) | (0.048) | (0.103) | (0.256) | (0.090) | (0.028) | (0.139) | (0.067) | (0.451) | (0.139) | (1.225) | (0.359) | (4.431) | (0.213) | (0.247) |
| **PCL** | 0.810 | 0.842 | 0.913 | 1.473 | 0.872 | 0.927 | 2.096 | 0.995 | 3.439 | 1.137 | 5.912 | 2.092 | 17.588 | 1.828 | 2.030 |
| | (0.014) | (0.027) | (0.050) | (0.171) | (0.040) | (0.020) | (0.153) | (0.042) | (0.426) | (0.079) | (1.031) | (0.269) | (3.736) | (0.192) | (0.199) |
| **PL** | 0.805 | 0.839 | 0.900 | 1.419 | 0.861 | 0.920 | 2.065 | 0.993 | 3.469 | 1.116 | 5.602 | 2.014 | 16.288 | 1.785 | 1.985 |
| | (0.014) | (0.027) | (0.044) | (0.154) | (0.036) | (0.020) | (0.152) | (0.041) | (0.432) | (0.072) | (0.918) | (0.253) | (3.214) | (0.182) | (0.185) |
| **PF** | 0.805 | 0.839 | 0.900 | 1.419 | 0.861 | 0.920 | 2.065 | 0.993 | 3.468 | 1.116 | **5.586** | 2.014 | **15.990** | **1.785** | 1.985 |
| | (0.014) | (0.027) | (0.044) | (0.154) | (0.036) | (0.020) | (0.152) | (0.041) | (0.432) | (0.072) | (0.909) | (0.252) | (3.043) | (0.182) | (0.185) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **0.257** | **0.441** | 5.051 | 20.728 | 4.152 | **0.343** | 2.873 | **0.661** | 6.500 | 8.464 | 135.110 | 32.188 | 530.159 | 21.746 | 21.020 |
| | (0.045) | (0.150) | (3.910) | (15.732) | (3.083) | (0.064) | (1.024) | (0.222) | (3.218) | (6.057) | (97.330) | (23.749) | (395.349) | (16.260) | (15.590) |
| **CLL** | 0.564 | 0.708 | 1.327 | 4.440 | 1.058 | 0.646 | 2.996 | 0.942 | 8.823 | 1.925 | 24.550 | 6.755 | 94.269 | 4.907 | 5.007 |
| | (0.085) | (0.254) | (0.744) | (2.927) | (0.597) | (0.131) | (1.432) | (0.398) | (5.598) | (1.211) | (16.787) | (4.496) | (64.439) | (3.020) | (3.255) |
| **CLF** | 0.560 | 0.702 | **1.320** | **4.357** | **1.051** | 0.638 | 2.942 | 0.930 | 7.888 | **1.910** | 24.726 | **6.727** | 98.390 | **4.841** | **4.835** |
| | (0.084) | (0.251) | (0.739) | (2.904) | (0.593) | (0.130) | (1.365) | (0.390) | (4.796) | (1.195) | (17.079) | (4.430) | (67.024) | (2.876) | (3.065) |
| **PCL** | 0.736 | 0.814 | 5.064 | 19.218 | 4.196 | 0.851 | **2.770** | 0.990 | **6.254** | 8.419 | 105.985 | 28.131 | 411.429 | 20.062 | 19.565 |
| | (0.070) | (0.163) | (3.844) | (13.700) | (3.039) | (0.098) | (0.988) | (0.261) | (3.226) | (5.943) | (76.030) | (19.135) | (309.178) | (13.953) | (13.558) |
| **PL** | 1.203 | 1.268 | 1.659 | 4.519 | 1.469 | 1.419 | 3.293 | 1.609 | 9.016 | 2.243 | 24.365 | 6.786 | 93.240 | 4.917 | 5.070 |
| | (0.138) | (0.416) | (0.886) | (3.014) | (0.761) | (0.208) | (1.377) | (0.639) | (5.648) | (1.355) | (16.706) | (4.483) | (64.907) | (3.048) | (3.308) |
| **PF** | 1.203 | 1.268 | 1.659 | 4.519 | 1.469 | 1.419 | 3.293 | 1.609 | 9.016 | 2.243 | **24.357** | 6.786 | **93.018** | 4.917 | 5.070 |
| | (0.138) | (0.416) | (0.886) | (3.014) | (0.761) | (0.208) | (1.377) | (0.639) | (5.648) | (1.355) | (16.703) | (4.483) | (64.481) | (3.048) | (3.308) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.19: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 1$ and $\varphi = 0.2$ with aggregation matrix $\mathbf{C}_2$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **0.793** | 1.103 | 1.926 | 5.948 | 1.577 | **0.906** | 3.839 | 1.542 | 11.046 | 2.878 | 32.954 | 9.152 | 118.497 | 2.351 | 26.583 |
| | (0.030) | (0.093) | (0.533) | (2.318) | (0.394) | (0.042) | (0.628) | (0.156) | (2.403) | (0.801) | (14.787) | (3.541) | (58.811) | (0.639) | (12.721) |
| **CLL** | 0.798 | 1.085 | 1.573 | 4.253 | 1.299 | 0.908 | 3.719 | 1.512 | 10.286 | 2.333 | 19.649 | 6.507 | 62.766 | 1.922 | 15.948 |
| | (0.030) | (0.074) | (0.206) | (0.836) | (0.163) | (0.040) | (0.467) | (0.124) | (1.710) | (0.325) | (4.839) | (1.274) | (17.716) | (0.269) | (3.951) |
| **CLF** | 0.798 | 1.082 | 1.519 | 4.002 | 1.266 | 0.908 | 3.582 | 1.504 | 9.838 | 2.238 | 18.519 | 6.117 | 59.151 | 1.845 | 14.990 |
| | (0.030) | (0.073) | (0.163) | (0.655) | (0.134) | (0.040) | (0.405) | (0.122) | (1.455) | (0.261) | (3.827) | (0.978) | (14.018) | (0.213) | (3.157) |
| **PCL** | 0.810 | **1.077** | 1.517 | 3.842 | 1.268 | 0.921 | **3.454** | **1.488** | **9.329** | 2.232 | 17.624 | 5.855 | 54.863 | 1.828 | 14.152 |
| | (0.030) | (0.068) | (0.155) | (0.581) | (0.123) | (0.039) | (0.379) | (0.115) | (1.310) | (0.255) | (3.428) | (0.843) | (11.715) | (0.192) | (2.599) |
| **PL** | 0.814 | 1.092 | 1.478 | 3.708 | 1.235 | 0.927 | 3.519 | 1.509 | 9.503 | 2.174 | 16.713 | 5.673 | 52.221 | 1.785 | 13.443 |
| | (0.030) | (0.071) | (0.143) | (0.516) | (0.113) | (0.040) | (0.389) | (0.119) | (1.351) | (0.233) | (3.087) | (0.787) | (10.927) | (0.182) | (2.383) |
| **PF** | 0.814 | 1.092 | **1.478** | **3.705** | **1.235** | 0.927 | 3.519 | 1.509 | 9.501 | **2.174** | **16.498** | **5.664** | **50.271** | **1.785** | **13.305** |
| | (0.030) | (0.071) | (0.143) | (0.515) | (0.113) | (0.040) | (0.389) | (0.119) | (1.350) | (0.233) | (2.979) | (0.782) | (9.702) | (0.182) | (2.313) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **0.829** | **1.364** | 17.130 | 66.392 | 14.005 | **0.965** | 5.844 | 1.992 | 19.827 | 27.013 | 427.684 | 103.075 | 1609.887 | 21.746 | 333.716 |
| | (0.149) | (0.470) | (12.550) | (50.394) | (9.943) | (0.217) | (2.724) | (0.744) | (10.567) | (20.408) | (322.850) | (77.606) | (1185.941) | (16.260) | (247.584) |
| **CLL** | 0.999 | 1.686 | 3.986 | 14.051 | 3.165 | 1.057 | 8.052 | 2.406 | 27.759 | 5.978 | 81.464 | 21.578 | 289.203 | 4.907 | 64.447 |
| | (0.205) | (0.722) | (2.456) | (9.057) | (1.932) | (0.282) | (4.948) | (1.140) | (17.598) | (3.809) | (55.980) | (14.479) | (202.098) | (3.020) | (43.366) |
| **CLF** | 0.994 | 1.661 | **3.894** | **13.857** | **3.122** | 1.055 | 7.134 | 2.345 | 24.810 | **5.778** | 81.983 | 21.568 | 300.801 | **4.841** | 65.326 |
| | (0.204) | (0.698) | (2.424) | (8.911) | (1.900) | (0.280) | (4.213) | (1.081) | (14.864) | (3.671) | (56.485) | (14.204) | (213.837) | (2.876) | (43.813) |
| **PCL** | 1.081 | 1.512 | 16.272 | 53.208 | 13.563 | 0.974 | **5.669** | **1.963** | **19.024** | 24.295 | 332.927 | 80.989 | 1257.589 | 20.062 | 260.499 |
| | (0.182) | (0.481) | (11.345) | (38.871) | (9.379) | (0.218) | (2.779) | (0.742) | (10.810) | (16.848) | (252.104) | (60.487) | (924.251) | (13.953) | (193.253) |
| **PL** | 1.460 | 2.084 | 4.018 | 13.903 | 3.165 | 1.160 | 8.168 | 2.604 | 27.900 | 5.966 | 80.679 | 21.437 | 287.237 | 4.917 | 63.659 |
| | (0.342) | (0.974) | (2.464) | (8.983) | (1.967) | (0.367) | (5.093) | (1.341) | (17.936) | (3.821) | (56.421) | (14.489) | (204.999) | (3.048) | (42.576) |
| **PF** | 1.460 | 2.084 | 4.018 | 13.902 | 3.165 | 1.163 | 8.168 | 2.604 | 27.896 | 5.966 | **80.473** | **21.435** | **283.165** | 4.917 | **63.580** |
| | (0.342) | (0.974) | (2.464) | (8.982) | (1.967) | (0.368) | (5.093) | (1.341) | (17.932) | (3.821) | (56.229) | (14.485) | (202.675) | (3.048) | (42.511) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.20: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 1$ and $\varphi = 0.8$ with aggregation matrix $\mathbf{C}_2$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **0.794** | **1.109** | 1.916 | 6.016 | 1.591 | **0.920** | 4.403 | **1.565** | 11.456 | 2.913 | 33.493 | 9.474 | 118.053 | 2.358 | 25.956 |
| | (0.029) | (0.091) | (0.533) | (2.317) | (0.404) | (0.043) | (0.867) | (0.157) | (2.718) | (0.834) | (15.059) | (3.754) | (58.079) | (0.639) | (11.331) |
| CLL | 0.901 | 1.166 | 1.610 | 4.277 | 1.352 | 1.018 | 3.959 | 1.583 | 10.477 | 2.375 | 19.978 | 6.608 | 62.928 | 1.949 | 16.220 |
| | (0.039) | (0.098) | (0.234) | (0.868) | (0.195) | (0.055) | (0.488) | (0.145) | (1.733) | (0.374) | (4.874) | (1.328) | (18.207) | (0.292) | (3.811) |
| CLF | 0.899 | 1.162 | **1.569** | 4.038 | **1.332** | 1.015 | **3.818** | 1.576 | 9.958 | **2.290** | 18.685 | 6.218 | 59.180 | **1.895** | 15.157 |
| | (0.039) | (0.097) | (0.198) | (0.669) | (0.171) | (0.055) | (0.416) | (0.141) | (1.462) | (0.303) | (3.816) | (1.033) | (14.345) | (0.247) | (2.932) |
| PCL | 1.103 | 1.316 | 1.684 | 3.934 | 1.477 | 1.256 | 3.861 | 1.718 | **9.403** | 2.384 | 17.553 | 5.964 | 54.764 | 2.030 | 14.333 |
| | (0.036) | (0.076) | (0.157) | (0.549) | (0.122) | (0.051) | (0.388) | (0.118) | (1.323) | (0.246) | (3.192) | (0.906) | (11.801) | (0.199) | (2.579) |
| PL | 1.103 | 1.325 | 1.652 | 3.802 | 1.449 | 1.256 | 3.899 | 1.736 | 9.536 | 2.335 | 16.760 | 5.796 | 52.181 | 1.985 | 13.661 |
| | (0.036) | (0.078) | (0.144) | (0.508) | (0.113) | (0.051) | (0.403) | (0.122) | (1.313) | (0.232) | (2.832) | (0.821) | (11.296) | (0.185) | (2.301) |
| PF | 1.103 | 1.325 | 1.652 | **3.799** | 1.449 | 1.256 | 3.899 | 1.736 | 9.534 | 2.335 | **16.562** | **5.786** | **50.225** | 1.985 | **13.531** |
| | (0.036) | (0.078) | (0.144) | (0.506) | (0.113) | (0.051) | (0.403) | (0.122) | (1.312) | (0.232) | (2.751) | (0.816) | (10.071) | (0.185) | (2.238) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CLCL | **0.829** | **1.364** | 16.404 | 69.083 | 13.290 | **0.982** | 6.374 | **1.976** | 20.339 | 26.344 | 408.146 | 103.234 | 1593.150 | 21.020 | 325.353 |
| | (0.149) | (0.470) | (12.046) | (50.844) | (9.574) | (0.216) | (2.887) | (0.734) | (10.780) | (19.131) | (310.140) | (79.066) | (1180.896) | (15.590) | (239.316) |
| CLL | 0.999 | 1.686 | 3.830 | 14.203 | 3.174 | 1.188 | 8.239 | 2.415 | 28.012 | 6.247 | 82.398 | 21.575 | 294.880 | 5.007 | 64.672 |
| | (0.205) | (0.722) | (2.343) | (9.489) | (2.037) | (0.319) | (4.851) | (1.139) | (17.623) | (3.853) | (55.741) | (14.478) | (202.833) | (3.255) | (43.454) |
| CLF | 0.994 | 1.661 | **3.744** | **13.904** | **3.171** | 1.179 | 7.396 | 2.342 | 24.852 | **6.100** | 82.739 | **21.421** | 307.673 | **4.835** | 65.173 |
| | (0.204) | (0.698) | (2.273) | (9.236) | (1.992) | (0.313) | (4.167) | (1.070) | (15.381) | (3.734) | (56.732) | (14.226) | (215.500) | (3.065) | (44.017) |
| PCL | 1.081 | 1.512 | 15.664 | 55.149 | 12.902 | 1.244 | **6.102** | 2.071 | **19.431** | 23.728 | 319.748 | 80.978 | 1244.666 | 19.565 | 253.241 |
| | (0.182) | (0.481) | (10.863) | (39.277) | (8.987) | (0.259) | (2.894) | (0.741) | (11.004) | (16.002) | (242.227) | (61.561) | (921.245) | (13.558) | (186.697) |
| PL | 1.460 | 2.084 | 3.910 | 14.037 | 3.342 | 1.767 | 8.340 | 2.833 | 27.852 | 6.289 | 81.712 | 21.579 | 292.838 | 5.070 | 64.246 |
| | (0.342) | (0.974) | (2.412) | (9.580) | (2.137) | (0.535) | (4.917) | (1.422) | (18.027) | (3.877) | (55.287) | (14.582) | (203.555) | (3.308) | (43.093) |
| PF | 1.460 | 2.084 | 3.910 | 14.036 | 3.342 | 1.767 | 8.340 | 2.833 | 27.848 | 6.288 | **81.489** | 21.575 | **290.589** | 5.070 | **64.203** |
| | (0.342) | (0.974) | (2.412) | (9.579) | (2.137) | (0.535) | (4.916) | (1.422) | (18.024) | (3.877) | (55.252) | (14.577) | (200.105) | (3.308) | (43.027) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.21: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 5$ and $\varphi = 0.2$ with aggregation matrix $\mathbf{C}_2$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | 1.769 | 2.478 | 4.292 | 13.505 | 3.552 | 2.027 | 8.520 | 3.483 | 25.087 | 6.379 | 73.372 | 20.513 | 260.702 | 5.305 | 5.278 |
| | (0.066) | (0.208) | (1.163) | (5.207) | (0.932) | (0.095) | (1.318) | (0.348) | (5.654) | (1.826) | (33.018) | (7.749) | (127.482) | (1.498) | (1.529) |
| **CLL** | 1.767 | 2.441 | 3.540 | 9.517 | 2.924 | 2.021 | 8.218 | 3.446 | 23.368 | 5.188 | 44.177 | 14.606 | 141.630 | 4.274 | 4.280 |
| | (0.064) | (0.170) | (0.451) | (1.899) | (0.354) | (0.090) | (1.024) | (0.290) | (3.820) | (0.755) | (10.892) | (2.959) | (40.020) | (0.581) | (0.625) |
| **CLF** | **1.766** | 2.417 | 3.402 | 8.976 | 2.822 | 2.018 | 7.932 | 3.377 | 22.227 | 4.973 | 41.352 | 13.698 | 131.896 | 4.099 | 4.118 |
| | (0.064) | (0.160) | (0.364) | (1.470) | (0.290) | (0.090) | (0.895) | (0.260) | (3.234) | (0.601) | (8.424) | (2.254) | (30.686) | (0.462) | (0.497) |
| **PCL** | 1.767 | **2.381** | 3.348 | 8.593 | 2.804 | **2.015** | **7.645** | **3.312** | **20.962** | 4.868 | 38.943 | 13.140 | 123.358 | 4.017 | 4.097 |
| | (0.064) | (0.151) | (0.339) | (1.279) | (0.275) | (0.088) | (0.854) | (0.242) | (2.996) | (0.540) | (7.282) | (2.014) | (25.951) | (0.416) | (0.432) |
| **PL** | 1.774 | 2.413 | 3.294 | 8.341 | 2.748 | 2.028 | 7.794 | 3.366 | 21.326 | 4.797 | 37.850 | 12.771 | 117.919 | 3.950 | 4.038 |
| | (0.065) | (0.157) | (0.322) | (1.185) | (0.253) | (0.090) | (0.882) | (0.253) | (3.080) | (0.536) | (7.091) | (1.905) | (25.312) | (0.393) | (0.405) |
| **PF** | 1.774 | 2.413 | **3.294** | **8.319** | **2.748** | 2.028 | 7.793 | 3.366 | 21.308 | **4.795** | **36.875** | **12.702** | **112.781** | **3.949** | **4.037** |
| | (0.065) | (0.157) | (0.322) | (1.175) | (0.253) | (0.090) | (0.882) | (0.253) | (3.075) | (0.535) | (6.603) | (1.869) | (22.027) | (0.392) | (0.404) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | 1.848 | 3.045 | 35.944 | 148.805 | 30.192 | 2.167 | 13.555 | 4.383 | 45.282 | 58.940 | 919.082 | 235.003 | 3612.535 | 45.365 | 46.796 |
| | (0.322) | (0.979) | (26.610) | (114.005) | (22.615) | (0.472) | (6.584) | (1.672) | (24.396) | (43.831) | (704.500) | (172.903) | (2675.236) | (34.747) | (34.884) |
| **CLL** | 1.942 | 3.726 | 9.023 | 31.507 | 7.404 | 2.338 | 18.492 | 5.823 | 61.757 | 13.688 | 180.169 | 49.012 | 674.923 | 11.197 | 11.189 |
| | (0.376) | (1.664) | (5.564) | (20.973) | (4.659) | (0.587) | (11.513) | (2.954) | (38.643) | (8.984) | (121.745) | (32.524) | (472.268) | (7.040) | (6.716) |
| **CLF** | 1.938 | 3.537 | **8.707** | **30.594** | **7.192** | 2.323 | 16.536 | 5.252 | 54.845 | **13.338** | 180.388 | 49.006 | 691.573 | **10.878** | **10.921** |
| | (0.374) | (1.475) | (5.317) | (20.751) | (4.482) | (0.574) | (9.721) | (2.447) | (33.475) | (8.602) | (122.964) | (32.620) | (482.595) | (6.756) | (6.458) |
| **PCL** | **1.844** | **2.965** | 31.022 | 116.371 | 26.632 | **2.147** | **13.169** | **4.278** | **44.334** | 47.617 | 717.827 | 183.343 | 2813.991 | 37.984 | 38.758 |
| | (0.323) | (0.967) | (21.338) | (88.846) | (18.372) | (0.471) | (6.690) | (1.677) | (25.068) | (33.669) | (549.734) | (134.894) | (2095.935) | (27.149) | (26.992) |
| **PL** | 2.024 | 3.923 | 8.935 | 30.980 | 7.315 | 2.469 | 18.639 | 5.927 | 61.766 | 13.653 | 178.632 | 48.500 | 677.095 | 11.081 | 11.267 |
| | (0.445) | (1.889) | (5.583) | (20.987) | (4.699) | (0.719) | (11.677) | (3.053) | (39.213) | (9.077) | (120.079) | (32.633) | (486.180) | (7.127) | (6.886) |
| **PF** | 2.024 | 3.923 | 8.935 | 30.966 | 7.315 | 2.469 | 18.638 | 5.927 | 61.713 | 13.652 | **176.767** | **48.439** | **659.881** | 11.080 | 11.266 |
| | (0.446) | (1.889) | (5.583) | (20.972) | (4.699) | (0.719) | (11.676) | (3.053) | (39.170) | (9.076) | (118.942) | (32.578) | (466.937) | (7.126) | (6.886) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.22: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 5$ and $\varphi = 0.8$ with aggregation matrix $\mathbf{C}_2$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **INTERPOLATION** | | | | | | | | | | | | | | | |
| **CLCL** | **1.774** | 2.466 | 4.307 | 13.364 | 3.536 | **2.035** | 8.854 | 3.448 | 24.859 | 6.511 | 72.416 | 20.712 | 262.635 | 7.405 | 7.429 |
| | (0.063) | (0.200) | (1.183) | (5.175) | (0.906) | (0.096) | (1.569) | (0.345) | (5.605) | (1.989) | (33.468) | (7.966) | (131.470) | (2.025) | (2.133) |
| **CLL** | 1.821 | 2.467 | 3.537 | 9.505 | 2.942 | 2.075 | 8.365 | 3.432 | 23.236 | 5.245 | 44.056 | 14.768 | 142.824 | 6.068 | 6.046 |
| | (0.068) | (0.185) | (0.465) | (1.927) | (0.367) | (0.102) | (1.073) | (0.298) | (3.866) | (0.771) | (10.767) | (2.941) | (41.951) | (0.831) | (0.840) |
| **CLF** | 1.819 | **2.446** | 3.407 | 8.967 | 2.848 | 2.072 | 8.066 | **3.375** | 22.075 | 5.017 | 41.298 | 13.858 | 133.016 | 5.808 | 5.805 |
| | (0.068) | (0.175) | (0.373) | (1.487) | (0.299) | (0.102) | (0.925) | (0.272) | (3.213) | (0.596) | (8.407) | (2.238) | (31.830) | (0.654) | (0.678) |
| **PCL** | 1.925 | 2.492 | 3.421 | 8.641 | 2.889 | 2.192 | **7.839** | 3.386 | **20.868** | 4.954 | 39.035 | 13.238 | 122.931 | 5.667 | 5.740 |
| | (0.068) | (0.158) | (0.339) | (1.341) | (0.261) | (0.100) | (0.864) | (0.254) | (2.839) | (0.537) | (7.382) | (1.986) | (26.771) | (0.592) | (0.596) |
| **PL** | 1.930 | 2.523 | 3.359 | 8.398 | 2.839 | 2.202 | 7.962 | 3.437 | 21.235 | 4.878 | 37.935 | 12.917 | 118.439 | 5.599 | 5.664 |
| | (0.068) | (0.163) | (0.319) | (1.225) | (0.242) | (0.100) | (0.890) | (0.262) | (2.908) | (0.505) | (7.065) | (1.846) | (25.471) | (0.592) | (0.586) |
| **PF** | 1.930 | 2.523 | **3.359** | **8.375** | **2.839** | 2.202 | 7.962 | 3.437 | 21.218 | **4.877** | **37.025** | **12.855** | **113.050** | **5.597** | **5.661** |
| | (0.068) | (0.163) | (0.319) | (1.213) | (0.242) | (0.100) | (0.890) | (0.262) | (2.903) | (0.504) | (6.548) | (1.822) | (22.116) | (0.590) | (0.585) |
| **EXTRAPOLATION** | | | | | | | | | | | | | | | |
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| **CLCL** | **1.835** | **3.033** | 37.549 | 146.575 | 29.795 | **2.166** | 13.609 | 4.339 | 45.222 | 58.846 | 905.648 | 229.177 | 3595.683 | 66.791 | 65.690 |
| | (0.345) | (1.018) | (28.539) | (113.943) | (20.966) | (0.466) | (6.779) | (1.603) | (25.164) | (43.585) | (678.628) | (169.597) | (2811.117) | (49.508) | (46.032) |
| **CLL** | 1.998 | 3.770 | 8.562 | 31.379 | 7.045 | 2.407 | 18.364 | 5.681 | 63.674 | 14.115 | 174.928 | 47.902 | 625.629 | 16.068 | 15.708 |
| | (0.409) | (1.730) | (5.244) | (21.193) | (4.186) | (0.618) | (10.979) | (2.873) | (40.071) | (9.047) | (123.466) | (32.451) | (440.881) | (9.778) | (9.739) |
| **CLF** | 1.990 | 3.575 | **8.221** | **30.735** | **6.828** | 2.383 | 16.293 | 5.175 | 56.370 | **13.648** | 175.787 | **47.372** | 641.208 | **15.529** | **15.321** |
| | (0.404) | (1.550) | (5.116) | (20.900) | (4.062) | (0.599) | (9.427) | (2.392) | (35.036) | (8.690) | (121.813) | (31.957) | (457.433) | (9.432) | (9.657) |
| **PCL** | 1.961 | 3.057 | 32.045 | 114.812 | 26.331 | 2.276 | **13.144** | **4.295** | **44.491** | 47.703 | 705.650 | 179.060 | 2792.766 | 53.340 | 52.926 |
| | (0.365) | (1.006) | (22.383) | (88.791) | (17.054) | (0.482) | (6.812) | (1.594) | (26.159) | (33.576) | (530.671) | (132.061) | (2196.379) | (38.326) | (35.561) |
| **PL** | 2.311 | 4.097 | 8.512 | 31.172 | 7.140 | 2.792 | 18.638 | 5.853 | 63.851 | 14.094 | 175.437 | 47.453 | 641.179 | 15.849 | 15.561 |
| | (0.582) | (2.000) | (5.124) | (21.237) | (4.329) | (0.866) | (11.181) | (3.025) | (40.379) | (9.007) | (123.522) | (32.686) | (452.840) | (9.932) | (9.817) |
| **PF** | 2.311 | 4.097 | 8.512 | 31.162 | 7.140 | 2.792 | 18.637 | 5.853 | 63.796 | 14.093 | **172.226** | 47.411 | **617.150** | 15.848 | 15.559 |
| | (0.582) | (2.000) | (5.124) | (21.223) | (4.328) | (0.866) | (11.180) | (3.025) | (40.333) | (9.005) | (120.942) | (32.652) | (437.837) | (9.930) | (9.815) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.23: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 10$ and $\varphi = 0.2$ with aggregation matrix $\mathbf{C}_2$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | 2.505 | 3.512 | 6.138 | 18.783 | 4.969 | 2.873 | 12.068 | 4.872 | 35.508 | 9.125 | 101.463 | 29.324 | 372.292 | 7.405 | 81.599 |
| | (0.095) | (0.299) | (1.673) | (7.091) | (1.218) | (0.133) | (2.008) | (0.458) | (7.695) | (2.517) | (45.710) | (11.228) | (192.278) | (2.025) | (35.959) |
| **CLL** | 2.499 | 3.469 | 4.953 | 13.404 | 4.146 | 2.865 | 11.623 | 4.839 | 32.815 | 7.312 | 62.343 | 20.634 | 197.971 | 6.068 | 50.070 |
| | (0.093) | (0.255) | (0.626) | (2.645) | (0.505) | (0.126) | (1.462) | (0.408) | (5.345) | (1.039) | (14.722) | (4.245) | (54.610) | (0.831) | (12.158) |
| **CLF** | 2.498 | 3.420 | 4.779 | 12.661 | 3.992 | 2.859 | 11.225 | 4.744 | 31.312 | 6.998 | 58.447 | 19.357 | 185.285 | 5.808 | 46.953 |
| | (0.093) | (0.232) | (0.509) | (2.099) | (0.407) | (0.125) | (1.267) | (0.373) | (4.503) | (0.826) | (11.356) | (3.164) | (43.961) | (0.654) | (9.594) |
| **PCL** | **2.494** | **3.368** | 4.699 | 12.127 | 3.937 | **2.848** | **10.834** | **4.667** | **29.516** | 6.867 | 55.227 | 18.420 | 171.362 | 5.667 | 44.378 |
| | (0.092) | (0.217) | (0.478) | (1.783) | (0.385) | (0.122) | (1.168) | (0.354) | (4.126) | (0.788) | (10.112) | (2.735) | (37.493) | (0.592) | (7.957) |
| **PL** | 2.503 | 3.410 | 4.624 | 11.837 | 3.882 | 2.866 | 11.023 | 4.739 | 30.001 | 6.759 | 54.022 | 18.013 | 165.368 | 5.599 | 43.000 |
| | (0.093) | (0.228) | (0.462) | (1.749) | (0.362) | (0.125) | (1.223) | (0.368) | (4.261) | (0.747) | (9.640) | (2.637) | (35.247) | (0.592) | (8.116) |
| **PF** | 2.503 | 3.410 | **4.623** | **11.783** | **3.882** | 2.866 | 11.022 | 4.739 | 29.961 | **6.754** | **52.335** | **17.874** | **157.888** | **5.597** | **41.824** |
| | (0.093) | (0.228) | (0.462) | (1.722) | (0.362) | (0.125) | (1.222) | (0.368) | (4.247) | (0.745) | (8.824) | (2.558) | (31.369) | (0.590) | (7.426) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | 2.582 | 4.260 | 52.273 | 203.764 | 42.436 | 3.101 | 18.607 | 6.281 | 65.614 | 83.436 | 1279.115 | 318.096 | 4837.577 | 66.791 | 1046.570 |
| | (0.464) | (1.367) | (38.453) | (153.774) | (31.632) | (0.664) | (8.933) | (2.298) | (36.113) | (60.603) | (951.729) | (248.804) | (3747.381) | (49.508) | (767.398) |
| **CLL** | 2.704 | 5.342 | 12.780 | 44.867 | 10.207 | 3.392 | 25.807 | 8.543 | 88.799 | 19.340 | 246.877 | 71.646 | 948.718 | 16.068 | 200.569 |
| | (0.527) | (2.359) | (8.043) | (28.583) | (6.105) | (0.889) | (15.006) | (4.163) | (58.452) | (12.018) | (169.517) | (47.101) | (631.369) | (9.778) | (143.451) |
| **CLF** | 2.694 | 4.932 | **12.281** | 43.959 | **9.828** | 3.347 | 22.835 | 7.659 | 79.414 | **18.878** | 248.313 | **69.334** | 983.171 | **15.529** | 203.011 |
| | (0.519) | (1.981) | (7.774) | (28.421) | (5.861) | (0.846) | (12.692) | (3.423) | (50.904) | (11.738) | (173.137) | (46.484) | (667.006) | (9.432) | (142.890) |
| **PCL** | **2.573** | **4.172** | 42.623 | 159.113 | 35.659 | **3.079** | **17.921** | **6.184** | **64.283** | 65.996 | 995.771 | 247.550 | 3757.481 | 53.340 | 815.243 |
| | (0.456) | (1.347) | (29.788) | (120.223) | (24.760) | (0.656) | (9.110) | (2.289) | (38.351) | (46.774) | (743.189) | (193.020) | (2936.218) | (38.326) | (598.037) |
| **PL** | 2.803 | 5.432 | 12.688 | 43.885 | 10.222 | 3.556 | 25.840 | 8.625 | 89.073 | 19.312 | 248.082 | 70.834 | 961.526 | 15.849 | 200.794 |
| | (0.606) | (2.467) | (8.115) | (28.077) | (6.208) | (1.045) | (15.335) | (4.306) | (59.952) | (11.938) | (174.318) | (47.020) | (634.799) | (9.932) | (145.311) |
| **PF** | 2.803 | 5.432 | 12.687 | **43.850** | 10.221 | 3.556 | 25.837 | 8.625 | 88.905 | 19.308 | **242.292** | 70.617 | **932.282** | 15.848 | **197.595** |
| | (0.606) | (2.467) | (8.114) | (28.046) | (6.208) | (1.045) | (15.332) | (4.306) | (59.797) | (11.936) | (168.838) | (46.814) | (627.754) | (9.930) | (141.963) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

Table 1.24: The average MSEs of the Monte Carlo HF estimates by the six models under various DGPs with $\sigma^2 = 10$ and $\varphi = 0.8$ with aggregation matrix $\mathbf{C}_2$

**INTERPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.512** | 3.486 | 6.064 | 18.807 | 5.052 | **2.873** | 12.296 | 4.889 | 35.234 | 9.272 | 103.285 | 29.753 | 373.114 | 7.429 | 82.111 |
| | (0.090) | (0.287) | (1.632) | (7.206) | (1.317) | (0.133) | (1.978) | (0.487) | (8.063) | (2.857) | (48.731) | (11.382) | (186.366) | (2.133) | (36.247) |
| **CLL** | 2.542 | 3.477 | 4.992 | 13.591 | 4.198 | 2.895 | 11.825 | 4.859 | 32.624 | 7.456 | 62.854 | 20.681 | 199.215 | 6.046 | 50.243 |
| | (0.095) | (0.256) | (0.681) | (2.869) | (0.559) | (0.135) | (1.542) | (0.422) | (5.494) | (1.134) | (14.744) | (4.347) | (54.317) | (0.840) | (12.197) |
| **CLF** | 2.540 | **3.432** | 4.791 | 12.774 | 4.040 | 2.889 | 11.378 | 4.762 | 31.070 | 7.142 | 58.977 | 19.436 | 186.796 | 5.805 | 47.236 |
| | (0.095) | (0.233) | (0.528) | (2.172) | (0.437) | (0.133) | (1.342) | (0.388) | (4.617) | (0.893) | (11.821) | (3.321) | (43.032) | (0.678) | (9.481) |
| **PCL** | 2.611 | 3.440 | 4.747 | 12.191 | 4.012 | 2.968 | **11.005** | **4.731** | **29.218** | 7.000 | 55.571 | 18.632 | 174.215 | 5.740 | 44.389 |
| | (0.095) | (0.215) | (0.458) | (1.838) | (0.369) | (0.131) | (1.276) | (0.372) | (4.244) | (0.769) | (10.121) | (2.864) | (37.053) | (0.596) | (8.145) |
| **PL** | 2.619 | 3.484 | 4.678 | 11.906 | 3.959 | 2.983 | 11.196 | 4.802 | 29.704 | 6.913 | 54.042 | 18.192 | 167.440 | 5.664 | 42.989 |
| | (0.096) | (0.221) | (0.451) | (1.742) | (0.359) | (0.133) | (1.315) | (0.386) | (4.346) | (0.738) | (9.606) | (2.751) | (34.834) | (0.586) | (7.670) |
| **PF** | 2.619 | 3.484 | **4.677** | **11.852** | **3.958** | 2.983 | 11.195 | 4.802 | 29.664 | **6.908** | **52.397** | **18.048** | **159.844** | **5.661** | **41.884** |
| | (0.096) | (0.221) | (0.450) | (1.716) | (0.359) | (0.133) | (1.314) | (0.386) | (4.336) | (0.736) | (8.857) | (2.692) | (30.388) | (0.585) | (7.071) |

**EXTRAPOLATION**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CLCL** | **2.594** | 4.299 | 53.208 | 207.847 | 41.883 | **3.042** | 19.206 | 6.193 | 64.232 | 83.814 | 1297.589 | 324.432 | 5152.244 | 65.690 | 1047.137 |
| | (0.454) | (1.468) | (39.396) | (153.094) | (31.614) | (0.653) | (9.402) | (2.332) | (36.409) | (62.020) | (1002.559) | (246.730) | (3732.258) | (46.032) | (773.592) |
| **CLL** | 2.779 | 5.503 | 12.914 | 45.764 | 9.929 | 3.353 | 26.266 | 8.101 | 87.585 | 19.585 | 251.588 | 66.777 | 916.470 | 15.708 | 203.613 |
| | (0.574) | (2.593) | (8.290) | (31.502) | (6.029) | (0.867) | (16.176) | (4.001) | (55.822) | (12.320) | (172.127) | (46.305) | (657.075) | (9.739) | (143.113) |
| **CLF** | 2.764 | 5.073 | **12.379** | **44.701** | **9.507** | 3.299 | 23.454 | 7.316 | 78.565 | **18.961** | 254.223 | **65.500** | 938.457 | **15.321** | 208.128 |
| | (0.561) | (2.165) | (7.935) | (30.667) | (5.853) | (0.823) | (13.730) | (3.325) | (47.998) | (11.756) | (176.022) | (45.414) | (681.952) | (9.657) | (148.044) |
| **PCL** | 2.679 | **4.244** | 43.543 | 162.411 | 35.036 | 3.101 | **18.487** | **6.064** | **63.218** | 66.323 | 1012.519 | 252.470 | 4001.301 | 52.926 | 817.163 |
| | (0.482) | (1.470) | (30.357) | (119.525) | (24.594) | (0.660) | (9.433) | (2.332) | (37.501) | (48.142) | (782.869) | (193.301) | (2917.998) | (35.561) | (606.422) |
| **PL** | 3.045 | 5.675 | 12.989 | 45.399 | 9.891 | 3.692 | 26.082 | 8.208 | 88.727 | 19.575 | 252.155 | 66.611 | 949.225 | 15.561 | 201.520 |
| | (0.754) | (2.737) | (8.365) | (31.198) | (6.058) | (1.125) | (16.394) | (4.183) | (57.309) | (12.358) | (175.987) | (46.254) | (681.914) | (9.817) | (145.461) |
| **PF** | 3.045 | 5.675 | 12.988 | 45.336 | 9.891 | 3.692 | 26.079 | 8.208 | 88.564 | 19.571 | **247.584** | 66.357 | **906.342** | 15.559 | **199.847** |
| | (0.754) | (2.737) | (8.364) | (31.160) | (6.057) | (1.125) | (16.392) | (4.182) | (57.159) | (12.355) | (171.467) | (46.064) | (662.608) | (9.815) | (143.432) |

The row represents the average and (standard deviation) of the MSEs of HF estimates by a particular model (row-wise) for 1000 artificial samples of a DGP (column-wise).

# Chapter 2

# The role of economic and financial indicators in forecasting stock volatility

## Abstract

This paper examines the value-added in forecasting volatility by utilising mixed frequency information from market sentiment indicators, economic variables, and activity measures. We consider high-frequency data to generate the volatility measures and low frequency to capture the economic indicators using a mixed sampling frequency approach. The data cover SPY and 100 representative stocks from ten business sectors from 2000 to 2016. Findings are presented on a multi-dimensional scale with three forecasting horizons (daily, weekly, and monthly) and three regimes, including the 2007-08 financial crisis. We find that (extrapolated) macroeconomic

and market sentiment indicators exhibit forecasting information for future volatility beyond its lagged values. Further, we find that utilising a representative factor(s) of all potential predictors results in significant forecast gains in predicting long-term financial volatility even during the financial crisis.

**Keywords:** Realised Volatility; Chow-Lin interpolation; Financial volatility predictors; HARX; Mixed Frequency

## 2.1 Introduction

Since accurate forecasts can inform action to mitigate financial risk, market traders and regulators take great interest in stock market price volatility. In a seminal approach by Engle (1982), the ARCH model captures the time-varying conditional volatility of financial returns. In an extension to that work, Bollerslev (1986) proposed to use the GARCH model. In a further development by Corsi (2009), the Heterogeneous Autoregressive Realised Volatility (HAR-RV) model presents daily volatility as a sequence of autoregressive components, as realised over daily, weekly and monthly horizons. The general hypothesis is that different categories of market participants have diverse horizons and trading activities, such that they react differently to similar news (Müller et al., 1997). Hence, different market indicators carry distinctive signals. This paper aims to combine financial and economic signals from high and low-frequency variables and test the value-added from each piece of information in forecasting financial volatility.

Considering that stock price volatility, unlike stock returns, is persistent, research papers tend to construct volatility models based on time series information; for example, see Corsi and Reno (2009), Ghysels, Santa-Clara, and Valkanov (2006), and Wang et al. (2016). Researchers explore other factors that can help predict financial volatility; see, for example, Hamilton and Lin (1996), Christiansen, Schmeling, and Schrimpf (2012), and Caporin, Rossi, and Magistris (2015), among others. The overall impact of the business cycle upon financial activity has motivated Officer (1973) followed by Schwert (1989), Hamilton and Susmel (1994), Hamilton and Lin (1996), and So, Lam, and Li (1998). More recently, Bollerslev and Zhou (2006) and Bollerslev and Todorov (2011) finds an association between risk-premium volatility and a number of macroeconomic variables. The general finding is one of increased volatility during recessions. All studies mentioned above have considered economic variables observed on a low-frequency (LF), i.e. monthly or annual. This paper investigates the role of a diverse set of economic and financial indicators in predicting high-frequency (HF), i.e. daily, financial volatility.

Where the record of financial volatility is typically available on (at least) a daily basis, less frequent weekly and monthly values are available for macroeconomic and other relevant variables. Most studies overcome the statistical challenge of modelling mixed frequency data by aggregating the HF variable to match the low-frequency (LF) variables or disaggregating from LF to HF, such as the Chow and Lin (1971) method. Ghysels, Santa-Clara, and Valkanov (2004) suggest a mixed data sampling model, MIDAS, to include HF explanatory variables to predict a LF one. Ghysels (2016) introduced and developed the mixed frequency vector autoregressive, MFVAR, to allow the study of bi-directional causality between LF and HF variables. However, when the frequency ratio of the variables is high, the MFVAR suffers from the curse of dimensionality. For example, if the model includes daily and monthly variables, the vector in the MFVAR consists of the daily and monthly observations within a month. Further, the announcement day of the economic data changes from one calendar month to another, posing a challenge in ordering the variables (according to their occurrence within the calendar month) in the VAR model.

To identify the impact of selected indicators on financial market volatility, we use the HARX framework (i.e., the HAR model with additional variable(s)). The HARX model can be applied directly using explanatory variables that are recorded on HF occurrences (e.g. daily). However, we consider two approaches for variables observed at a lower frequency. The first is the reverse MIDAS by Foroni, Guérin, and Marcellino (2018) where the LF explanatory variables are added to the HARX model without any frequency conversion; the second approach is to use the temporal-extrapolation method of Chow and Lin (1971) to disaggregate the LF series into a HF series. The disaggregated series can then be added as daily explanatory variables in the HARX model. We find the latter approach to generally yield better forecasting results.

A growing body of literature suggests that a subset selection and the derived components of variables from a feasible set improve forecast competence (see Fuentes, Poncela, and Rodríguez (2015) and references therein). Therefore, in addition to studying each indicator's added value, we also examine the benefits of combining all the predictors and adding them to the HARX model. Chun and Keleş (2010) propose the Sparse Partial Least Square (SPLS) approach, which imposes a sparsity constraint, similar to that of LASSO, on the Partial Least Square (PLS) method of Wold (1966). The constraint leads to sparse linear combinations of the original predictors. Since this paper is concerned with forecasting, we employ the SPLS to derive orthogonal unobserved components based on the covariance between the set of predictors and the dependent variable. The derived SPLS factors enter the HARX model as explanatory variables. Overall, we find that all indicators' combined information produces the best forecasts. Information about the economic and market conditions, in addition to the lagged volatility, provide further gains in predicting future stock volatility.

In this paper, we contribute to the literature by empirically investigating which economic and financial indicators exhibit essential data in forecasting HF financial volatility of the market and 100 individual stocks in ten business sectors. We also examine when unique and combined factors are most significant in providing pertinent information to predicting future volatility. We do so by studying the forecasting performance over three regimes: before, during, and post the 2007-08 financial crisis. We find that market sentiments hold valuable information in predicting financial volatility before the financial crisis. The value-added of macroeconomic variables, such as new housing projects and industrial production indices, in forecasting stock volatility increased since the financial crisis.

This paper is organized as follows. Section 2.2 provides background on the various predictors of financial volatility investigated in the literature. Section 2.3 presents the volatility measures and HARX models employed in this paper. Section 2.4 describes the data used in this study. Section 2.5 presents the findings and analysis using forecast evaluation methods. Section 2.6 concludes.

## 2.2 Predictors of the financial volatility

Due to stocks' persistent financial volatility, various papers often construct volatility models with distributed lags. Hence, the forecasting accuracy can often be improved by incorporating information beyond those included in the volatility's distributed lag. Therefore, exploring factors that contain significant information in forecasting volatility has occupied a number of researchers, see for example Ghysels, Santa-Clara, and Valkanov (2006), Hamilton and Lin (1996), and Christiansen, Schmeling, and Schrimpf (2012) among others. We categorise indicators as macroeconomic, market sentiment and stock-specific activity.

*Macroeconomic Indicators.* Green (2004) and Andersen et al. (2003b) observe significant intraday effects of macroeconomic announcements. Corradi, Distaso, and Mele (2012) find that macroeconomic factors can explain most of the long-term variation in overall stock volatility. Beltratti and Morana (2006) found that the causality from macroeconomics to financial volatility is stronger than vice versa. In general, any strong contemporaneous relationship between financial volatility and economic conditions would suggest that the use of lagged economic variables might improve volatility forecasts.

We consider macroeconomic indicators with different classifications relative to the business cycle. In particular, the term spread, which is the difference between the three-month LIBOR rate and T-Bill rate, is a forward-looking indicator as it reflects the investors' expectations for future economic activity, demand for credit and monetary policy (see Conrad and Loch (2015) and the references therein). Similarly, the federal fund rates variable is shown to be predictive of volatility jumps (Caporin, Rossi, and Magistris, 2015). Another leading indicator is the housing starts index (HOUST), defined as the total new privately-owned housing units started. It signals the residential investments, so it leads GDP as shown by Kydland, Rupert, and Šustek (2016), among others. The industrial production index (INDPRO) is a coincident indicator, where an increase in the total industrial output yields a decrease in financial volatility (Officer, 1973; Schwert, 1989).

Another coincident indicator we consider is the Auroba-Diebold-Scotti index (ADS) by Aruoba, Diebold, and Scotti (2009). It reflects the prevailing economic conditions combining jobless claims, growth of payroll enrollment, industrial production growth, real manufacturing, trade sales, real personal income, and real GDP. The unemployment rate is a lagging indicator of the business cycle. For example, due to a recession, people have a more challenging time landing jobs (Conrad and Loch, 2015). Similarly, inflation and inflation uncertainty, Producer Price Index (PPI), are both lagging indicators representing the change in prices (Engle, Ghysels, and Sohn, 2013).

*Market Sentiments.* Market sentiments signal current and expected business conditions. Andersen, Bollerslev, and Cai (2000) show that indices for consumer confidence (CCI) and business confidence (BCI) can be used to predict financial volatility in the Japanese market. Further, option prices, as reflected in implied volatility, possess information about future stock market volatility (Andersen and Bondarenko, 2007; Chernov, 2007). Becker, Clements, and McClelland (2009) show that the Chicago Board Options Exchange (CBOE) volatility index (VIX) delivers information relevant to future jump activity. Estimated from options prices, squared VIX is interpreted as the conditional return variance using a risk-neutral probability measure. Hence, it is an unbiased predictor of future financial volatility because it incorporates a risk premium (Bekaert and Hoerova, 2014).

*Activity Measures.* Researchers examined the empirical relationship between stock volatility and trading volume. On the one hand, the mixture of distribution hypothesis (MDH) explains that the volatility-volume relation is a correlation one because any new information yields contemporaneous changes in volume and volatility (Clark, 1973; Harris, 1987). On the other hand, the sequential information arrival hypothesis (SIAH) illustrates that lagged values of volatility may predict current trading volume and vice versa due to the time gap between news and investors' different reactions (Copeland, 1976; Smirlock and Starks, 1988). A positive lead-lag relationship between financial volatility and volume or number of trades has been demonstrated in several empirical studies (Darrat, Zhong, and Cheng, 2007; Girard and Biswas, 2007).

## 2.3 Measures and Models

### Volatility measures

Advances in computing and data technology make it possible to observe HF transaction data. This has led to the so-called realised volatility, RV, which is a non-parametric ex-post estimate of the return variation. The daily RV measure is the sum of intraday squared returns of sub-intervals of the day. Andersen et al. (2003b) show that in the absence of microstructure noise, RV, calculated using HF data, is a consistent estimator of the quadratic variation. Sub-intervals of length 300 sec (e.g. 5 min), which is 78 sub-intervals a day[1], resemble a balance between information gains from high-frequency data and micro-structure effects (Andersen et al., 2001).

Owing to its stylised facts and other desirable statistical properties, RV is preferable over the parametric volatility measures generated from GARCH and Stochastic Volatility (SV) models (Andersen et al., 2003a; Barndorff-Nielsen and Shephard,

[1]See Zhang, Mykland, and Aït-Sahalia (2005) for a discussion on optimising the sampling frequency for the estimation of RV.

2002; Liu, Patton, and Sheppard, 2015). The superiority of RV arises from utilising information available at small intervals, which is lost at the lower frequencies of GARCH and SV models.

The price at the start of the $jth$ interval is computed as the average of the closing and opening prices of intervals $j-1$ and $j$, respectively. The $jth$ return over an interval of length $h$ is:

$$r_j = log(P_{jh}) - log(P_{(j-1)h})$$

Then, dividing the day into $M$ sub-intervals, the intra-day return of the $jth$ sub-interval within the $ith$ day is:

$$r_{j,i} = log(P_{(i-1)M+j}) - log(P_{(i-1)M+(j-1)}), j = 1, 2, ..., M$$

where $M = 78$ which is the length of the day with a 5 min sampling frequency.

As a result the daily return is

$$r_t = \Sigma_{j=1}^{M} r_{j,t}$$

Hence, the daily realised volatility is

$$RV_t^{(d)} := \Sigma_{j=1}^{M} r_{j,t}^2$$

$RV_t^{(w)}$ and $RV_t^{(m)}$ are the weekly and monthly average of RV calculated as follow:

$$RV_t^{(w)} = \frac{1}{5}\Sigma_{i=1}^{5} RV_{t-i}^{(d)} \quad ; \qquad RV_t^{(m)} = \frac{1}{22}\Sigma_{i=1}^{22} RV_{t-i}^{(d)}$$

## HAR-X models

The Heterogeneous Autoregressive Realised Volatility (HAR) model introduced by Corsi (2009) represents daily volatility as a sequence of autoregressive volatility components realised over daily, weekly and monthly time horizons:

$$RV_t^{(d)} = \beta_0 + \beta_d RV_{t-1}^{(d)} + \beta_w RV_{t-1}^{(w)} + \beta_m RV_{t-1}^{(m)} + \epsilon_t^{(d)} \tag{2.1}$$

Consider the HARX model:

$$RV_t^{(d)} = \beta_0^{(d)} + \beta_d^{(d)} RV_{t-1}^{(d)} + \beta_w^{(d)} RV_{t-1}^{(w)} + \beta_m^{(d)} RV_{t-1}^{(m)} + \beta_x^{(d)} X_{t-1} + \epsilon_t \tag{2.2}$$

Similarly, for weekly or monthly realised volatility:

$$RV_t^{(w)} = \beta_0^{(w)} + \beta_d^{(w)} RV_{t-5}^{(d)} + \beta_w^{(w)} RV_{t-5}^{(w)} + \beta_m^{(w)} RV_{t-5}^{(m)} + \beta_x^{(w)} X_{t-5} + \epsilon_t^{(w)} \tag{2.3}$$

$$RV_t^{(m)} = \beta_0^{(m)} + \beta_d^{(m)} RV_{t-22}^{(d)} + \beta_w^{(m)} RV_{t-22}^{(w)} + \beta_m^{(m)} RV_{t-22}^{(m)} + \beta_x^{(m)} X_{t-22} + \epsilon_t^{(m)} \tag{2.4}$$

where $X$ is an exogenous variable sampled at a daily level. Some variables sampled at a LF can be incorporated in the model using one of two approaches.

The first approach is the reverse MIDAS suggested by Foroni, Guérin, and Marcellino (2018) to use LF data in predicting HF variables:

$$RV_t^{(d)} = \beta_0^{(d)} + \beta_1 B(L, \theta) RV_{t-1}^{(d)} + \beta_x^{(d)} X_{t-1} + \epsilon_t \tag{2.5}$$

where $B(L, \theta)$ is a weighting function. Note that the HAR model can be thought of as a particular weighting scheme. While it is interesting to study the optimal restricted AR in modelling RV, it is beyond the scope of this paper. Our approach is to apply Foroni, Guérin, and Marcellino (2018)'s reverse approach to the HARX model. Where this approach holds the LF observation constant within each LF period (month) it allows its coefficient to vary according to the day of the month:

$$\beta_x X_t := \Sigma_{i=1}^{k} \gamma_i D_i X_{\tau}^{LF}$$

where $k$ is the maximum frequency ratio (i.e number of days in a month). $\gamma_i$ is the coefficient to be estimated, $D_i$ is a dummy variable that takes the value one when $t$ falls on the $ith$ day of the month and zero otherwise. $X_{\tau}^{LF}$ is the low-frequency observation of variable $X$ in month $\tau$ such that $t$ falls in month $\tau$.

The second approach to incorporating a LF (monthly) variable into the HARX model is to disaggregate the LF variable to daily values before adding these as a HF series. Among the numerous extrapolation methods, the Chow and Lin (1971) (CL) is selected because macroeconomic variables' statistical characteristics align with the method's assumptions. Particularly, macroeconomic variables are thought to have low variance within the month. In a nutshell, the CL method uses a statistical relationship between LF data and HF indicator variable(s) to create a HF time series that is consistent with the LF data. By our choice of the Aruoba-Diebold-Scotti Business Conditions Index (ADS) and the daily Term Spread (TS) as daily indicators, we obtain signals of the current state of the business cycle (Berge and Jordà, 2011). Note that, unlike interpolation, no future observation is used in extrapolation, an important feature of the estimated HF series in the out-of-sample window to be used in real-time forecasting. For a more elaborate technical explanation see Dagum and Cholette (2006) and Chow and Lin (1971).

Finally, we add all the explanatory variables to the HARX model. We use the Spare Partial Least Square (SPLS) method of Chun and Keleş (2010) to extract components from the original set of predictors. The advantage of this method is that instead of adding all the predictors to the HARX model, we replace them with a smaller set of factors that capture their joint variation while also considering the in-sample covariation with the dependent variable. The HARX model with SPLS factors is as follows:

$$\begin{aligned} RV_t^{(d)} &= \beta_0^{(d)} + \beta_d^{(d)} RV_{t-1}^{(d)} + \beta_w^{(d)} RV_{t-1}^{(w)} + \beta_m^{(d)} RV_{t-1}^{(m)} + \beta_f^{'(d)} f_{t-1} + \epsilon_t \\ f_t &= W z_t \end{aligned} \tag{2.6}$$

where $z_t$ is the $(M \times 1)$ vector of all $M$ predictors listed in table (2.1) at time $t$. Note that in this model the $X = f_{t-1}$ where $f_t$ is a $(k \times 1)$ vector of latent unobserved factors obtained using SPLS and $\beta_f$ is its $(k \times 1)$ coefficients vector. $W$ is a $(k \times M)$ matrix of the weights assigned to each one of the predictor variables in each of the SPLS component in $f_t$.

Firstly, the direction matrix, W in equation (2.6), is found using SPLS. Secondly, once the factor(s), $\hat{f}_t$, is estimated, it enters the HARX model, equation (2.6), to serve as a reduced set of explanatory variables. We do not include RV lags in the $z_t$ and instead keep the HAR framework.

Estimation is undertaken using OLS. While the error terms suffer from autocorrelation and heteroskedasticity, the GLS estimation shows no significant difference in the coefficients or the forecasting evaluation measures. Therefore, we use OLS for its simplicity and computational convenience. Lastly, we follow Swanson and White (1995) and use an "insanity filter" to ensure that the volatility forecasts are not negative in all the models. In particular, we replace negative forecasted values of any model by the the lagged RV.

## 2.4 Data

We consider 17 years of high-frequency data for the period 2000 to 2016. The dataset includes the SPY and 100 stocks representing ten sectors. The financial data comprises 4277 trading days observed at the tick level. The use of cleaned data from *TickWrite*[2] makes our results easier to authenticate and replicate. The sample range and sector coverage allow us to examine the sensitivity of the forecasting performance of our models across different market regimes and individual stocks. Table (2.5) provides descriptive statistics of the RV of each stock.

Table (2.1) provides a summary of the explanatory variables, $X$, in the HARX models (eq. 2.2, 2.3, 2.4) and Table (2.6) presents their descriptive statistics. The variables were downloaded from multiple sources namely: *ALFRED*, *OECD*, and *QUANDL*. Data revisions can be substantial for macroeconomic variables. Thus, employing revised instead of first-release real-time data can be misleading regarding forecasting evaluation (see, for example, Stark (2010)). Hence, for variables that undergo revision, we use *ALFRED*, which allows retrieving vintage versions of the economic data

[2]TickWrite is a database that provides data on a commercial basis for futures, Index and equity markets. Tick Data is sourced from NYSE's TAQ (Trade and Quote) database. Tick adjusts the TAQ database for ticker mapping, code filtering, price splits, and dividend payments https://www.tickdata.com/.

recorded at the initial release date of each observation. In addition to macroeconomic variables growth rates, we also considered business indices and squared VIX as market sentiments variables and activity measures.

Table 2.1: List of Predictors

| **Macroeconomic Variables** | **Frequency** | **Functional Form** | **Source** |
|---|---|---|---|
| Aruoba-Diebold-Scotti Business Conditions Index (ADS) | Daily | Level | *QUANDL* |
| Effective Federal Funds Rate (FED-FUNDS) | Daily | Logarithmic-square | *ALFRED* |
| Housing Starts Index (HOUST) | Monthly | Growth | *ALFRED* |
| Industrial Production Index (INDPRO) | Monthly | Growth | *ALFRED* |
| Produce Price Index (PPIACO) | Monthly | Level | *ALFRED* |
| Term Spread (TS) | Daily | Logarithmic-square | *ALFRED* |
| Unemployment Rate (UNRATE) | Monthly | Growth | *ALFRED* |
| **Market Sentiments** | **Frequency** | **Functional Form** | **Source** |
| Economic Policy Uncertainty Index (EPU) | Daily | Level | *Policy Uncertainty* |
| Business Confidence Index (BCI) | Monthly | Level | *OECD* |
| Consumer Confidence Index (CCI) | Monthly | Level | *OECD* |
| Composite Leading Indicator (CLI) | Monthly | Level | *OECD* |
| CBOE Volatility Index (VIX) | Daily | Squared | *ALFRED* |
| **Activity Measures** | **Frequency** | **Functional Form** | **Source** |
| Number of Trades | Daily | Level/($10^4$) | *TickWrite* |
| Volume | Daily | Level/($10^8$) | *TickWrite* |

## 2.5 Empirical Results

The explanatory variables under consideration are observed at either daily or monthly frequency. Hence, given that RV's frequency is daily, we have two scenarios based on the frequency of the explanatory variable, $X$. For explanatory variables with daily frequency, we employ the HARX models (eq. (2.2), (2.3), (2.4)) directly. For explanatory variables with monthly frequency, we use its extrapolated daily data. We also considered the reverse MIDAS approach and found no significant differences with the former in the forecasting results of weekly and monthly horizons. However, for the daily forecast horizon, the CL extrapolation approach, which accounts for daily variations, yields better predictions. Therefore, we only report the results using the CL approach for conciseness.

In what follows, we report the results using the full sample period and sub-samples. Since the time frame of the data includes the 2007-08 financial crisis, we dissect the full sample into three periods: before the financial crisis (Pre-Crisis) from 2000 to 2006, during the financial crisis (Crisis) from 2007 to 2010, and after the financial crisis (Post-Crisis) from 2011 to 2016.

### 2.5.1 In-sample Estimation

Tables (2.7), (2.8), and (2.9) report the in-sample estimation results of the HAR and HARX models (eq. 2.2, 2.3, & 2.4) with each of the exogenous variables listed in table (2.1) for three sample periods (pre-crisis, crisis, and post-crisis), respectively. The scale of the HAR model coefficients varies across forecast horizons. In particular, $\hat{\beta}_d$ is the highest for equation (2.2) than (2.3) and (2.4). That is to say, more recent daily information, lagged $RV^{(d)}$, is more relevant for the daily forecaster than the longer-term forecaster. Similarly, $\hat{\beta}_w$ and $\hat{\beta}_m$ are higher in long-term predictions models. Further, the magnitude and significance of the HAR and HARX coefficients change over the three sample periods. During the crisis period, $\hat{\beta}_w$ increases while $\hat{\beta}_m$ decreases, which suggests that recent information becomes more valuable during the crisis than the historical average.

Few magnitude and sign changes in the exogenous variables' coefficients are identified across the three periods, but the pattern is consistent across the three forecast horizons. Generally, the magnitude of most of the X variables' coefficients increases during the crisis. Regarding market sentiments, BCI and CLI reflect current conditions and signals about the economy's future, respectively. Hence, an increase in their values yields a decline in the volatility forecasts. Also, CCI measures future developments in households' consumption and saving. Hence, there has been a significant negative relationship between CCI and volatility forecasts since the crisis. Further, the VIX's coefficient is positive and significant in all three sample periods. Lastly, the coefficients of volume and trades become positive and significant since the crisis, which implies that high volume and trades signal high RV of the next day, week, and month.

In terms of the macroeconomic variables, overall, the sign of their coefficients is in line with the counter-cyclical property of long-term volatility observed by Engle, Ghysels, and Sohn (2013) and Conrad and Loch (2015), among others, even at a daily level. We find that the signs of the coefficients of leading indicators such as TS and HOUST are consistently positive and negative, respectively, across the three sample periods. As for lagging indicators, a rise in UE or PPI is linked with an increase in market volatility. Further, coincident indicators such as ADS and the INDPRO signal the current business conditions. Their coefficients are negative before and during the crisis, i.e. an improvement in economic conditions generally reduce market volatility. However, the relationship is positive but insignificant after the crisis. The latter observation suggests that an increase in these particular coincident indicators might not have an instantaneous counter-cyclical effect on HF financial volatility.

The in-sample estimation results over the three periods suggest that the relationship between some exogenous variables and RV has changed over the three periods. Hence, looking at the overall out-of-sample (OOS) performance over the whole sample could be misleading. Therefore, we are interested in examining the OOS forecasting performance over the three subsample periods representing the: pre-crisis, crisis, and post-crisis regimes.

### 2.5.2 Out-of-Sample Analysis

This paper examines the role of a set of economic and financial variables in predicting financial volatility. To do so, we evaluate the OOS performance of the HARX model against the HAR model as the benchmark. We estimate the models in a rolling window of 1000 daily observations. Specifically, we estimate the model using the first 1000 observations to produce the first out-of-sample forecast value corresponding to the $1001^{st}$ RV observation. Then, we estimate the model from $2^{nd}$ to the $1001^{st}$ observation to have the second OOS forecast corresponding to the $1002^{nd}$ RV observation. We proceed this way to produce 3277 OOS forecasts corresponding to RV's $1001^{st}$ to $4277^{th}$ observation.

Many studies are focused on evaluating and comparing volatility models, see, for example, Poon and Granger (2003) for an extensive review on the matter. Most apply a loss function, where model-based predictions of the conditional variance are compared to estimates of the conditional variance (e.g. realised volatility). Many researchers have considered the Mean Squared Forecast Error (MSFE) as a forecast evaluation method in volatility models (Andersen, Bollerslev, and Lange, 1999; Corsi and Reno, 2009):

$$MSFE = \frac{1}{N}\Sigma_{t=1}^{N}[RV_{t+1} - \hat{RV}_{t+1}]^2$$

To ease the comparison between the models, we compute the relative loss function (relative MSFE) defined as:

$$relative\ MSFE_{HARX} = \frac{MSFE_{HARX}}{MSFE_{HAR}} \tag{2.7}$$

where a value below 1 indicates that the HARX outperforms the HAR model.

Table (2.2) displays the relative $MSFE$ of the various HARX models relative to the HAR model. The combined information using SPLS results in the highest gains over the entire sample for all forecasting horizons. We observe a distinct change in variable predictability gains across the forecast horizons and periods per individual variable. Overall, the variables yield higher forecast gains for the long-term RV, where all variables, except TS and FEDFD, improve the forecast of monthly volatility. Some information is more relevant during particular regimes than others.

For example, we note that market sentiments and macroeconomic variables are more relevant for the weekly and monthly forecasts than daily forecasts. These variables produce the most significant improvements in the RV forecasts during the crisis. These variables flag valuable information beyond the lagged RV during unprecedented times.

VIX and activity measures, on the other hand, were better predictors before the financial crisis. Overall, as expected, combined information yields the best forecasts during periods of uncertainty.

To investigate the findings by sectors, we report, for each business sector over each regime using each set of X variables, the OOS median of the daily, weekly, and monthly forecast gains in Tables (2.9) to (2.20). Overall, the financial sector has the highest forecast gains from incorporating market and economic conditions information. There are no significant differences among the other sectors. Hence, we look at the overall impact on forecasting volatility for the 100 individual stocks.

In Table (2.3), we report the mean and the skewness of the relative MSFE. Consistent with the SPY, the findings suggest that information beyond the lagged RV can be more relevant for the long-term horizons of individual stocks' volatility during the financial crisis. We find that, on average, combined information (SPLS factors) yields the highest forecast gains for full-sample, particularly during the crisis. The skewness also enables us to know more about the distribution. If it is negatively skewed, it means that while for some stocks, we get even lower relative MSFE than the average, with the majority of the stocks, the relative MSFE is higher than the mean. We mainly observe negative skewness during the financial crisis.

Table 2.2: Relative MSFE of HAR-X in Forecasting Market Volatility

| | **Daily** | | | | **Weekly** | | | | **Monthly** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **FS** | **Pre** | **Crisis** | **Post** | **FS** | **Pre** | **Crisis** | **Post** | **FS** | **Pre** | **Crisis** | **Post** |
| **ADS** | 1.009 | **0.987** | 1.011 | **0.999** | **0.940** | **0.971** | **0.933** | **0.996** | **0.734** | **0.990** | **0.717** | **0.908** |
| **BCI** | 1.023 | **0.980** | 1.026 | 1.003 | **0.970** | **0.947** | **0.966** | 1.006 | **0.635** | **0.890** | **0.593** | 1.108 |
| **CCI** | 1.017 | 1.001 | 1.020 | **0.998** | **0.984** | 1.011 | **0.984** | **0.987** | **0.711** | 1.043 | **0.685** | **0.980** |
| **CLI** | 1.021 | **0.990** | 1.024 | 1.001 | **0.970** | **0.976** | **0.966** | 1.002 | **0.653** | **0.975** | **0.621** | **0.995** |
| **EPU** | 1.023 | **0.950** | 1.026 | 1.001 | 1.006 | **0.983** | 1.007 | **0.993** | **0.950** | 1.009 | **0.940** | 1.060 |
| **FEDFD** | 1.002 | 1.004 | 1.002 | 1.005 | 1.002 | 1.048 | 1.003 | **0.999** | 1.005 | 1.097 | 1.008 | **0.958** |
| **HOUST** | 1.006 | 1.182 | 1.006 | 1.002 | 1.003 | 1.125 | 1.003 | **0.998** | **0.983** | 1.075 | **0.986** | **0.929** |
| **INDPRO** | 1.008 | 1.018 | 1.007 | 1.012 | **0.966** | 1.093 | **0.955** | 1.039 | **0.810** | 1.232 | **0.775** | 1.170 |
| **PPI** | 1.003 | **0.945** | **1.000** | 1.026 | **0.945** | **0.935** | **0.922** | 1.114 | **0.728** | **0.931** | **0.681** | 1.268 |
| **TRD** | 1.024 | **0.989** | 1.052 | **0.815** | **0.986** | **0.646** | **0.998** | **0.913** | **0.925** | **0.379** | **0.929** | **0.977** |
| **TS** | 1.007 | 1.129 | 1.008 | **1.000** | 1.008 | 1.233 | 1.007 | 1.001 | 1.009 | 1.165 | 1.005 | 1.034 |
| **UE** | 1.013 | 1.013 | 1.015 | 1.002 | **0.989** | 1.036 | **0.984** | 1.022 | **0.779** | 1.051 | **0.737** | 1.245 |
| **VIX** | **0.859** | **0.757** | **0.859** | **0.861** | **0.904** | **0.461** | **0.908** | **0.892** | **0.987** | **0.471** | **0.998** | **0.948** |
| **VOL** | 1.024 | 1.138 | 1.043 | **0.874** | 1.007 | **0.893** | 1.019 | **0.924** | **0.972** | **0.648** | **0.977** | **0.961** |
| **SPLS** | **0.842** | 1.037 | **0.844** | **0.826** | **0.839** | **0.901** | **0.829** | **0.912** | **0.652** | **0.578** | **0.630** | **0.932** |

This table presents the MSE values of each HAR-X model relative to the MSFE of the HAR model in forecasting the RV of SPY. The results are listed over the Full-sample period and the sub-samples defined as pre-crisis, crisis, and post-crisis period. A value of less than one in bold indicates forecast gains over the HAR model. The lowest value is underlined, showing the best model in the relevant period.

Table 2.3: Mean and Skewness of Relative MSE of HAR-X in Forecasting Individual Stocks Volatility

| | **Daily** | | | | **Weekly** | | | | **Monthly** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **FS** | **Pre** | **Crisis** | **Post** | **FS** | **Pre** | **Crisis** | **Post** | **FS** | **Pre** | **Crisis** | **Post** |
| **ADS** | **0.991** | **0.994** | **0.991** | **0.998** | **0.930** | **0.989** | **0.920** | <u>**0.980**</u> | **0.762** | **0.988** | **0.734** | <u>**0.823**</u> |
| | -3.214 | -6.739 | -2.562 | 0.466 | -3.667 | -0.579 | -3.963 | -2.993 | -0.524 | 1.804 | -1.018 | -1.250 |
| **BCI** | 1.002 | 1.003 | 1.002 | 1.043 | **0.957** | **0.999** | **0.946** | 1.099 | **0.718** | **0.973** | **0.657** | 1.291 |
| | -3.271 | 6.362 | -2.906 | 5.855 | -4.164 | 2.224 | -4.028 | 4.536 | -0.695 | 0.568 | -0.125 | 2.254 |
| **CCI** | 1.002 | 1.015 | 1.003 | 1.029 | **0.972** | 1.042 | **0.964** | 1.050 | **0.765** | 1.074 | **0.712** | 1.157 |
| | -3.621 | 3.213 | -3.115 | 6.023 | -4.146 | 1.663 | -4.098 | 4.796 | -0.777 | 0.373 | -0.418 | 2.635 |
| **CLI** | 1.000 | 1.005 | 1.000 | 1.018 | **0.953** | 1.014 | **0.943** | 1.041 | **0.711** | 1.031 | **0.655** | 1.075 |
| | -3.152 | 2.696 | -2.812 | 6.341 | -4.471 | 2.935 | -4.244 | 3.679 | -0.811 | 3.869 | -0.576 | 2.108 |
| **EPU** | 1.001 | <u>**0.979**</u> | 1.001 | 1.044 | **0.986** | **0.964** | **0.985** | 1.049 | **0.953** | **0.995** | **0.940** | 1.138 |
| | -0.430 | -2.552 | -0.180 | 4.856 | -2.676 | -2.165 | -2.664 | 3.212 | -1.195 | -0.339 | -1.073 | 2.363 |
| **FEDFD** | 1.005 | 1.038 | 1.003 | <u>**0.996**</u> | 1.011 | 1.128 | 1.004 | **0.971** | 1.027 | 1.324 | 1.013 | **0.874** |
| | 5.616 | 5.442 | 1.286 | -6.047 | 5.835 | 2.620 | 0.625 | -3.489 | 5.153 | 1.905 | -0.056 | -1.014 |
| **HOUST** | 1.003 | 1.048 | 1.001 | **0.997** | 1.001 | 1.080 | **0.995** | **0.976** | **0.985** | 1.116 | **0.976** | **0.887** |
| | 3.360 | 5.149 | -1.710 | -4.517 | 4.856 | 3.720 | -2.031 | -3.474 | 3.755 | 3.021 | -2.265 | -1.764 |
| **INDPRO** | **0.998** | 1.015 | **0.997** | 1.042 | **0.969** | 1.045 | **0.956** | 1.116 | **0.864** | 1.132 | **0.813** | 1.278 |
| | -4.043 | 6.178 | -3.612 | 5.160 | -3.372 | 2.311 | -4.060 | 3.553 | 0.095 | 2.252 | -0.199 | 1.827 |
| **PPI** | **0.997** | 1.010 | **0.993** | 1.070 | **0.956** | 1.009 | **0.933** | 1.217 | **0.789** | 1.010 | **0.725** | 1.396 |
| | 2.967 | 9.705 | 0.139 | 4.576 | 1.248 | 8.331 | -0.195 | 4.198 | -0.004 | 2.920 | 0.604 | 2.710 |
| **TRD** | **0.987** | 1.175 | **0.996** | 1.030 | **0.969** | 1.019 | **0.967** | 1.050 | **0.933** | **0.887** | **0.928** | 1.059 |
| | -3.531 | 4.615 | -3.451 | 2.894 | -3.993 | 2.855 | -4.589 | 4.389 | -3.001 | 1.752 | -3.376 | 4.233 |
| **TS** | 1.005 | 1.062 | 1.003 | 1.005 | 1.011 | 1.155 | 1.003 | 1.042 | 1.033 | 1.243 | 1.016 | 1.162 |
| | 5.272 | 8.101 | -3.524 | 4.086 | 4.824 | 3.497 | -2.288 | 6.206 | 2.118 | 1.068 | -0.624 | 3.926 |
| **UE** | 1.004 | 1.009 | 1.004 | 1.054 | **0.983** | 1.028 | **0.975** | 1.153 | **0.834** | 1.068 | **0.770** | 1.535 |
| | -3.076 | 1.488 | -2.626 | 5.195 | -3.893 | 1.377 | -3.266 | 4.687 | -0.330 | 2.003 | -0.117 | 2.398 |
| **VIX** | **0.935** | **0.940** | **0.929** | 1.076 | <u>**0.892**</u> | <u>**0.802**</u> | <u>**0.887**</u> | 1.095 | **0.913** | <u>**0.748**</u> | **0.926** | 1.051 |
| | -1.980 | -2.685 | -1.552 | 5.009 | -2.126 | -0.177 | -1.773 | 5.633 | -1.817 | 0.440 | -1.980 | 7.019 |
| **VOL** | 1.006 | 1.174 | 1.008 | 1.036 | 1.009 | 1.073 | 1.002 | 1.023 | 1.002 | 1.008 | **0.991** | 1.023 |
| | -2.201 | 6.444 | -2.148 | 5.844 | -1.521 | 5.653 | -3.512 | 2.914 | 1.395 | 3.080 | -5.630 | 6.441 |
| **SPLS** | <u>**0.933**</u> | 1.062 | <u>**0.923**</u> | 1.077 | **0.898** | **0.830** | **0.888** | 1.087 | <u>**0.674**</u> | **0.824** | <u>**0.629**</u> | 1.071 |
| | -1.666 | 4.943 | -1.152 | 5.091 | -1.138 | -0.185 | -0.934 | 5.612 | -0.345 | 0.907 | -0.399 | 2.446 |

This table presents the mean and skewness of MSE values of each HAR-X model relative to the MSFE of the HAR model in forecasting the RV of the 100 individual stocks. The results are listed over the Full-sample period, and the sub-samples are defined as pre-crisis, crisis, and post-crisis periods. A mean of less than one in bold indicates forecast gains over the HAR model. The lowest value is underlined, showing the best model in the relevant period.

We note that macroeconomic variables, particularly the unemployment rate, became more valuable during the crisis. Similarly, on average, economic conditions indicators are more valuable during unprecedented times, except for a few individual stocks. In the case of SPY, activity measures are more valuable during the financial crisis than pre-and post-crisis regimes. Hence, we only find supporting evidence for the volatility-volume and the trades-volatility lead-lag relationship among the individual stocks during the crisis period. The finding suggests that, on average, the stock activity measures do not necessarily add valuable information beyond those included in the lagged volatility for all individual stocks.

### 2.5.3 Forecasting Significance

To test for forecasting significance, we consider the Clark and West (2007) (CW) statistics on MSFE which has been widely adopted in the literature for comparing nested models (see, for example, Audrino and Hu (2016) and Audrino, Sigrist, and Ballinari (2019) for forecasting RV). The test adds an adjustment term to the OOS difference in MSFE that accounts for parameter estimation noise. The null hypothesis involves the population difference in MSFE between the two nested models for a given forecasting horizon. The test statistic is defined as:

$$\hat{f}_{t+1} = (RV_{t+1} - R\hat{V}_{t+1}^{HAR})^2 - [(RV_{t+1} - R\hat{V}_{t+1}^{HARX})^2 - (R\hat{V}_{t+1}^{HAR} - R\hat{V}_{t+1}^{HARX})^2] \quad (2.8)$$

Using the resulting $\hat{f}_{t+1}$ (2.8), we estimate the regression $\hat{f}_{t+1} = \mu + \epsilon_t$ , then test the null hypothesis: $\mu \leq 0$ which implies that the HARX model is not better than the HAR model. We reject the null if the t-statistic is greater than +1.282 for 10% significance or +1.645 for 5% significance. The results are reported in table (2.4) for all HARX models across the three forecasting horizons.

According to the CW test, lagging and coincident indicators such as UE and INDPRO show significant gains during the crisis. In contrast, leading indicators such as HOUST show significant gains after the crisis. HOUST's long-term forecasting gains are also significant during the crisis. Further, most market sentiments are significant for long-term forecasting horizons during and after the crisis. However, VIX is the only variable that yields significant gains during the crisis period for the daily forecasts. Volume and number of trades variables are found to predict the market RV before and after the crisis. Lastly, SPLS is the only one that significantly improves the prediction of the market volatility consistently in all three regimes and three forecasting horizons.

The previously mentioned forecast evaluation methods examine the global forecasting performance over a chosen interval. However, our entire sample period includes unprecedented times, which caused instability in the economy. While dividing the sample into three sub-samples relative to the financial crisis period, namely the pre-crisis, crisis, and post-crisis period, mitigates part of the regime-change problem, it still suffers from other challenges. For example, it is nearly impossible to depict the exact day of the beginning and end of a regime.

Table 2.4: Clark-West MSFE significance test of forecasting RV of SPY using HARX model compared to the HAR model

| | Daily | | | | Weekly | | | | Monthly | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **FS** | **Pre** | **Crisis** | **Post** | **FS** | **Pre** | **Crisis** | **Post** | **FS** | **Pre** | **Crisis** | **Post** |
| **ADS** | 0.016 | 0.001* | 0.032 | 0.013** | 0.218* | 0.002 | 0.657* | 0.034*** | 0.994* | 0.002 | 3.015* | 0.132*** |
| **BCI** | -0.053 | 0.001*** | -0.175 | 0.001 | 0.100** | 0.004*** | 0.308* | 0.009 | 1.016* | 0.019*** | 3.236* | 0.020 |
| **CCI** | -0.040 | 0.000 | -0.140 | 0.007 | 0.059** | 0.000 | 0.160 | 0.021 | 0.765* | 0.001 | 2.368* | 0.069** |
| **CLI** | -0.049 | 0.001*** | -0.162 | 0.002 | 0.091* | 0.002*** | 0.283* | 0.008 | 0.948* | 0.008*** | 3.007* | 0.034* |
| **EPU** | -0.050 | 0.004*** | -0.177 | 0.007 | 0.017 | 0.001*** | 0.031 | 0.015 | 0.214 | 0.000 | 0.656 | 0.024 |
| **FEDFD** | 0.004 | 0.002 | 0.003 | 0.005 | 0.008 | 0.003 | -0.002 | 0.017** | 0.062 | 0.006 | 0.128 | 0.047* |
| **HOUST** | -0.013 | -0.004 | -0.053 | 0.010*** | 0.008 | -0.003 | -0.008 | 0.025*** | 0.067*** | -0.003 | 0.087* | 0.089** |
| **INDPRO** | -0.001 | 0.000 | 0.004 | -0.006 | 0.121* | 0.000 | 0.400* | -0.006 | 0.549* | -0.006 | 1.800* | -0.015 |
| **PPI** | 0.012 | 0.003*** | 0.032 | 0.002 | 0.149* | 0.005*** | 0.464* | 0.010 | 0.694* | 0.018*** | 2.122* | 0.071** |
| **TRD** | 0.097 | 0.003*** | -0.090 | 0.270* | 0.140** | 0.028*** | 0.298* | 0.091* | 0.340 | 0.128*** | 0.973 | 0.020* |
| **TS** | -0.013 | -0.004 | -0.053 | 0.010 | 0.008 | -0.003 | -0.008 | 0.025 | 0.067 | -0.003 | 0.087 | 0.089 |
| **UE** | -0.035 | 0.000 | -0.115 | 0.002 | 0.040** | -0.001 | 0.124** | 0.004 | 0.563* | -0.002 | 1.769* | 0.034 |
| **VIX** | 1.513* | 0.035*** | 4.489** | 0.269** | 0.563** | 0.043*** | 1.645** | 0.100** | 0.165 | 0.070*** | 0.452 | 0.020*** |
| **VOL** | 0.009 | -0.003 | -0.231 | 0.176** | 0.056 | 0.010*** | 0.040 | 0.089** | 0.135* | 0.069*** | 0.326 | 0.039*** |
| **SPLS** | 1.467* | 0.021*** | 4.420** | 0.220* | 0.592** | 0.043*** | 1.641** | 0.166* | 1.063* | 0.076*** | 3.278* | 0.067*** |

*** 1% significance, ** 5% significance, and * 10% significance.

Therefore, we examine local superiority using the fluctuation test developed by Giacomini and Rossi (2010) (GR). The methods are designed to examine the relative forecasting performance of two competing models in the presence of instabilities. Thus, we can more accurately detect any reversal in the relative forecasting ability of HAR and HARX models. The general formula of the local relative performance statistic is defined as:

$$F_{t,m}^{OOS} = \hat{\sigma}^{-1} m^{-1} \Sigma_{j=t-m/2}^{t+m/2-1} \Delta L_j(\hat{\theta}_{j-h,R}^{HAR}, \hat{\theta}_{j-h,R}^{HARX}), \quad t = R+h+m/2, .., T-m/2+1 \quad (2.9)$$

Where $m$ is the length of the of the moving window, set to 250 (i.e. approximately one year of data). $\Delta L_j$ is the loss function difference between the two models, HAR and HARX. With the CW statistic for nested models, the local relative performance MSFE becomes:

$$F_{t,m}^{OOS} = \hat{\sigma}^{-1} m^{-1} \Sigma_{j=t-m/2}^{t+m/2-1} \hat{f}_{t+\tau}, \quad t = R + h + m/2, .., T - m/2 + 1 \quad (2.10)$$

where $\hat{f}_{t+\tau}$ is defined in equation (2.8) and $R = 1000$ is the number of observations in the in-sample. For every $t$, the parameters of the models are re-estimated over the in-sample period in a rolling window scheme: $t-h-R+1, ..., t-h$. The null hypothesis for this test is that the two competing models, HAR and HARX, have equal OOS performance at each point in time. Wherever the local relative MSFE, $F_{t,m}^{OOS}$, exceeds the critical value we reject the null hypothesis and conclude that the HARX produced better forecasts than the HAR model over the period of length $m$ centered at $t$.

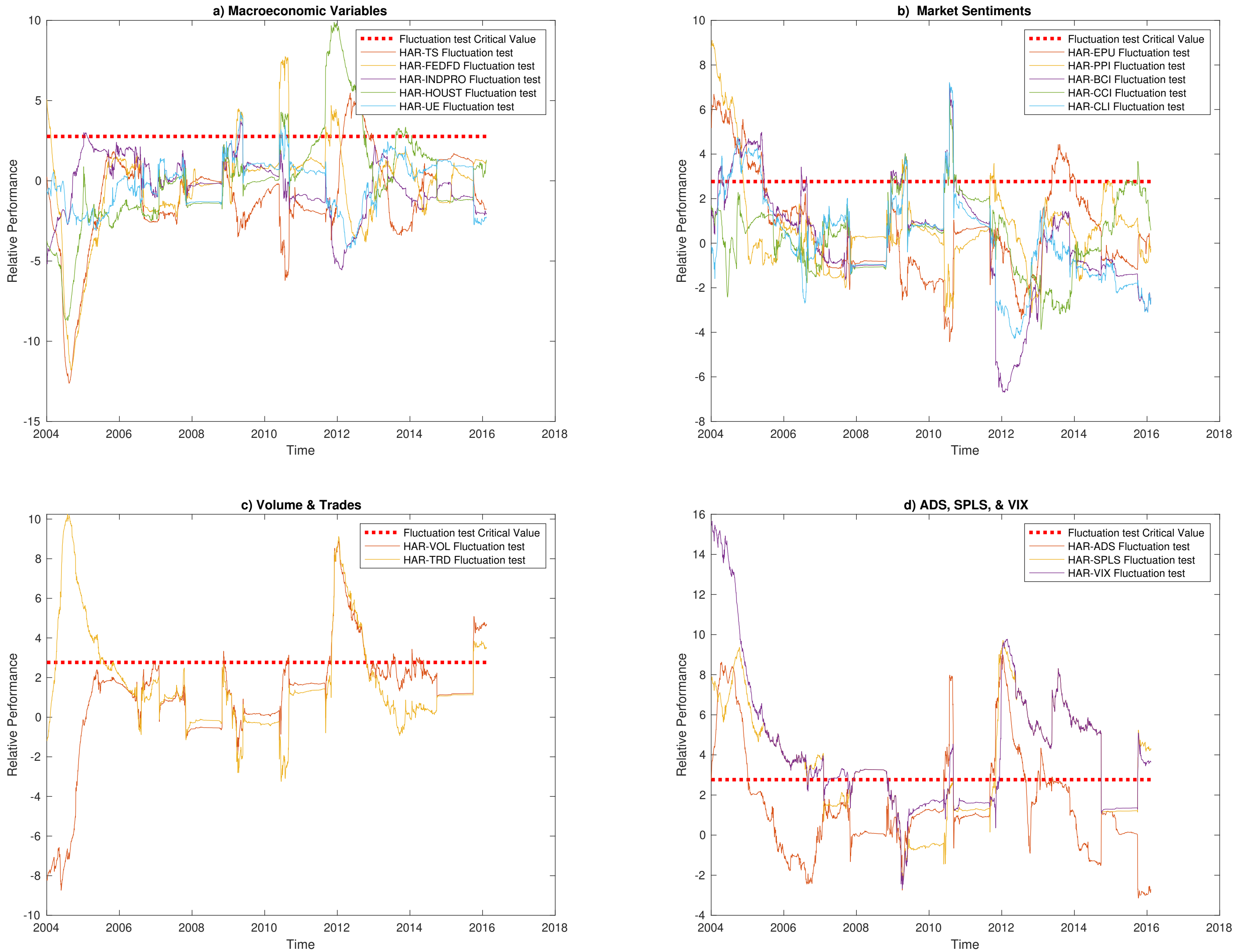



Figure 2.1: GR statistic tests plot for HARX models with 5% confidence level for forecasting $RV^{(d)}$ of SPY

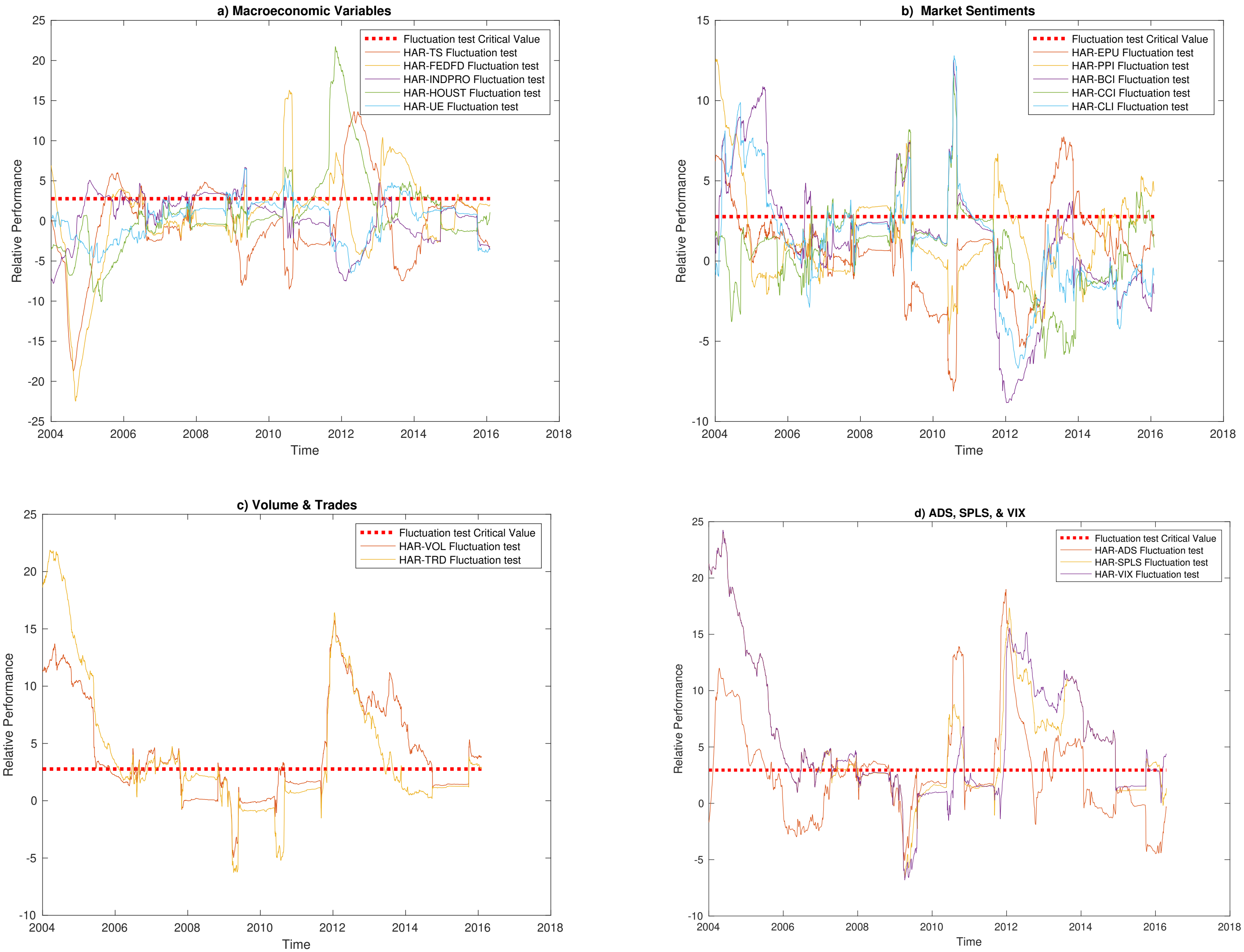



Figure 2.2: GR statistic tests plot for HARX models with 5% confidence level for forecasting $RV^{(w)}$ of SPY

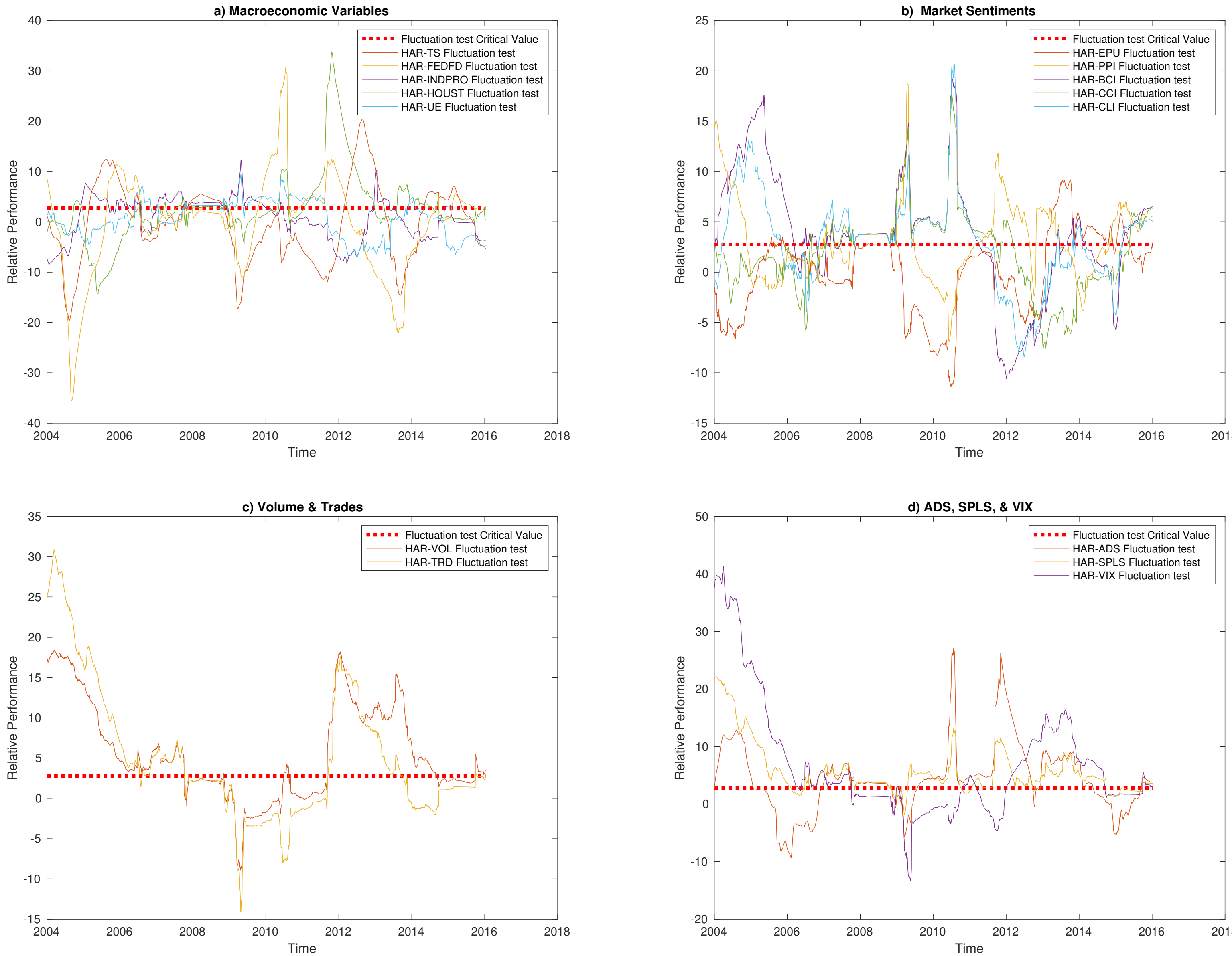



Figure 2.3: GR statistic tests plot for HARX models with 5% confidence level for forecasting $RV^{(m)}$ of SPY

Figures (2.1), (2.2), and (2.3) show the significance of the forecast gains of HARX, with each explanatory variable ($X$), for the daily, weekly, and monthly forecasts, respectively. The figures reveal when the HARX models outperform the HAR model. It is evident that the overall level of significance falls during the crisis period and increases afterwards. It falls around 2014 but starts to rise again by the end of 2015. The decline in significance around 2014 can be explained by economic trends not reflected explicitly in any of the HARX models, such as the plunge in oil prices and the dollar's appreciation.

Most macroeconomic variables yield significant long-term forecast gains over the covered years, although it generally falls during the crisis. The window over which the HARX has significant outperformance increases with the forecast horizon, including the crisis period for the monthly forecast. In particular, as leading indicators, both TS and HOUST show predictability for market volatility, especially directly after the crisis. Similarly, INDPRO, a coincident indicator, and UE, a lagging indicator, yield significant forecast gains for market volatility even during the crisis for the weekly and monthly forecast. Another lagging indicator, the PPI, shows significant forecast gains before and towards the end of the crisis period. ADS, a daily indicator of the current business environment, shows significant weekly and monthly forecast improvement mainly after the crisis.

The significant-high gains observed in the weekly and monthly forecasts using macroeconomic factors confirm that the latter can explain long-term variation in the stock market. We have shown that this observation holds not only for calendar months volatility, as used in previous literature, but also at a higher frequency level, i.e. for one week, $RV_{t+5}^{(w)}$, or one month, $RV_{t+22}^{(m)}$, ahead from any given day, t.

Furthermore, market sentiments yield significant forecast gains for most of the sample period, with lower gains during the crisis. BCI and CLI yield the highest significant gains over the majority of the sample period for the weekly and monthly forecast, while VIX shows higher significance for the daily forecast. Moreover, we find evidence of the SIAH using SPY before and after the crisis period only and more significant for weekly and monthly forecast horizons. The number of trades shows higher significant forecast gains than the volume before the crisis, while the latter yields higher significant gains after the crisis. Lastly, the forecast gains of using combined information (SPLS factors) are significant for most of the sample period, particularly for the long-term forecasting horizon.

## 2.6 Conclusion

This study examines the value-added from utilising information for various financial-economic indicators observed at different frequencies. We do so by adding "X" variable(s) to the HAR model, hence, HARX. We consider macroeconomic indicators, market sentiments, and activity measures that reflect the economic and market conditions. The findings reveal that economic variables include essential information beyond the lagged and historical average of volatility that significantly improves the RV's forecast. The market and economic conditions are particularly valuable for individual stocks during the crisis. The forecast gains are higher when forecasting long-term volatility than short-term ones.

One challenge of studying the relationship between financial markets and macroeconomy is the frequency of the variables. While the former's data are available at a high-frequency level, the latter's are recorded at a lower frequency. Hence, researchers suggest different methods to handle the frequency mismatch. One approach that can be used to study a LF variable's effect on a HF variable is employing the long-known Chow-Lin extrapolation method (Chow and Lin, 1971). A recent approach is the reverse MIDAS by Foroni, Guérin, and Marcellino (2018). The former extrapolates the LF into the HF series, allowing the monthly variable's value to fluctuate within the month while its coefficient in the HARX model is constant in a given estimation window. However, the latter approach keeps the value of the LF variable constant during the month but allows its coefficient to vary based on the day of the month. There is no significant difference in forecasting gains in the long-term forecasts while favouring the extrapolation method in the daily forecasts as it incorporates daily variations.

The forecast gains of including macroeconomic variables in the HARX model are more pronounced for long-term forecasting (weekly and monthly) than one day ahead. While unemployment and industrial production growth rates yield significant forecasting gains for most of the sample period, the housing starts growth rate, a leading indicator, showed more persistent significance after the financial crisis. Similarly, most market sentiments yield significant forecasting gains, particularly leading indicators. The CLI (VIX) produces significant forecasting gains for most of the OOS period for the weekly and monthly (daily) forecast. Also, the predictive power of volume and number of trades for RV is significant at all forecast horizons but not during a recession.

This study shows that incorporating financial and macroeconomic indicators in forecasting the RV produces significant forecasting gains. We note that combining the information using SPLS components in the HARX framework has the lowest average forecast errors over the sample period and is particularly beneficial for long-term investors during an unprecedented time, such as the financial crisis. Hence, a study of the economic gains from incorporating this information in portfolio risk management would be interesting to examine for future research. Furthermore, speaking of blended information, it is of interest to develop a resilient economic-financial conditions index robust to periods of high uncertainty such as financial crises or pandemics. Ideally, such an index would reflect the overall environment by including new information and adjusting the measurements accordingly.

# Appendix

## 2.A Tables

Table 2.5: Descriptive Statistics of Stocks

| | Ticker | Mean | Median | Minimum | Maximum | St. Dev. |
|---|---|---|---|---|---|---|
| Amazon.com Inc | AMZN | 8.284 | 3.344 | 0.225 | 229.244 | 14.846 |
| Best Buy Co. Inc. | BBY | 5.760 | 3.110 | 0.244 | 1439.571 | 23.491 |
| Gap (The) | GPS | 2.966 | 1.452 | 0.143 | 145.706 | 5.015 |
| Interpublic Group | IPG | 2.974 | 1.478 | 0.060 | 170.100 | 6.221 |
| Marriott International | MAR | 5.088 | 2.830 | 0.146 | 178.723 | 7.586 |
| McDonald's | MCD | 3.121 | 1.573 | 0.156 | 103.477 | 4.938 |
| News Corporation NWS | FOX | 5.524 | 2.632 | 0.235 | 615.148 | 13.956 |
| The Home Depot | HD | 3.537 | 1.782 | 0.154 | 104.578 | 5.364 |
| Time Warner Inc. | TWX | 2.164 | 1.090 | 0.087 | 161.156 | 4.333 |
| Walt Disney | DIS | 4.099 | 1.722 | 0.163 | 148.875 | 7.289 |
| **Consumer Discretionary Sector Avg.** | | **4.352** | **2.101** | **0.161** | **329.658** | **9.304** |
| Avon Products | AVP | 5.098 | 2.159 | 0.140 | 342.665 | 11.720 |
| Brown-Forman Corp. | BFb | 1.920 | 1.152 | 0.074 | 240.414 | 4.861 |
| Coca-Cola | KO | 3.087 | 1.497 | 0.126 | 83.955 | 4.728 |
| Costco | COST | 2.679 | 1.505 | 0.158 | 62.689 | 3.863 |
| Estee Lauder Cos. | EL | 1.566 | 0.856 | 0.061 | 100.855 | 2.745 |
| Kimberly-Clark | KMB | 1.561 | 0.836 | 0.046 | 58.808 | 2.535 |
| PepsiCo Inc. | PEP | 1.693 | 0.828 | 0.083 | 132.093 | 3.638 |
| Procter & Gamble | PG | 1.492 | 0.766 | 0.101 | 79.549 | 2.907 |
| Unilever | UL | 1.302 | 0.646 | 0.078 | 25.319 | 2.054 |
| Wal-Mart | WMT | 2.045 | 0.976 | 0.090 | 71.485 | 3.277 |
| **Consumer Staples Sector Avg.** | | **2.244** | **1.122** | **0.096** | **119.783** | **4.233** |

This table reports the descriptive statistics of daily Realised Variance (RV) for each stock over the period 01/2000 to 12/2016.

Table 2.5: Descriptive Statistics of Stocks (cont.)

| | Ticker | Mean | Median | Minimum | Maximum | St. Dev. |
|---|---|---|---|---|---|---|
| Baker Huges | BHI | 4.312 | 2.699 | 0.318 | 155.091 | 6.361 |
| Chesapeaks Energy | CHK | 5.609 | 3.584 | 0.321 | 218.401 | 7.742 |
| Chevron Corporation | CVX | 13.008 | 4.537 | 0.418 | 2216.263 | 41.965 |
| Devon Energy Corp | DVN | 2.113 | 1.263 | 0.112 | 137.535 | 4.181 |
| Exxon Mobil | XOM | 4.595 | 2.747 | 0.274 | 136.649 | 7.012 |
| Halliburton Co. | HAL | 6.305 | 3.579 | 0.229 | 265.432 | 10.443 |
| Hess Corporation | HES | 4.474 | 2.570 | 0.211 | 271.511 | 8.675 |
| Occidental Petroleum | OXY | 3.743 | 2.179 | 0.208 | 161.023 | 6.349 |
| Sunoco Inc. | APA | 10.216 | 3.338 | 0.267 | 1633.848 | 42.946 |
| Williams Cos. | WMB | 1.987 | 1.141 | 0.107 | 141.130 | 3.955 |
| **Energy Sector Avg.** | | **5.636** | **2.764** | **0.246** | **533.688** | **13.963** |
| Allstate Corp | ALL | 3.188 | 1.173 | 0.098 | 255.375 | 8.091 |
| American Express | AXP | 4.020 | 1.465 | 0.088 | 299.968 | 9.195 |
| Bank of America | BAC | 5.387 | 1.772 | 0.085 | 406.683 | 16.736 |
| Citigroup Inc. | C | 5.086 | 1.771 | 0.156 | 1059.901 | 23.703 |
| Goldman Sachs Group | GS | 6.742 | 2.110 | 0.137 | 975.858 | 27.240 |
| JPMorgan Chase | JPM | 4.186 | 1.757 | 0.153 | 400.346 | 11.978 |
| Morgan Stanley | MS | 4.615 | 1.770 | 0.114 | 252.877 | 10.848 |
| The Bank of New York Mellon | BK | 8.191 | 2.797 | 0.236 | 1732.782 | 46.295 |
| Travelers Group Inc | TRV | 2.968 | 1.186 | 0.102 | 263.929 | 7.866 |
| Wells Fargo | WFC | 4.302 | 1.330 | 0.104 | 226.609 | 12.139 |
| **Financials Sector Avg.** | | **4.869** | **1.713** | **0.127** | **587.433** | **17.409** |

This table reports the descriptive statistics of daily Realised Variance (RV) for each stock over the period 01/2000 to 12/2016.

Table 2.5: Descriptive Statistics of Stocks (cont.)

| **Stock** | **Ticker** | **Mean** | **Median** | **Minimum** | **Maximum** | **St. Deviation** |
|---|---|---|---|---|---|---|
| Abbott Laboratories | ABT | 2.168 | 1.226 | 0.119 | 65.347 | 3.045 |
| Amgen Inc | AMGN | 3.542 | 1.835 | 0.203 | 94.858 | 5.259 |
| Boston Scientific | BSX | 5.010 | 2.908 | 0.200 | 151.610 | 7.026 |
| Gilead Sciences | GILD | 6.308 | 2.839 | 0.198 | 187.286 | 10.469 |
| Humana Inc. | HUM | 6.679 | 2.609 | 0.240 | 157.529 | 11.366 |
| Johnson & Johnson | JNJ | 1.385 | 0.692 | 0.076 | 179.016 | 3.482 |
| Medtronic Inc | MDT | 2.311 | 1.236 | 0.109 | 181.758 | 4.348 |
| Merck | MRK | 2.431 | 1.359 | 0.135 | 223.255 | 5.227 |
| Pzer | PFE | 2.332 | 1.382 | 0.150 | 62.697 | 3.224 |
| United Health Group | UNH | 3.407 | 1.745 | 0.129 | 225.956 | 6.883 |
| **Health Care Sector Avg.** | | **3.557** | **1.783** | **0.156** | **152.931** | **6.033** |
| Boeing | BA | 2.812 | 1.602 | 0.086 | 55.570 | 3.900 |
| Caterpillar | CAT | 3.239 | 1.873 | 0.185 | 105.908 | 4.889 |
| Cummins Inc. | CMI | 4.905 | 2.566 | 0.157 | 199.866 | 9.070 |
| General Dynamics | GD | 2.237 | 1.281 | 0.081 | 63.282 | 3.259 |
| General Electric | GE | 3.020 | 1.303 | 0.108 | 180.389 | 6.982 |
| Honeywell Int'l Inc. | HON | 3.253 | 1.609 | 0.104 | 268.331 | 6.513 |
| Minnesota Mining & Mfg Co | MMM | 4.844 | 3.030 | 0.285 | 150.858 | 6.413 |
| Southwest Airlines | LUV | 1.853 | 1.008 | 0.082 | 91.955 | 3.216 |
| United Parcel Service | UPS | 1.649 | 0.851 | 0.081 | 216.939 | 4.140 |
| United Technologies | UTX | 2.300 | 1.211 | 0.104 | 75.915 | 3.793 |
| **Industrials Sector Avg.** | | **3.011** | **1.633** | **0.127** | **140.901** | **5.217** |

This table reports the descriptive statistics of daily Realised Variance (RV) for each stock over the period 01/2000 to 12/2016.

Table 2.5: Descriptive Statistics of Stocks (cont.)

| Stock | Ticker | Mean | Median | Minimum | Maximum | St. Deviation |
|---|---|---|---|---|---|---|
| Apple Inc. | AAPL | 5.292 | 2.578 | 0.079 | 126.172 | 7.806 |
| Cisco Systems | CSCO | 4.275 | 1.986 | 0.184 | 87.833 | 6.707 |
| Dell Inc. | EBAY | 6.511 | 2.782 | 0.202 | 236.419 | 12.762 |
| EMC Corp. | YAHOO | 4.302 | 2.267 | 0.122 | 156.740 | 7.066 |
| Hewlett-Packard | HPQ | 2.026 | 0.986 | 0.102 | 71.293 | 3.527 |
| IBM | IBM | 4.075 | 2.038 | 0.154 | 89.885 | 5.754 |
| Intel Corp. | INTC | 2.679 | 1.416 | 0.083 | 62.386 | 3.854 |
| Microsoft | MSFT | 5.021 | 2.174 | 0.129 | 123.804 | 8.381 |
| Oracle Corp. | ORCL | 6.548 | 2.864 | 0.299 | 276.588 | 13.732 |
| Xerox Corp. | XRX | 7.419 | 3.100 | 0.240 | 220.822 | 12.760 |
| **Information Technology Sector Avg.** | | **4.815** | **2.219** | **0.159** | **145.194** | **8.235** |
| AK Steel Holding Corp | AKS | 16.137 | 10.585 | 0.872 | 559.612 | 21.802 |
| Alcoa | ARNC | 5.288 | 3.070 | 0.339 | 291.089 | 9.601 |
| Dow Chemical | DOW | 2.761 | 1.491 | 0.100 | 83.487 | 4.076 |
| DuPont | DD | 3.976 | 2.039 | 0.146 | 216.937 | 7.353 |
| Freeport-McMoran | FCX | 8.034 | 4.327 | 0.317 | 188.580 | 12.279 |
| International Paper | IP | 4.290 | 2.103 | 0.162 | 171.959 | 7.500 |
| Newmont Mining | NEM | 5.098 | 3.435 | 0.470 | 109.879 | 6.395 |
| Nucor Corp. | NUE | 4.910 | 2.754 | 0.334 | 266.824 | 10.582 |
| United States Steel Corp. | X | 3.699 | 2.043 | 0.252 | 131.603 | 5.916 |
| Weyerhauser Co | WY | 8.696 | 5.557 | 0.736 | 344.652 | 12.579 |
| **Materials Sector Avg.** | | **6.289** | **3.740** | **0.373** | **236.462** | **9.808** |

This table reports the descriptive statistics of daily Realised Variance (RV) for each stock over the period 01/2000 to 12/2016.

Table 2.5: Descriptive Statistics of Stocks (cont.)

| **Stock** | **Ticker** | **Mean** | **Median** | **Minimum** | **Maximum** | **St. Deviation** |
|---|---|---|---|---|---|---|
| Ameren Corp | AEE | 1.828 | 1.061 | 0.109 | 113.488 | 3.453 |
| American Electric Power | AEP | 2.444 | 1.208 | 0.151 | 207.378 | 6.592 |
| Constellation Energy Group | CEG | 4.497 | 1.825 | 0.168 | 772.838 | 20.605 |
| Duke Energy | DUK | 2.433 | 1.182 | 0.051 | 189.935 | 6.070 |
| Entergy Corp. | ETR | 2.230 | 1.177 | 0.085 | 118.560 | 4.028 |
| Exelon Corp. | EXC | 2.635 | 1.429 | 0.158 | 130.875 | 4.791 |
| PG&E Corp. | PCG | 3.402 | 1.668 | 0.160 | 411.055 | 8.590 |
| Progress Energy, Inc. | CNP | 4.162 | 1.265 | 0.149 | 1532.118 | 27.721 |
| Public Serv. Enterprise Inc. | PEG | 2.461 | 1.376 | 0.148 | 122.428 | 4.827 |
| The Southern Company | SO | 1.744 | 0.937 | 0.092 | 97.041 | 2.773 |
| **Utilities Sector Avg.** | | **2.784** | **1.313** | **0.127** | **369.572** | **8.945** |
| American Tower Corp A | AMT | 9.628 | 1.957 | 0.165 | 1048.656 | 32.874 |
| AT&T | T | 2.311 | 1.162 | 0.100 | 59.568 | 3.228 |
| BT Group plc (ADR) | BT | 2.052 | 1.063 | 0.082 | 65.965 | 3.608 |
| CenturyTel Inc | CTL | 2.720 | 1.346 | 0.152 | 279.942 | 6.856 |
| Frontier Communications | FTR | 5.426 | 2.933 | 0.232 | 1738.725 | 27.431 |
| Qwest Communication Int | CHL | 20.499 | 10.917 | 0.242 | 1159.384 | 35.569 |
| Sprint Nextel Corp | LVLT | 2.655 | 1.184 | 0.108 | 141.846 | 4.767 |
| Telefonica S.A. (ADR) | TEF | 1.820 | 1.021 | 0.117 | 162.204 | 3.601 |
| Verizon Communications | VZ | 1.986 | 0.926 | 0.110 | 70.936 | 3.063 |
| Vodafone Group Plc (ADR) | VOD | 2.336 | 1.162 | 0.122 | 102.221 | 3.860 |
| **Telecom Sector Avg.** | | **5.143** | **2.367** | **0.143** | **482.945** | **12.486** |
| SPDR ETF | SPY | 1.037 | 0.485 | 0.013 | 59.863 | 3.453 |

This table reports the descriptive statistics of daily Realised Variance (RV) for each stock over the period 01/2000 to 12/2016.

Table 2.6: Descriptive Statistics

| **Variable** | **Functional form** | **Mean** | **Median** | **Minimum** | **Maximum** | **Std. dev** | **Skewness** | **Kurtosis** |
|---|---|---|---|---|---|---|---|---|
| ADS | Level | -0.331 | -0.192 | -4.234 | 0.958 | 0.759 | -2.418 | 10.907 |
| BCI | Level | 99.807 | 99.852 | 96.022 | 101.997 | 1.069 | -0.919 | 4.893 |
| CCI | Level | 99.657 | 99.809 | 96.707 | 102.702 | 1.376 | -0.221 | 2.671 |
| CLI | Level | 99.751 | 100.005 | 94.614 | 101.892 | 1.409 | -1.351 | 5.339 |
| FEDFD | Log of Squared | -0.796 | -0.020 | -6.438 | 3.900 | 3.186 | 0.013 | 1.394 |
| HOUST | % Growth | 0.199 | 0.109 | -18.830 | 24.922 | 8.294 | 0.193 | 2.979 |
| INDPRO | % Growth | 0.046 | 0.098 | -4.337 | 1.459 | 0.676 | -2.029 | 12.843 |
| PPI | Level | 171.332 | 174.850 | 128.100 | 208.300 | 26.372 | -0.251 | 1.644 |
| TS | Log of Squared | 0.791 | 1.493 | -9.210 | 2.696 | 1.998 | -1.945 | 6.945 |
| UE | % Growth | 0.116 | 0.000 | -7.463 | 8.000 | 2.721 | 0.499 | 3.325 |
| VIX | Squared | 491.688 | 335.989 | 97.812 | 6538.340 | 539.523 | 4.562 | 33.363 |

This table reports the descriptive statistics of the variables in the specified functional form over the period 01/2000 to 12/2016.

Table 2.7: a) In-sample estimation of HARX (eq. (2.2)) using SPY data for $y = RV^{(d)}$, Pre-Crisis Period

| | $\hat{\beta}_0^{(d)}$ | | $\hat{\beta}_d^{(d)}$ | | $\hat{\beta}_w^{(d)}$ | | $\hat{\beta}_m^{(d)}$ | | $\hat{\beta}_x^{(d)}$ | | **Adj. $R^2$** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.131 | *** | 0.236 | ** | 0.241 | ** | 0.393 | *** | | | 36.598% |
| | (0.042) | | (0.117) | | (0.119) | | (0.115) | | | | |
| HAR-ue | 0.148 | *** | 0.236 | ** | 0.234 | ** | 0.379 | *** | 0.055 | | 36.753% |
| | (0.053) | | (0.116) | | (0.117) | | (0.106) | | (0.064) | | |
| HAR-ind | 0.171 | ** | 0.235 | ** | 0.240 | ** | 0.375 | *** | -3.226 | | 36.673% |
| | (0.068) | | (0.117) | | (0.118) | | (0.109) | | (3.157) | | |
| HAR-hst | 0.132 | *** | 0.236 | ** | 0.234 | ** | 0.401 | *** | -0.015 | | 36.607% |
| | (0.042) | | (0.116) | | (0.119) | | (0.115) | | (0.011) | | |
| HAR-ts | 0.131 | *** | 0.236 | ** | 0.241 | ** | 0.393 | *** | 0.002 | | 36.563% |
| | (0.042) | | (0.116) | | (0.119) | | (0.117) | | (0.012) | | |
| HAR-cci | -2.445 | | 0.236 | * | 0.240 | ** | 0.391 | *** | 0.026 | | 36.578% |
| | (4.633) | | (0.121) | | (0.119) | | (0.097) | | (0.046) | | |
| HAR-cli | 6.508 | * | 0.235 | * | 0.242 | ** | 0.360 | *** | -0.064 | | 36.714% |
| | (3.936) | | (0.121) | | (0.118) | | (0.099) | | (0.039) | | |
| HAR-bci | 8.055 | | 0.234 | * | 0.241 | ** | 0.352 | *** | -0.079 | | 36.766% |
| | (5.055) | | (0.122) | | (0.119) | | (0.089) | | (0.050) | | |
| HAR-ppi | 1.524 | *** | 0.234 | ** | 0.239 | ** | 0.339 | *** | -0.009 | *** | 36.864% |
| | (0.382) | | (0.118) | | (0.115) | | (0.102) | | (0.002) | | |
| HAR-ads | 0.124 | *** | 0.236 | ** | 0.241 | ** | 0.379 | *** | -0.076 | | 36.608% |
| | (0.038) | | (0.117) | | (0.118) | | (0.116) | | (0.106) | | |
| HAR-fd | 0.092 | * | 0.236 | ** | 0.240 | ** | 0.388 | *** | 0.023 | | 36.596% |
| | (0.051) | | (0.117) | | (0.119) | | (0.113) | | (0.026) | | |
| HAR-vol | 0.185 | * | 0.238 | ** | 0.242 | ** | 0.380 | *** | -0.130 | | 36.591% |
| | (0.106) | | (0.112) | | (0.118) | | (0.131) | | (0.169) | | |
| HAR-trd | 0.251 | ** | 0.237 | ** | 0.242 | ** | 0.354 | *** | -0.212 | ** | 36.727% |
| | (0.099) | | (0.111) | | (0.116) | | (0.127) | | (0.097) | | |
| HAR-vix | -0.120 | ** | 0.203 | ** | 0.172 | * | 0.143 | | 0.503 | *** | 39.368% |
| | (0.060) | | (0.103) | | (0.089) | | (0.096) | | (0.112) | | |

*** 1% significance, ** 5% significance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.7: b) In-sample estimation of HARX (eq. (2.2)) using SPY data for $y = RV^{(d)}$, Crisis Period

| | $\hat{\beta}_0^{(d)}$ | | $\hat{\beta}_d^{(d)}$ | $\hat{\beta}_w^{(d)}$ | | $\hat{\beta}_m^{(d)}$ | $\hat{\beta}_x^{(d)}$ | | **Adj. $R^2$** |
|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.165 | * | 0.221 | 0.493 | ** | 0.194 | | | 54.048% |
| | (0.097) | | (0.138) | (0.199) | | (0.135) | | | |
| HAR-ue | 0.121 | | 0.218 | 0.499 | ** | 0.153 | 0.140 | | 54.173% |
| | (0.078) | | (0.138) | (0.203) | | (0.151) | (0.099) | | |
| HAR-ind | 0.184 | * | 0.217 | 0.495 | ** | 0.157 | -5.142 | * | 54.224% |
| | (0.101) | | (0.137) | (0.203) | | (0.146) | (3.078) | | |
| HAR-hst | 0.080 | | 0.219 | 0.502 | ** | 0.161 | -0.089 | | 54.072% |
| | (0.082) | | (0.138) | (0.203) | | (0.159) | (0.114) | | |
| HAR-ts | 0.157 | * | 0.221 | 0.493 | ** | 0.193 | 0.012 | | 54.006% |
| | (0.091) | | (0.138) | (0.199) | | (0.137) | (0.019) | | |
| HAR-cci | 11.414 | | 0.219 | 0.498 | *** | 0.173 | -0.114 | | 54.079% |
| | (7.998) | | (0.155) | (0.188) | | (0.153) | (0.080) | | |
| HAR-cli | 5.195 | | 0.220 | 0.499 | *** | 0.174 | -0.050 | | 54.053% |
| | (4.711) | | (0.155) | (0.188) | | (0.156) | (0.046) | | |
| HAR-bci | 20.314 | * | 0.216 | 0.506 | *** | 0.127 | -0.202 | * | 54.246% |
| | (11.836) | | (0.155) | (0.189) | | (0.164) | (0.118) | | |
| HAR-ppi | -1.279 | | 0.221 | 0.489 | *** | 0.201 | 0.008 | | 54.039% |
| | (1.146) | | (0.155) | (0.188) | | (0.151) | (0.006) | | |
| HAR-ads | 0.086 | | 0.218 | 0.498 | ** | 0.149 | -0.177 | | 54.212% |
| | (0.074) | | (0.138) | (0.203) | | (0.151) | (0.114) | | |
| HAR-fd | 0.184 | * | 0.220 | 0.486 | ** | 0.208 | 0.036 | * | 54.082% |
| | (0.097) | | (0.138) | (0.197) | | (0.136) | (0.022) | | |
| HAR-vol | -0.813 | *** | 0.122 | 0.476 | ** | 0.160 | 0.639 | *** | 55.274% |
| | (0.265) | | (0.146) | (0.190) | | (0.143) | (0.203) | | |
| HAR-trd | -0.376 | ** | 0.161 | 0.493 | *** | 0.131 | 0.168 | ** | 54.728% |
| | (0.188) | | (0.147) | (0.190) | | (0.147) | (0.069) | | |
| HAR-vix | -0.589 | *** | 0.093 | 0.443 | ** | -0.357 | 0.907 | *** | 58.407% |
| | (0.227) | | (0.147) | (0.210) | | (0.228) | (0.257) | | |

*** 1% significance, ** 5% siginificance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.7: c) In-sample estimation of HARX (eq. (2.2)) using SPY data for $y = RV^{(d)}$, Post-Crisis Period

| | $\hat{\beta}_0^{(d)}$ | | $\hat{\beta}_d^{(d)}$ | | $\hat{\beta}_w^{(d)}$ | $\hat{\beta}_m^{(d)}$ | | $\hat{\beta}_x^{(d)}$ | | **Adj. $R^2$** |
|---|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.103 | *** | 0.422 | ** | 0.169 | 0.204 | ** | | | 35.336% |
| | (0.036) | | (0.197) | | (0.153) | (0.092) | | | | |
| HAR-ue | 0.127 | *** | 0.422 | ** | 0.168 | 0.202 | ** | 0.058 | | 35.329% |
| | (0.030) | | (0.196) | | (0.152) | (0.092) | | (0.059) | | |
| HAR-ind | 0.087 | * | 0.422 | ** | 0.168 | 0.202 | ** | 1.685 | | 35.315% |
| | (0.046) | | (0.196) | | (0.152) | (0.092) | | (1.863) | | |
| HAR-hst | 0.117 | *** | 0.422 | ** | 0.167 | 0.202 | ** | -0.030 | | 35.338% |
| | (0.031) | | (0.195) | | (0.152) | (0.092) | | (0.029) | | |
| HAR-ts | 0.083 | | 0.422 | ** | 0.168 | 0.205 | ** | 0.014 | | 35.298% |
| | (0.055) | | (0.197) | | (0.153) | (0.091) | | (0.022) | | |
| HAR-cci | 5.128 | * | 0.420 | ** | 0.168 | 0.175 | * | -0.050 | * | 35.536% |
| | (2.865) | | (0.193) | | (0.153) | (0.096) | | (0.029) | | |
| HAR-cli | 8.241 | ** | 0.420 | ** | 0.170 | 0.177 | * | -0.081 | ** | 35.465% |
| | (3.290) | | (0.197) | | (0.156) | (0.093) | | (0.033) | | |
| HAR-bci | 4.557 | | 0.422 | ** | 0.169 | 0.195 | ** | -0.045 | | 35.336% |
| | (3.819) | *** | (0.198) | | (0.157) | (0.091) | | (0.038) | | |
| HAR-ppi | 0.110 | | 0.422 | ** | 0.169 | 0.204 | ** | 0.000 | | 35.293% |
| | (0.499) | | (0.198) | | (0.157) | (0.090) | | (0.002) | | |
| HAR-ads | 0.105 | *** | 0.422 | ** | 0.168 | 0.205 | ** | 0.015 | | 35.295% |
| | (0.033) | | (0.196) | | (0.153) | (0.092) | | (0.042) | | |
| HAR-fd | 0.062 | | 0.422 | ** | 0.169 | 0.201 | ** | -0.011 | | 35.307% |
| | (0.056) | | (0.197) | | (0.154) | (0.092) | | (0.010) | | |
| HAR-vol | -0.382 | ** | 0.241 | | 0.082 | 0.072 | | 0.619 | ** | 40.999% |
| | (0.188) | | (0.157) | | (0.087) | (0.106) | | (0.250) | | |
| HAR-trd | -0.458 | * | 0.181 | | 0.050 | 0.097 | | 0.223 | * | 40.556% |
| | (0.277) | | (0.247) | | (0.084) | (0.108) | | (0.119) | | |
| HAR-vix | -0.138 | ** | 0.284 | | 0.080 | -0.333 | | 0.690 | *** | 41.753% |
| | (0.065) | | (0.184) | | (0.098) | (0.204) | | (0.218) | | |

*** 1% significance, ** 5% significance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.8: a) In-sample estimation of HARX (eq. (2.3)) using SPY data for $y = RV^{(w)}$, Pre-Crisis Period

| | $\hat{\beta}_0^{(w)}$ | | $\hat{\beta}_d^{(w)}$ | | $\hat{\beta}_w^{(w)}$ | | $\hat{\beta}_m^{(w)}$ | | $\hat{\beta}_x^{(w)}$ | | **Adj.** $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.189 | *** | 0.096 | * | 0.324 | *** | 0.392 | *** | | | 52.638% |
| | (0.052) | | (0.050) | | (0.082) | | (0.111) | | | | |
| HAR-ue | 0.207 | *** | 0.095 | * | 0.317 | *** | 0.378 | *** | 0.060 | | 53.001% |
| | (0.065) | | (0.050) | | (0.086) | | (0.104) | | (0.074) | | |
| HAR-ind | 0.247 | *** | 0.094 | * | 0.323 | *** | 0.366 | *** | -4.726 | | 53.027% |
| | (0.076) | | (0.050) | | (0.081) | | (0.108) | | (3.305) | | |
| HAR-hst | 0.190 | *** | 0.095 | * | 0.315 | *** | 0.403 | *** | -0.021 | | 52.766% |
| | (0.052) | | (0.050) | | (0.078) | | (0.104) | | (0.015) | | |
| HAR-ts | 0.189 | *** | 0.096 | * | 0.324 | *** | 0.392 | *** | 0.000 | | 52.611% |
| | (0.052) | | (0.050) | | (0.082) | | (0.111) | | (0.017) | | |
| HAR-cci | -3.623 | | 0.096 | * | 0.324 | *** | 0.389 | *** | 0.038 | | 52.676% |
| | (6.415) | | (0.050) | | (0.081) | | (0.106) | | (0.064) | | |
| HAR-cli | 9.287 | | 0.094 | * | 0.326 | *** | 0.346 | *** | -0.091 | | 53.152% |
| | (5.749) | | (0.049) | | (0.079) | | (0.121) | | (0.057) | | |
| HAR-bci | 11.388 | ** | 0.094 | * | 0.325 | *** | 0.334 | *** | -0.111 | ** | 53.323% |
| | (5.214) | | (0.049) | | (0.080) | | (0.104) | | (0.052) | | |
| HAR-ppi | 2.188 | *** | 0.093 | * | 0.322 | *** | 0.315 | *** | -0.013 | *** | 53.694% |
| | (0.629) | | (0.048) | | (0.078) | | (0.108) | | (0.004) | | |
| HAR-ads | 0.180 | *** | 0.095 | * | 0.325 | *** | 0.375 | *** | -0.100 | | 52.750% |
| | (0.048) | | (0.050) | | (0.081) | | (0.118) | | (0.125) | | |
| HAR-fd | 0.133 | ** | 0.095 | * | 0.324 | *** | 0.386 | *** | 0.033 | | 52.735% |
| | (0.060) | | (0.051) | | (0.083) | | (0.107) | | (0.032) | | |
| HAR-vol | 0.345 | *** | 0.102 | ** | 0.329 | *** | 0.355 | *** | -0.378 | ** | 53.038% |
| | (0.105) | | (0.052) | | (0.082) | | (0.104) | | (0.175) | | |
| HAR-trd | 0.390 | *** | 0.097 | * | 0.326 | *** | 0.328 | *** | -0.036 | *** | 53.429% |
| | (0.104) | *** | (0.051) | | (0.080) | | (0.100) | | (0.011) | | |
| HAR-vix | -0.040 | | 0.066 | * | 0.262 | *** | 0.166 | | 0.456 | *** | 56.641% |
| | (0.065) | | (0.035) | | (0.081) | | (0.123) | | (0.135) | | |

*** 1% significance, ** 5% significance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.8: b) In-sample estimation of HARX (eq. (2.3)) using SPY data for $y = RV^{(w)}$, Crisis Period

| | $\hat{\beta}_0^{(w)}$ | | $\hat{\beta}_d^{(w)}$ | | $\hat{\beta}_w^{(w)}$ | | $\hat{\beta}_m^{(w)}$ | | $\hat{\beta}_x^{(w)}$ | | **Adj. $R^2$** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.263 | *** | 0.212 | *** | 0.344 | ** | 0.298 | ** | | | 64.863% |
| | (0.095) | | (0.048) | | (0.147) | | (0.118) | | | | |
| HAR-ue | 0.192 | ** | 0.208 | *** | 0.353 | ** | 0.230 | * | 0.229 | * | 65.484% |
| | (0.082) | | (0.047) | | (0.157) | | (0.136) | | (0.125) | | |
| HAR-ind | 0.296 | *** | 0.206 | *** | 0.348 | ** | 0.233 | * | -9.017 | ** | 65.798% |
| | (0.084) | | (0.047) | | (0.155) | | (0.134) | | (3.809) | | |
| HAR-hst | 0.134 | | 0.210 | *** | 0.357 | ** | 0.248 | ** | -0.135 | | 65.057% |
| | (0.129) | | (0.047) | | (0.152) | | (0.123) | | (0.127) | | |
| HAR-ts | 0.253 | *** | 0.212 | *** | 0.344 | ** | 0.296 | ** | 0.015 | | 64.837% |
| | (0.087) | | (0.048) | | (0.147) | | (0.118) | | (0.028) | | |
| HAR-cci | 18.453 | * | 0.210 | *** | 0.352 | ** | 0.263 | ** | -0.184 | * | 65.118% |
| | (10.584) | | (0.043) | | (0.143) | | (0.115) | | (0.107) | | |
| HAR-cli | 7.679 | | 0.210 | *** | 0.353 | ** | 0.267 | ** | -0.074 | | 64.985% |
| | (5.381) | | (0.043) | | (0.143) | | (0.116) | | (0.053) | | |
| HAR-bci | 31.012 | * | 0.205 | *** | 0.364 | ** | 0.196 | | -0.308 | * | 65.632% |
| | (16.481) | | (0.042) | | (0.147) | | (0.127) | | (0.165) | | |
| HAR-ppi | -2.596 | | 0.212 | *** | 0.335 | ** | 0.311 | *** | 0.016 | | 65.034% |
| | (2.090) | | (0.044) | | (0.136) | | (0.106) | | (0.012) | | |
| HAR-ads | 0.132 | | 0.207 | *** | 0.352 | ** | 0.223 | | -0.292 | ** | 65.642% |
| | (0.094) | | (0.047) | | (0.157) | | (0.137) | | (0.147) | | |
| HAR-fd | 0.283 | *** | 0.211 | *** | 0.336 | ** | 0.314 | *** | 0.041 | | 64.980% |
| | (0.103) | | (0.048) | | (0.143) | | (0.113) | | (0.031) | | |
| HAR-vol | -0.485 | * | 0.136 | *** | 0.331 | ** | 0.271 | ** | 0.489 | ** | 65.888% |
| | (0.281) | *** | (0.044) | | (0.144) | | (0.120) | | (0.203) | | |
| HAR-trd | -0.100 | | 0.172 | *** | 0.344 | ** | 0.255 | ** | 0.011 | * | 65.295% |
| | (0.165) | | (0.043) | | (0.148) | | (0.125) | *** | (0.006) | | |
| HAR-vix | -0.138 | | 0.144 | *** | 0.317 | * | 0.004 | | 0.484 | *** | 66.615% |
| | (0.115) | | (0.042) | | (0.167) | | (0.152) | | (0.128) | | |

*** 1% significance, ** 5% significance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.8: c) In-sample estimation of HARX (eq. (2.3)) using SPY data for $y = RV^{(w)}$, Post-Crisis Period

| | $\hat{\beta}_0^{(w)}$ | | $\hat{\beta}_d^{(w)}$ | | $\hat{\beta}_w^{(w)}$ | | $\hat{\beta}_m^{(w)}$ | | $\hat{\beta}_x^{(w)}$ | | **Adj.** $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.167 | *** | 0.232 | ** | 0.133 | * | 0.302 | *** | | | 32.982% |
| | (0.042) | | (0.112) | | (0.069) | | (0.076) | | | | |
| HAR-ue | 0.206 | *** | 0.231 | ** | 0.133 | * | 0.299 | *** | 0.093 | | 33.101% |
| | (0.050) | | (0.111) | | (0.070) | | (0.076) | | (0.094) | | |
| HAR-ind | 0.143 | *** | 0.232 | ** | 0.133 | * | 0.299 | *** | 2.521 | | 33.024% |
| | (0.052) | | (0.111) | | (0.070) | | (0.076) | | (3.212) | | |
| HAR-hst | 0.190 | *** | 0.231 | ** | 0.131 | * | 0.298 | *** | -0.049 | | 33.149% |
| | (0.046) | | (0.110) | | (0.070) | | (0.076) | | (0.048) | | |
| HAR-ts | 0.091 | | 0.232 | ** | 0.132 | ** | 0.305 | *** | 0.051 | | 33.060% |
| | (0.075) | | (0.109) | | (0.053) | | (0.050) | | (0.048) | | |
| HAR-cci | 8.105 | | 0.228 | ** | 0.132 | ** | 0.256 | *** | -0.080 | | 33.984% |
| | (5.838) | | (0.104) | | (0.054) | | (0.065) | | (0.058) | | |
| HAR-cli | 12.929 | * | 0.229 | ** | 0.135 | *** | 0.259 | *** | -0.127 | | 33.667% |
| | (7.834) | | (0.107) | | (0.051) | | (0.059) | | (0.078) | | |
| HAR-bci | 7.409 | | 0.231 | ** | 0.135 | *** | 0.287 | *** | -0.072 | | 33.137% |
| | (5.530) | | (0.109) | | (0.051) | | (0.052) | | (0.055) | | |
| HAR-ppi | 0.119 | | 0.232 | ** | 0.133 | ** | 0.302 | *** | 0.000 | | 32.938% |
| | (0.863) | | (0.109) | | (0.052) | | (0.050) | | (0.004) | | |
| HAR-ads | 0.170 | *** | 0.232 | ** | 0.133 | * | 0.303 | *** | 0.022 | | 32.945% |
| | (0.042) | | (0.112) | | (0.070) | | (0.076) | | (0.064) | | |
| HAR-fd | 0.124 | | 0.232 | ** | 0.134 | *** | 0.299 | *** | -0.011 | | 32.963% |
| | (0.086) | | (0.109) | | (0.051) | | (0.051) | | (0.019) | | |
| HAR-vol | -0.268 | | 0.070 | | 0.056 | | 0.184 | ** | 0.554 | ** | 40.813% |
| | (0.179) | | (0.051) | | (0.058) | | (0.078) | | (0.252) | | |
| HAR-trd | -0.275 | | 0.042 | | 0.039 | | 0.218 | *** | 0.018 | ** | 38.587% |
| | (0.168) | *** | (0.095) | | (0.065) | | (0.073) | | (0.008) | | |
| HAR-vix | -0.009 | | 0.131 | | 0.069 | * | -0.088 | | 0.501 | *** | 38.839% |
| | (0.037) | | (0.099) | | (0.040) | | (0.114) | | (0.129) | | |

*** 1% significance, ** 5% significance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.9: a) In-sample estimation of HARX (eq. (2.4)) using SPY data for $y = RV^{(m)}$, Pre-Crisis Period

| | $\hat{\beta}_0^{(m)}$ | | $\hat{\beta}_d^{(m)}$ | | $\hat{\beta}_w^{(m)}$ | | $\hat{\beta}_m^{(m)}$ | | $\hat{\beta}_x^{(m)}$ | | **Adj. $R^2$** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.313 | *** | 0.070 | ** | 0.253 | *** | 0.366 | *** | | | 50.834% |
| | (0.083) | | (0.034) | | (0.064) | | (0.093) | | | | |
| HAR-ue | 0.320 | *** | 0.070 | ** | 0.251 | *** | 0.361 | *** | 0.001 | | 50.875% |
| | (0.090) | | (0.034) | | (0.067) | | (0.092) | | (0.001) | | |
| HAR-ind | 0.391 | *** | 0.069 | ** | 0.251 | *** | 0.331 | *** | -0.063 | | 51.832% |
| | (0.114) | | (0.033) | | (0.063) | | (0.100) | | (0.049) | | |
| HAR-hst | 0.316 | *** | 0.070 | ** | 0.238 | *** | 0.385 | *** | -0.001 | | 51.429% |
| | (0.082) | | (0.033) | | (0.060) | | (0.082) | | (0.001) | | |
| HAR-ts | 0.314 | *** | 0.070 | ** | 0.253 | *** | 0.366 | *** | -0.004 | | 50.815% |
| | (0.088) | | (0.034) | | (0.062) | | (0.093) | | (0.025) | | |
| HAR-cci | -7.644 | | 0.070 | ** | 0.252 | *** | 0.360 | *** | 0.079 | | 51.191% |
| | (8.838) | | (0.034) | | (0.058) | | (0.083) | | (0.088) | | |
| HAR-cli | 14.266 | | 0.068 | ** | 0.256 | *** | 0.295 | ** | -0.139 | | 52.533% |
| | (9.723) | | (0.033) | | (0.057) | | (0.129) | | (0.097) | | |
| HAR-bci | 17.631 | ** | 0.067 | ** | 0.254 | *** | 0.275 | *** | -0.172 | ** | 53.125% |
| | (6.883) | | (0.034) | | (0.058) | | (0.095) | | (0.068) | | |
| HAR-ppi | 3.655 | *** | 0.066 | ** | 0.249 | *** | 0.238 | ** | -0.022 | *** | 54.870% |
| | (1.150) | | (0.031) | | (0.054) | | (0.096) | | (0.007) | | |
| HAR-ads | 0.302 | *** | 0.070 | ** | 0.254 | *** | 0.345 | *** | -0.122 | | 51.085% |
| | (0.077) | | (0.034) | | (0.063) | | (0.104) | | (0.169) | | |
| HAR-fd | 0.217 | ** | 0.070 | ** | 0.253 | *** | 0.354 | *** | 0.058 | | 51.329% |
| | (0.100) | | (0.034) | | (0.063) | | (0.085) | | (0.047) | | |
| HAR-vol | 0.666 | *** | 0.084 | ** | 0.263 | *** | 0.282 | *** | -0.857 | *** | 53.801% |
| | (0.138) | | (0.039) | | (0.066) | | (0.080) | | (0.230) | | |
| HAR-trd | 0.677 | *** | 0.073 | ** | 0.256 | *** | 0.249 | *** | -0.065 | *** | 54.472% |
| | (0.144) | | (0.035) | | (0.064) | | (0.079) | | (0.016) | | |
| HAR-vix | 0.104 | | 0.043 | ** | 0.197 | *** | 0.162 | | 0.414 | *** | 55.351% |
| | (0.078) | | (0.021) | | (0.064) | | (0.142) | | (0.160) | | |

*** 1% significance, ** 5% significance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.9: b) In-sample estimation of HARX (eq. (2.4)) using SPY data for $y = RV^{(m)}$, Crisis Period

| | $\hat{\beta}_0^{(m)}$ | | $\hat{\beta}_d^{(m)}$ | | $\hat{\beta}_w^{(m)}$ | | $\hat{\beta}_m^{(m)}$ | | $\hat{\beta}_x^{(m)}$ | | **Adj. $R^2$** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.509 | *** | 0.122 | *** | 0.348 | * | 0.249 | | | | 57.038% |
| | (0.173) | | (0.033) | | (0.180) | | (0.179) | ** | | | |
| HAR-ue | 0.353 | *** | 0.111 | *** | 0.369 | ** | 0.090 | | 0.005 | * | 61.496% |
| | (0.122) | | (0.028) | | (0.183) | | (0.247) | *** | (0.003) | | |
| HAR-ind | 0.596 | *** | 0.104 | *** | 0.359 | ** | 0.070 | | -0.247 | ** | 66.151% |
| | (0.155) | | (0.025) | | (0.155) | | (0.221) | * | (0.118) | | |
| HAR-hst | 0.180 | | 0.115 | *** | 0.383 | * | 0.117 | | -0.004 | | 58.952% |
| | (0.229) | | (0.031) | | (0.201) | | (0.261) | | (0.003) | | |
| HAR-ts | 0.481 | *** | 0.122 | *** | 0.348 | * | 0.241 | | 0.048 | | 57.110% |
| | (0.153) | | (0.032) | | (0.178) | | (0.180) | | (0.054) | | |
| HAR-cci | 35.711 | | 0.117 | *** | 0.363 | *** | 0.179 | | -0.357 | | 58.464% |
| | (23.154) | ** | (0.020) | | (0.124) | | (0.119) | | (0.233) | ** | |
| HAR-cli | 11.680 | | 0.119 | *** | 0.361 | *** | 0.202 | * | -0.112 | | 57.446% |
| | (9.511) | | (0.020) | | (0.125) | | (0.117) | | (0.094) | | |
| HAR-bci | 48.044 | | 0.111 | *** | 0.380 | *** | 0.093 | | -0.476 | | 59.347% |
| | (29.977) | | (0.020) | | (0.128) | | (0.153) | | (0.299) | | |
| HAR-ppi | -8.676 | | 0.120 | *** | 0.318 | *** | 0.288 | *** | 0.051 | | 59.753% |
| | (6.545) | | (0.020) | | (0.107) | | (0.085) | | (0.037) | | |
| HAR-ads | 0.218 | | 0.110 | *** | 0.366 | *** | 0.080 | | -0.652 | * | 62.126% |
| | (0.163) | | (0.022) | | (0.136) | | (0.175) | | (0.374) | | |
| HAR-fd | 0.529 | *** | 0.121 | *** | 0.337 | * | 0.271 | | 0.054 | | 57.328% |
| | (0.187) | | (0.033) | | (0.173) | | (0.169) | | (0.046) | | |
| HAR-vol | -0.163 | | 0.053 | | 0.337 | ** | 0.224 | | 0.442 | | 58.112% |
| | (0.365) | | (0.044) | | (0.162) | | (0.170) | | (0.310) | | |
| HAR-trd | 0.232 | | 0.091 | *** | 0.348 | ** | 0.215 | | 0.009 | | 57.356% |
| | (0.192) | | (0.034) | | (0.160) | | (0.166) | | (0.008) | | |
| HAR-vix | 0.481 | ** | 0.117 | *** | 0.346 | * | 0.228 | | 0.034 | | 57.006% |
| | (0.236) | | (0.042) | | (0.179) | | (0.195) | | (0.142) | | |

*** 1% significance, ** 5% siginificance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.8: c) In-sample estimation of HARX (eq. (2.4)) using SPY data for $y = RV^{(m)}$, Post-Crisis Period

| | $\hat{\beta}_0^{(m)}$ | | $\hat{\beta}_d^{(m)}$ | | $\hat{\beta}_w^{(m)}$ | | $\hat{\beta}_m^{(m)}$ | | $\hat{\beta}_x^{(m)}$ | | **Adj. $R^2$** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| HAR | 0.235 | *** | 0.100 | * | 0.068 | ** | 0.360 | *** | | | 29.782% |
| | (0.058) | | (0.052) | | (0.031) | | (0.073) | | | | |
| HAR-ue | 0.295 | *** | 0.100 | * | 0.067 | ** | 0.355 | *** | 0.001 | | 30.433% |
| | (0.079) | | (0.051) | | (0.029) | | (0.065) | | (0.001) | | |
| HAR-ind | 0.213 | *** | 0.100 | * | 0.068 | ** | 0.357 | *** | 0.024 | | 29.872% |
| | (0.064) | | (0.052) | | (0.031) | | (0.073) | | (0.053) | | |
| HAR-hst | 0.269 | *** | 0.100 | ** | 0.066 | ** | 0.354 | *** | -0.001 | | 30.579% |
| | (0.071) | | (0.051) | | (0.030) | | (0.067) | | (0.001) | | |
| HAR-ts | 0.052 | | 0.101 | * | 0.067 | ** | 0.367 | *** | 0.123 | | 31.014% |
| | (0.115) | | (0.053) | | (0.030) | | (0.067) | | (0.087) | | |
| HAR-cci | 10.471 | | 0.096 | ** | 0.067 | ** | 0.302 | *** | -0.103 | | 32.756% |
| | (8.794) | | (0.046) | | (0.031) | | (0.046) | | (0.088) | | |
| HAR-cli | 16.468 | | 0.097 | * | 0.071 | ** | 0.305 | *** | -0.162 | | 31.808% |
| | (12.267) | | (0.050) | | (0.029) | | (0.070) | | (0.122) | | |
| HAR-bci | 9.341 | * | 0.099 | * | 0.070 | ** | 0.342 | *** | -0.091 | | 30.320% |
| | (5.599) | | (0.053) | | (0.029) | | (0.077) | | (0.056) | | |
| HAR-ppi | -0.197 | | 0.101 | * | 0.067 | ** | 0.363 | *** | 0.002 | | 29.800% |
| | (1.294) | | (0.053) | | (0.031) | | (0.071) | | (0.007) | | |
| HAR-ads | 0.241 | *** | 0.100 | * | 0.068 | ** | 0.361 | *** | 0.035 | | 29.771% |
| | (0.064) | | (0.052) | | (0.031) | | (0.072) | | (0.102) | | |
| HAR-fd | 0.080 | | 0.100 | * | 0.069 | ** | 0.350 | *** | -0.040 | | 30.296% |
| | (0.166) | *** | (0.052) | *** | (0.028) | * | (0.058) | | (0.046) | | |
| HAR-vol | -0.061 | | -0.009 | | 0.016 | | 0.280 | *** | 0.376 | ** | 36.154% |
| | (0.120) | *** | (0.034) | *** | (0.037) | *** | (0.065) | | (0.173) | | |
| HAR-trd | -0.029 | | -0.013 | | 0.012 | | 0.310 | *** | 0.010 | ** | 33.297% |
| | (0.092) | | (0.040) | | (0.044) | * | (0.073) | | (0.004) | | |
| HAR-vix | 0.118 | ** | 0.034 | | 0.026 | | 0.102 | | 0.332 | *** | 34.307% |
| | (0.052) | | (0.040) | | (0.028) | | (0.084) | | (0.101) | | |

*** 1% significance, ** 5% siginificance, and * 10% significance

This table reports the estimated coefficients with its (standard deviation) and level of significance and the adjusted $R^2$ for each equation.

Table 2.9: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(d)}$ of ten individual stocks in each sector over the full-sample period

| | **ADS** | **BCI** | **CCI** | **CLI** | **EPU** | **FEDFD** | **HOUST** | **INDPRO** | **PPI** | **TRD** | **TS** | **UE** | **VIX** | **VOL** | **SPLS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.001 | 1.005 | **0.997** | **0.937** | **0.992** | 1.005 | 1.001 | 1.003 | **0.998** | 1.003 | 1.006 | 1.003 | 1.004 | **0.994** | **0.950** |
| **CS** | 1.002 | **0.999** | **0.999** | **0.899** | **0.981** | 1.000 | **0.991** | **0.997** | **0.995** | **0.989** | **0.991** | **0.985** | **0.991** | **0.965** | **0.921** |
| **Energy** | 1.002 | 1.002 | 1.002 | **0.951** | **0.993** | 1.000 | **0.999** | 1.002 | **0.995** | 1.003 | 1.003 | 1.003 | 1.017 | **0.993** | **0.911** |
| **FIN** | 1.004 | 1.001 | **0.998** | **0.974** | **0.979** | 1.004 | **0.994** | **0.996** | **0.992** | **0.987** | **0.989** | **0.985** | **0.964** | **0.948** | **0.958** |
| **HC** | 1.007 | 1.002 | **0.999** | **0.912** | **0.992** | 1.002 | **0.996** | 1.002 | **0.998** | 1.001 | 1.002 | **0.998** | **0.998** | **0.895** | **0.917** |
| **IND** | 1.003 | 1.007 | 1.012 | **0.883** | 1.006 | 1.003 | 1.005 | 1.013 | 1.004 | 1.018 | 1.014 | 1.019 | 1.044 | 1.030 | **0.886** |
| **IT** | 1.007 | 1.008 | 1.001 | **0.965** | **0.992** | 1.006 | **0.999** | 1.002 | **0.994** | **0.998** | **0.999** | **0.997** | 1.008 | 1.003 | **0.974** |
| **MAT** | 1.003 | 1.006 | 1.003 | **0.960** | **0.996** | 1.003 | 1.000 | 1.007 | **0.998** | 1.007 | 1.007 | 1.006 | 1.014 | 1.015 | **0.953** |
| **TELE** | 1.008 | 1.006 | 1.001 | **0.927** | **0.997** | 1.003 | 1.003 | 1.007 | **0.997** | 1.008 | 1.008 | 1.006 | 1.014 | 1.009 | **0.919** |
| **UTI** | 1.010 | 1.017 | **0.996** | **0.943** | **0.982** | 1.005 | **0.996** | 1.008 | 1.002 | 1.004 | 1.005 | 1.000 | 1.008 | 1.022 | **0.943** |

Table 2.10: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(w)}$ of ten individual stocks in each sector over the full-sample period

| | **ADS** | **BCI** | **CCI** | **CLI** | **EPU** | **FEDFD** | **HOUST** | **INDPRO** | **PPI** | **TRD** | **TS** | **UE** | **VIX** | **VOL** | **SPLS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.004 | 1.011 | **0.978** | **0.888** | **0.940** | 1.004 | **0.981** | **0.990** | **0.966** | **0.970** | **0.989** | **0.968** | 1.009 | **0.976** | **0.889** |
| **CS** | 1.004 | 1.001 | **0.976** | **0.819** | **0.912** | **0.995** | **0.953** | **0.973** | **0.959** | **0.935** | **0.949** | **0.926** | **0.994** | **0.934** | **0.864** |
| **Energy** | 1.002 | 1.001 | **0.985** | **0.943** | **0.932** | **0.997** | **0.964** | **0.973** | **0.945** | **0.949** | **0.964** | **0.949** | 1.046 | **0.994** | **0.962** |
| **FIN** | 1.004 | 1.003 | **0.992** | **0.933** | **0.879** | **0.995** | **0.946** | **0.950** | **0.936** | **0.901** | **0.925** | **0.895** | **0.938** | **0.888** | **0.910** |
| **HC** | 1.011 | 1.003 | **0.991** | **0.880** | **0.947** | **1.000** | **0.973** | **0.993** | **0.970** | **0.976** | **0.985** | **0.970** | 1.045 | **0.929** | **0.893** |
| **IND** | 1.005 | 1.010 | **0.994** | **0.842** | **0.937** | 1.000 | **0.974** | **0.997** | **0.955** | **0.979** | **0.989** | **0.978** | 1.038 | 1.001 | **0.838** |
| **IT** | 1.029 | 1.026 | **0.991** | **0.953** | **0.941** | 1.009 | **0.977** | **0.981** | **0.955** | **0.950** | **0.967** | **0.949** | 1.014 | 1.002 | **0.956** |
| **MAT** | 1.003 | 1.010 | **0.992** | **0.957** | **0.940** | **1.000** | **0.973** | **0.987** | **0.952** | **0.967** | **0.981** | **0.964** | 1.011 | 1.006 | **0.950** |
| **TELE** | 1.029 | 1.015 | **0.985** | **0.845** | **0.950** | **1.000** | **0.985** | **0.994** | **0.962** | **0.971** | **0.990** | **0.970** | 1.001 | **0.963** | **0.859** |
| **UTI** | 1.017 | 1.033 | **0.975** | **0.863** | **0.918** | 1.012 | **0.963** | **0.995** | **0.962** | **0.972** | **0.981** | **0.962** | **0.997** | **0.992** | **0.855** |

Table 2.11: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(m)}$ of ten individual stocks in each sector over the full-sample period

| | ADS | BCI | CCI | CLI | EPU | FEDFD | HOUST | INDPRO | PPI | TRD | TS | UE | VIX | VOL | SPLS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.022 | 1.050 | **0.954** | **0.903** | **0.809** | **0.985** | **0.906** | **0.875** | **0.835** | **0.781** | **0.824** | **0.767** | **1.000** | **0.931** | **0.660** |
| **CS** | 1.013 | 1.022 | **0.943** | **0.855** | **0.731** | **0.975** | **0.831** | **0.820** | **0.791** | **0.689** | **0.742** | **0.675** | **0.989** | **0.881** | **0.638** |
| **Energy** | 1.006 | 1.008 | **0.947** | **0.943** | **0.761** | **0.982** | **0.827** | **0.782** | **0.744** | **0.656** | **0.704** | **0.667** | 1.009 | **0.920** | **0.656** |
| **FIN** | 1.008 | 1.027 | **0.981** | **0.936** | **0.723** | **0.973** | **0.851** | **0.805** | **0.771** | **0.669** | **0.715** | **0.653** | **0.910** | **0.875** | **0.656** |
| **HC** | 1.021 | 1.019 | **0.957** | **0.923** | **0.766** | **0.973** | **0.873** | **0.870** | **0.811** | **0.770** | **0.805** | **0.758** | 1.087 | **0.959** | **0.698** |
| **IND** | 1.014 | 1.018 | **0.943** | **0.908** | **0.730** | **0.980** | **0.852** | **0.827** | **0.764** | **0.721** | **0.768** | **0.710** | 1.021 | **0.935** | **0.669** |
| **IT** | 1.092 | 1.079 | **0.972** | **0.983** | **0.794** | 1.016 | **0.901** | **0.838** | **0.800** | **0.713** | **0.768** | **0.709** | 1.016 | **0.981** | **0.757** |
| **MAT** | 1.009 | 1.032 | **0.961** | **0.981** | **0.756** | **0.981** | **0.857** | **0.817** | **0.769** | **0.701** | **0.752** | **0.690** | **0.998** | **0.968** | **0.700** |
| **TELE** | 1.064 | 1.021 | **0.948** | **0.839** | **0.813** | **0.984** | **0.908** | **0.876** | **0.826** | **0.765** | **0.820** | **0.769** | **0.992** | **0.906** | **0.672** |
| **UTI** | 1.023 | 1.049 | **0.921** | **0.858** | **0.734** | **0.999** | **0.831** | **0.830** | **0.775** | **0.720** | **0.751** | **0.709** | **0.998** | **0.971** | **0.633** |



Table 2.12: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(d)}$ of ten individual stocks in each sector during the Pre-Crisis period

| *Pre-Crisis* | ADS | BCI | CCI | CLI | EPU | FEDFD | HOUST | INDPRO | PPI | TRD | TS | UE | VIX | VOL | SPLS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.020 | 1.039 | **0.985** | **0.893** | **0.989** | 1.060 | 1.010 | 1.008 | **0.998** | 1.015 | 1.024 | 1.009 | 1.033 | **0.998** | 1.018 |
| **CS** | 1.011 | 1.028 | **0.987** | **0.927** | **0.999** | 1.028 | 1.006 | 1.005 | 1.000 | **0.995** | 1.009 | 1.001 | 1.017 | **1.000** | **0.972** |
| **Energy** | 1.022 | 1.002 | **0.977** | 1.008 | 1.007 | 1.010 | 1.021 | 1.001 | 1.002 | **0.994** | 1.004 | 1.005 | 1.401 | 1.225 | 1.151 |
| **FIN** | 1.046 | 1.132 | **0.951** | **0.881** | **0.980** | 1.089 | 1.051 | 1.015 | **0.995** | 1.001 | 1.025 | 1.005 | 1.998 | 1.557 | 1.210 |
| **HC** | 1.005 | 1.007 | **0.996** | **0.970** | 1.005 | 1.019 | 1.006 | 1.003 | 1.001 | **0.999** | 1.003 | 1.002 | 1.019 | 1.013 | 1.000 |
| **IND** | 1.006 | 1.034 | **0.987** | **0.917** | **0.991** | 1.028 | 1.007 | 1.006 | **0.994** | 1.003 | 1.010 | 1.003 | 1.030 | 1.035 | **0.969** |
| **IT** | 1.084 | 1.095 | **0.974** | **0.908** | 1.002 | 1.094 | 1.022 | 1.018 | 1.003 | **0.996** | 1.016 | **0.999** | 1.033 | 1.049 | **0.991** |
| **MAT** | 1.009 | 1.023 | **0.995** | 1.032 | 1.001 | 1.026 | 1.002 | 1.003 | **1.000** | 1.003 | 1.008 | 1.002 | 1.012 | 1.014 | 1.028 |
| **TELE** | 1.064 | 1.041 | **0.968** | **0.916** | 1.029 | 1.043 | 1.032 | 1.023 | 1.002 | **0.999** | 1.036 | 1.015 | 1.039 | 1.146 | **0.989** |
| **UTI** | 1.111 | 1.215 | **0.967** | **0.947** | **0.938** | 1.088 | **0.995** | 1.005 | 1.107 | 1.029 | 1.017 | 1.009 | 1.161 | 1.711 | 1.289 |

Table 2.13: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(w)}$ of ten individual stocks in each sector during the Pre-Crisis period

| | **ADS** | **BCI** | **CCI** | **CLI** | **EPU** | **FEDFD** | **HOUST** | **INDPRO** | **PPI** | **TRD** | **TS** | **UE** | **VIX** | **VOL** | **SPLS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.128 | 1.161 | **0.946** | **0.707** | **0.959** | 1.099 | 1.033 | 1.028 | **0.995** | 1.020 | 1.078 | 1.025 | 1.045 | **0.922** | **0.795** |
| **CS** | 1.066 | 1.108 | **0.977** | **0.780** | **0.994** | 1.051 | 1.028 | 1.018 | **0.999** | **0.979** | 1.026 | 1.004 | 1.023 | **0.901** | **0.811** |
| **Energy** | 1.050 | **0.998** | **0.963** | **0.962** | **0.995** | 1.017 | 1.045 | **0.993** | 1.007 | **0.995** | 1.001 | 1.010 | 1.186 | 1.153 | 1.016 |
| **FIN** | 1.192 | 1.350 | **0.937** | **0.655** | **0.944** | 1.120 | 1.033 | 1.032 | **0.992** | **0.997** | 1.054 | 1.015 | 1.325 | 1.010 | **0.690** |
| **HC** | 1.026 | 1.021 | **0.996** | **0.890** | 1.017 | 1.040 | 1.028 | 1.011 | **0.993** | **0.994** | 1.014 | 1.012 | 1.004 | **0.977** | **0.904** |
| **IND** | 1.061 | 1.143 | **0.973** | **0.732** | **0.958** | 1.030 | 1.019 | 1.026 | **0.983** | 1.011 | 1.033 | 1.011 | 1.009 | **0.919** | **0.770** |
| **IT** | 1.429 | 1.347 | **0.986** | **0.780** | **0.999** | 1.186 | 1.123 | 1.069 | 1.014 | **0.980** | 1.062 | **0.999** | 1.053 | **0.979** | **0.787** |
| **MAT** | 1.035 | 1.090 | **0.987** | 1.006 | **0.996** | 1.051 | 1.012 | 1.012 | **0.993** | 1.010 | 1.026 | 1.005 | 1.012 | **0.994** | **0.962** |
| **TELE** | 1.145 | 1.052 | **0.971** | **0.655** | 1.116 | 1.052 | 1.125 | 1.073 | 1.020 | **0.991** | 1.089 | 1.047 | 1.004 | 1.073 | **0.787** |
| **UTI** | 1.145 | 1.284 | **0.907** | **0.858** | **0.916** | 1.156 | 1.010 | 1.016 | 1.089 | 1.014 | 1.035 | 1.016 | 1.067 | 1.263 | **0.782** |

Table 2.14: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(m)}$ of ten individual stocks in each sector during the Pre-Crisis period

| | **ADS** | **BCI** | **CCI** | **CLI** | **EPU** | **FEDFD** | **HOUST** | **INDPRO** | **PPI** | **TRD** | **TS** | **UE** | **VIX** | **VOL** | **SPLS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.435 | 1.323 | **0.970** | **0.632** | **0.916** | 1.107 | 1.111 | 1.072 | **0.986** | **0.968** | 1.102 | 1.024 | 1.027 | **0.785** | **0.805** |
| **CS** | 1.277 | 1.242 | 1.020 | **0.683** | **0.996** | 1.098 | 1.109 | 1.058 | **0.976** | **0.921** | 1.070 | 1.013 | **0.993** | **0.750** | **0.807** |
| **Energy** | 1.096 | **0.976** | **0.984** | **0.933** | **0.989** | 1.023 | 1.064 | 1.015 | 1.050 | 1.012 | **0.969** | 1.038 | **1.000** | 1.013 | **0.813** |
| **FIN** | 1.476 | 1.519 | **0.980** | **0.586** | **0.920** | 1.141 | 1.096 | 1.061 | **0.994** | **0.956** | 1.095 | 1.019 | 1.018 | **0.858** | **0.629** |
| **HC** | 1.172 | 1.060 | 1.027 | **0.849** | 1.051 | 1.071 | 1.139 | 1.047 | **0.999** | **0.975** | 1.064 | 1.038 | **0.999** | **0.882** | **0.863** |
| **IND** | 1.274 | 1.295 | **1.000** | **0.691** | **0.892** | 1.044 | 1.062 | 1.065 | **0.975** | **0.984** | 1.065 | 1.022 | **0.951** | **0.776** | **0.749** |
| **IT** | 1.920 | 1.495 | 1.039 | **0.965** | **0.953** | 1.306 | 1.318 | 1.108 | 1.053 | **0.937** | 1.122 | **0.997** | 1.090 | **0.864** | 1.050 |
| **MAT** | 1.175 | 1.242 | 1.005 | **0.924** | **0.995** | 1.118 | 1.054 | 1.070 | **0.961** | 1.006 | 1.075 | 1.019 | 1.026 | **0.973** | 1.052 |
| **TELE** | 1.214 | **0.955** | 1.014 | **0.581** | 1.243 | 1.066 | 1.276 | 1.126 | 1.057 | **0.982** | 1.134 | 1.102 | **0.926** | **0.887** | **0.825** |
| **UTI** | 1.199 | 1.319 | **0.911** | **0.632** | **0.922** | 1.185 | 1.093 | 1.062 | 1.044 | **0.989** | 1.046 | 1.033 | 1.046 | 1.084 | **0.647** |

Table 2.15: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(d)}$ of ten individual stocks in each sector during the Crisis period

| *Crisis* | **ADS** | **BCI** | **CCI** | **CLI** | **EPU** | **FEDFD** | **HOUST** | **INDPRO** | **PPI** | **TRD** | **TS** | **UE** | **VIX** | **VOL** | **SPLS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.001 | 1.005 | **0.997** | **0.940** | **0.992** | **1.000** | **0.998** | 1.003 | **0.995** | 1.002 | 1.005 | 1.002 | 1.005 | 1.003 | **0.943** |
| **CS** | 1.002 | **0.999** | 1.001 | **0.881** | **0.980** | **0.999** | **0.989** | **0.999** | **0.991** | **0.990** | **0.991** | **0.986** | **0.997** | **0.973** | **0.905** |
| **Energy** | 1.002 | 1.003 | 1.002 | **0.946** | **0.992** | **1.000** | **0.997** | 1.004 | **0.989** | 1.004 | 1.004 | 1.004 | 1.013 | **0.989** | **0.897** |
| **FIN** | 1.004 | 1.001 | **0.997** | **0.971** | **0.978** | 1.004 | **0.992** | **0.994** | **0.990** | **0.985** | **0.988** | **0.984** | **0.963** | **0.947** | **0.955** |
| **HC** | 1.004 | 1.004 | 1.000 | **0.920** | **0.996** | **1.000** | **0.998** | 1.009 | **0.996** | 1.006 | 1.006 | 1.003 | 1.023 | **0.997** | **0.926** |
| **IND** | 1.003 | 1.007 | 1.013 | **0.876** | 1.005 | 1.003 | 1.003 | 1.013 | 1.001 | 1.018 | 1.014 | 1.019 | 1.044 | 1.030 | **0.879** |
| **IT** | 1.004 | 1.005 | 1.001 | **0.961** | **0.990** | 1.002 | **0.995** | 1.001 | **0.989** | **0.997** | **0.998** | **0.996** | 1.009 | 1.003 | **0.969** |
| **MAT** | 1.003 | 1.006 | 1.002 | **0.948** | **0.995** | 1.002 | **0.998** | 1.007 | **0.992** | 1.007 | 1.007 | 1.006 | 1.015 | 1.017 | **0.937** |
| **TELE** | 1.002 | 1.003 | 1.002 | **0.911** | **0.996** | 1.001 | 1.000 | 1.006 | **0.994** | 1.008 | 1.006 | 1.006 | 1.013 | 1.007 | **0.904** |
| **UTI** | 1.001 | **0.999** | **0.996** | **0.934** | **0.986** | **0.998** | **0.995** | 1.009 | **0.988** | 1.000 | 1.005 | **1.000** | **0.999** | **0.990** | **0.918** |

Table 2.16: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(w)}$ of ten individual stocks in each sector during the Crisis period

| | **ADS** | **BCI** | **CCI** | **CLI** | **EPU** | **FEDFD** | **HOUST** | **INDPRO** | **PPI** | **TRD** | **TS** | **UE** | **VIX** | **VOL** | **SPLS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.001 | 1.007 | **0.982** | **0.895** | **0.938** | **0.990** | **0.969** | **0.982** | **0.948** | **0.960** | **0.977** | **0.957** | 1.007 | **0.986** | **0.893** |
| **CS** | 1.003 | **0.998** | **0.975** | **0.801** | **0.899** | **0.992** | **0.940** | **0.966** | **0.937** | **0.925** | **0.943** | **0.915** | **0.991** | **0.931** | **0.844** |
| **Energy** | 1.003 | 1.001 | **0.984** | **0.948** | **0.916** | **0.996** | **0.949** | **0.963** | **0.911** | **0.937** | **0.955** | **0.937** | 1.031 | **0.974** | **0.962** |
| **FIN** | 1.005 | **0.999** | **0.989** | **0.929** | **0.876** | **0.996** | **0.938** | **0.940** | **0.927** | **0.894** | **0.921** | **0.890** | **0.938** | **0.888** | **0.906** |
| **HC** | 1.005 | 1.005 | **0.984** | **0.855** | **0.927** | **0.992** | **0.960** | **0.992** | **0.942** | **0.962** | **0.976** | **0.956** | 1.005 | **0.968** | **0.858** |
| **IND** | 1.005 | 1.008 | **0.993** | **0.834** | **0.930** | **1.000** | **0.965** | **0.990** | **0.939** | **0.972** | **0.986** | **0.971** | 1.037 | **0.998** | **0.827** |
| **IT** | 1.007 | 1.008 | **0.990** | **0.961** | **0.926** | **0.997** | **0.956** | **0.964** | **0.927** | **0.935** | **0.954** | **0.935** | 1.013 | 1.001 | **0.961** |
| **MAT** | 1.005 | 1.010 | **0.990** | **0.950** | **0.928** | **0.998** | **0.958** | **0.976** | **0.925** | **0.955** | **0.974** | **0.951** | 1.014 | 1.008 | **0.942** |
| **TELE** | 1.003 | 1.004 | **0.984** | **0.835** | **0.938** | **0.995** | **0.968** | **0.985** | **0.944** | **0.963** | **0.978** | **0.959** | 1.001 | **0.961** | **0.829** |
| **UTI** | 1.001 | **0.994** | **0.978** | **0.861** | **0.923** | **0.994** | **0.956** | **0.992** | **0.926** | **0.959** | **0.977** | **0.955** | **0.985** | **0.953** | **0.852** |

Table 2.17: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(m)}$ of ten individual stocks in each sector during the Crisis period

| | ADS | BCI | CCI | CLI | EPU | FEDFD | HOUST | INDPRO | PPI | TRD | TS | UE | VIX | VOL | SPLS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | 1.011 | 1.045 | **0.951** | **0.935** | **0.799** | **0.968** | **0.860** | **0.819** | **0.783** | **0.727** | **0.776** | **0.715** | **0.995** | **0.945** | **0.620** |
| **CS** | 1.010 | 1.010 | **0.930** | **0.864** | **0.706** | **0.970** | **0.783** | **0.763** | **0.735** | **0.636** | **0.697** | **0.626** | **0.986** | **0.882** | **0.592** |
| **Energy** | 1.009 | 1.007 | **0.938** | **0.957** | **0.703** | **0.981** | **0.768** | **0.702** | **0.654** | **0.574** | **0.628** | **0.591** | 1.010 | **0.909** | **0.614** |
| **FIN** | 1.012 | 1.007 | **0.974** | **0.935** | **0.724** | **0.980** | **0.827** | **0.766** | **0.749** | **0.640** | **0.694** | **0.635** | **0.910** | **0.874** | **0.644** |
| **HC** | 1.014 | 1.015 | **0.922** | **0.901** | **0.702** | **0.971** | **0.804** | **0.790** | **0.724** | **0.677** | **0.727** | **0.668** | **0.994** | **0.943** | **0.610** |
| **IND** | 1.012 | 1.008 | **0.930** | **0.917** | **0.715** | **0.982** | **0.818** | **0.778** | **0.720** | **0.678** | **0.735** | **0.674** | 1.022 | **0.933** | **0.640** |
| **IT** | 1.022 | 1.033 | **0.958** | **0.989** | **0.763** | **0.981** | **0.821** | **0.755** | **0.720** | **0.641** | **0.698** | **0.642** | 1.009 | **0.988** | **0.683** |
| **MAT** | 1.015 | 1.027 | **0.947** | **0.980** | **0.730** | **0.980** | **0.815** | **0.750** | **0.706** | **0.636** | **0.700** | **0.637** | **0.999** | **0.965** | **0.652** |
| **TELE** | 1.013 | 1.016 | **0.937** | **0.885** | **0.771** | **0.974** | **0.841** | **0.803** | **0.759** | **0.703** | **0.764** | **0.703** | **0.997** | **0.926** | **0.609** |
| **UTI** | 1.009 | **0.992** | **0.910** | **0.898** | **0.723** | **0.976** | **0.788** | **0.774** | **0.700** | **0.661** | **0.707** | **0.662** | **0.985** | **0.913** | **0.627** |

Table 2.18: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(d)}$ of ten individual stocks in each sector during the Post-Crisis period

| | ADS | BCI | CCI | CLI | EPU | FEDFD | HOUST | INDPRO | PPI | TRD | TS | UE | VIX | VOL | SPLS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | **0.992** | **0.995** | 1.032 | 1.200 | **0.980** | **0.990** | 1.026 | 1.016 | 1.040 | 1.018 | 1.008 | 1.009 | **0.994** | **0.989** | 1.227 |
| **CS** | **0.981** | **0.996** | 1.017 | **0.985** | **0.975** | **0.982** | 1.044 | 1.104 | 1.051 | 1.069 | 1.048 | 1.018 | **0.966** | **0.982** | **0.974** |
| **Energy** | 1.001 | 1.002 | 1.007 | 1.004 | **0.999** | 1.003 | 1.011 | 1.005 | 1.038 | 1.004 | **1.000** | 1.002 | **0.996** | **0.996** | 1.009 |
| **FIN** | **0.989** | 1.056 | 1.279 | 1.249 | 1.073 | **0.980** | 1.240 | 1.370 | 1.308 | 1.286 | 1.219 | 1.123 | 1.331 | 1.213 | 1.379 |
| **HC** | 1.002 | **1.000** | 1.008 | **0.972** | **0.991** | 1.003 | 1.005 | 1.003 | 1.011 | 1.005 | 1.003 | 1.001 | **0.979** | **0.892** | **0.973** |
| **IND** | 1.000 | 1.001 | 1.020 | 1.026 | **0.999** | 1.006 | 1.027 | 1.009 | 1.068 | 1.013 | 1.004 | 1.008 | **0.989** | 1.015 | **0.988** |
| **IT** | 1.001 | 1.003 | 1.008 | **0.988** | **0.997** | 1.001 | 1.012 | 1.008 | 1.021 | 1.004 | **1.000** | 1.002 | **0.993** | **0.986** | **0.990** |
| **MAT** | **0.999** | 1.003 | 1.017 | 1.049 | **0.998** | 1.006 | 1.020 | 1.008 | 1.051 | 1.011 | 1.004 | 1.008 | **0.986** | **0.991** | 1.036 |
| **TELE** | 1.000 | 1.001 | 1.011 | 1.114 | **0.995** | 1.005 | 1.012 | 1.006 | 1.031 | 1.008 | 1.002 | 1.004 | 1.019 | 1.039 | 1.063 |
| **UTI** | **0.994** | **0.994** | 1.046 | 1.171 | **0.973** | **0.989** | 1.027 | 1.012 | 1.080 | 1.016 | 1.001 | 1.008 | 1.101 | 1.198 | 1.134 |

Table 2.19: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(w)}$ of ten individual stocks in each sector during the POst-Crisis period

| | **ADS** | **BCI** | **CCI** | **CLI** | **EPU** | **FEDFD** | **HOUST** | **INDPRO** | **PPI** | **TRD** | **TS** | **UE** | **VIX** | **VOL** | **SPLS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | **0.971** | **0.991** | 1.071 | 1.303 | **0.940** | **0.966** | 1.071 | 1.067 | 1.125 | 1.048 | 1.012 | 1.022 | 1.039 | 1.044 | 1.292 |
| **CS** | **0.960** | 1.018 | 1.022 | **0.981** | **0.950** | **0.967** | 1.076 | 1.169 | 1.108 | 1.100 | 1.062 | 1.027 | **0.969** | **0.998** | **0.969** |
| **Energy** | **0.990** | 1.013 | 1.008 | **0.985** | **0.997** | **0.996** | 1.031 | 1.031 | 1.132 | 1.011 | **1.000** | 1.000 | 1.001 | 1.013 | **0.985** |
| **FIN** | **0.906** | 1.360 | 1.246 | 1.256 | 1.065 | **0.864** | 1.609 | 1.978 | 1.795 | 1.600 | 1.391 | 1.252 | **1.000** | 1.002 | 1.259 |
| **HC** | **0.995** | **1.000** | 1.013 | 1.029 | **0.973** | 1.001 | 1.020 | 1.017 | 1.063 | 1.021 | 1.007 | 1.008 | 1.074 | **0.955** | **0.980** |
| **IND** | **0.984** | 1.008 | 1.025 | 1.039 | **0.999** | **0.994** | 1.088 | 1.056 | 1.204 | 1.047 | 1.006 | 1.027 | **0.999** | 1.074 | 1.062 |
| **IT** | **0.987** | 1.013 | 1.002 | **0.985** | **0.979** | **0.993** | 1.040 | 1.044 | 1.083 | 1.012 | **0.993** | 1.001 | **0.998** | 1.006 | **0.998** |
| **MAT** | **0.976** | 1.016 | 1.028 | 1.053 | **0.998** | 1.004 | 1.079 | 1.071 | 1.203 | 1.050 | 1.020 | 1.031 | **0.986** | **0.990** | 1.050 |
| **TELE** | **0.983** | 1.004 | 1.012 | 1.160 | **0.977** | 1.009 | 1.055 | 1.044 | 1.153 | 1.040 | 1.008 | 1.021 | 1.033 | 1.097 | 1.124 |
| **UTI** | **0.959** | **0.993** | 1.064 | 1.160 | **0.926** | **0.963** | 1.093 | 1.056 | 1.305 | 1.056 | 1.003 | 1.023 | 1.127 | 1.319 | 1.150 |

Table 2.20: OOS median MSE % gains of HARX relative to the HAR model in forecasting $RV^{(m)}$ of ten individual stocks in each sector during the POst-Crisis period

| | **ADS** | **BCI** | **CCI** | **CLI** | **EPU** | **FEDFD** | **HOUST** | **INDPRO** | **PPI** | **TRD** | **TS** | **UE** | **VIX** | **VOL** | **SPLS** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **CD** | **0.886** | 1.035 | 1.159 | 1.250 | **0.795** | **0.929** | 1.216 | 1.388 | 1.255 | 1.211 | 1.099 | 1.048 | 1.062 | 1.101 | 1.014 |
| **CS** | **0.889** | 1.118 | 1.060 | **0.973** | **0.819** | **0.912** | 1.199 | 1.369 | 1.303 | 1.196 | 1.086 | 1.026 | **0.979** | 1.027 | **0.969** |
| **Energy** | **0.943** | 1.077 | 1.068 | 1.015 | **0.952** | **0.952** | 1.134 | 1.311 | 1.289 | 1.138 | 1.076 | 1.011 | **0.977** | 1.008 | **0.954** |
| **FIN** | **0.743** | 1.959 | 1.269 | 1.136 | **0.662** | **0.629** | 1.866 | 2.700 | 2.001 | 2.023 | 1.699 | 1.381 | **0.973** | 1.001 | 1.291 |
| **HC** | **0.944** | 1.029 | 1.085 | 1.040 | **0.888** | **0.937** | 1.131 | 1.217 | 1.155 | 1.127 | 1.051 | 1.040 | 1.219 | 1.031 | **0.987** |
| **IND** | **0.909** | 1.080 | 1.138 | **0.983** | **0.890** | **0.904** | 1.272 | 1.439 | 1.342 | 1.236 | 1.082 | 1.069 | **0.992** | 1.087 | 1.051 |
| **IT** | **0.919** | 1.093 | 1.051 | **0.989** | **0.853** | **0.944** | 1.188 | 1.340 | 1.273 | 1.148 | 1.051 | 1.005 | **0.998** | 1.015 | 1.084 |
| **MAT** | **0.887** | 1.094 | 1.150 | 1.052 | **0.898** | **0.942** | 1.258 | 1.504 | 1.378 | 1.265 | 1.161 | 1.079 | **0.964** | **0.973** | 1.086 |
| **TELE** | **0.870** | 1.057 | 1.108 | 1.051 | **0.807** | **0.939** | 1.216 | 1.398 | 1.346 | 1.237 | 1.101 | 1.056 | 1.011 | 1.104 | 1.088 |
| **UTI** | **0.747** | 1.074 | 1.296 | 1.020 | **0.665** | **0.781** | 1.299 | 1.690 | 1.613 | 1.334 | 1.162 | 1.032 | 1.058 | 1.244 | 1.181 |

# Chapter 3

# SHARP: A State-Space HAR model using Particle Gibbs Sampling


## Abstract

We propose a general state-space autoregressive (AR) model with time-varying coefficients that follow an AR process with stochastic volatility. We implement these new specifications in the HAR framework to capture the time-varying salient feature of volatility using a two-state representation via a) allowing the time-varying coefficients to follow an AR(1) specification. b) introducing stochastic volatility for the innovations of the coefficients. Using high-frequency data of the SPY-ETF and representative NYSE stocks from 2000 to 2016, we show that the proposed model estimated using particle Gibbs sampling consistently outperforms different HAR

model specifications in forecasting financial volatility.

**Key words:** Forecasting; Heterogeneous Autoregressive Realised Volatility model; Particle GIBBS sampling; Sequential Monte Carlo; State Space models.

## 3.1 Introduction

There is considerable evidence of time-varying parameters and volatility in time series macroeconomic and financial data. For out-of-sample analyses, models that fail to anticipate changes in the data generating process produce inaccurate predictions. Hence, analysing and accounting for parameters' behaviour is essential in the statistical modelling of such data. Much attention has been given to forecast evaluation (Giacomini and Rossi, 2009), estimation approaches (Inoue, Jin, and Rossi, 2017; Pesaran and Timmermann, 2005), and dynamic models (Bekierman and Manner (2018), Buccheri and Corsi (2021), Chen et al. (2018), Clark and Ravazzolo (2015), and D'Agostino, Gambetti, and Giannone (2013), among others).

This work presents a novel state-space model where the coefficients are assumed to be latent factors that follow an autoregressive (AR) process with Stochastic Volatility (SV). Such features can be utilised for any model specification, including exogenous regressors, trends, seasonal dummies, or autoregressive structures. We use Bayesian inference to estimate the model with Particle Gibbs sampling (Andrieu, Doucet, and Holenstein, 2010) following the Creal and Tsay (2015) procedure. This method permits an efficient computation of the latent variables. We implement the new specifications in modelling and forecasting the volatility of financial assets returns.

Financial volatility is a preeminent topic in theoretical and empirical finance due to its significance in risk management, portfolio selection, and asset pricing. In a seminal work, to capture the time-varying conditional volatility of financial returns, Engle (1982) introduced the ARCH model. Extending that work, Bollerslev (1986) proposed the GARCH model, which became widely accepted. In further development, the non-parametric realised volatility (RV) - defined as the sum of intraday returns - is preferred to parametric volatility measures generated by GARCH and SV models. The improvement derives from the delivery of information over much smaller intervals of time (Andersen and Bollerslev, 1998; Barndorff-Nielsen and Shephard, 2002; Liu, Patton, and Sheppard, 2015).

Since then, significant work has been done on linear reduced-form modelling of the $RV$ series, where the heterogeneous autoregressive (HAR) model by Corsi (2009) has become the preferred specification for that purpose. The model is distinctive by its heterogeneous market hypothesis, i.e., that, with diverse horizons and trading activities, market participants may react differently to identical news. It models daily volatility as a sequence of RV's daily, weekly, and monthly historical average. By its design, it parsimoniously captures RV's persistence. Further, Corsi (2009) notes that modelling the RV's log-transformation in the HAR model (henceforth, HARL) ensures the partial volatilities' positiveness, improves the dynamic specification for the RV and delivers more promising forecasts.

Despite its empirical success, the linear HARL model does not account for measurement errors and nonlinear dependencies, which may lead to biased OLS estimates, autocorrelated and highly heteroskedastic residuals, and time-varying OLS coefficients. Several proposals have emerged to address these miss-specifications in the

HARL model. Inspired by the work of Bollerslev, Patton, and Quaedvlieg (2016), but instead of explicitly using the realise quarticity ($RQ$), a proxy of the measurement error variance, Bekierman and Manner (2018) presents a state-space HARL model that specifies the first-order AR coefficient as a latent Gaussian AR(1) process. This specification also captures additional sources of temporal variations in the coefficient of interest. In parallel work, Chen et al. (2018) propose, for the HARL model, time-varying coefficients of an unknown functional form. Further, in an attempt to model both time-varying coefficients and measurement errors, Buccheri and Corsi (2021) propose score-driven coefficients and heteroskedastic innovations. Several other papers address nonlinear dependency in different approaches, such as regime-changing (McAleer and Medeiros, 2008) and dynamic model averaging (Wang et al., 2016), among others.

Our work is mostly related to that of Bekierman and Manner (2018), Buccheri and Corsi (2021), and Chen et al. (2018) in the HARL framework. We model all the coefficients (including the intercept) as an AR(1) process either with or without SV. Hence, we account for the model's nonlinearity and relax the assumption of homoskedastic errors in the coefficients' process. We also account for autocorrelation and heteroskedasticity of the residuals in the main equation. We estimate the state-space model in the HARL framework by using Particle Gibbs sampling. We refer to the proposed model i) without SV by SHARP and ii) with SV by SHARP-sv.

We compare the models' forecasting performance with a recent (parallel) work within the HARL framework. In an empirical study using daily $RV$ of SPY-ETF, as a tradable US market index, and twenty representative NYSE individual stocks over seventeen years from 2000 to 2016, both the SHARP and SHARP-sv models significantly outperform the HARL model and its existing time-varying extensions in predicting $RV$. While the SHARP-sv model offers a moderate forecasting advantage over the SHARP model, both: (i) are consistently included in the model confidence set of Hansen, Lunde, and Nason (2011); (ii) significantly outperform the HARL model in forecasting $RV$ according to the predictive ability test by Giacomini and White (2006).

Our contribution to the literature can be summarised as follows: i) we introduce a general state-space model with relaxed features (time-varying coefficients that follow an AR(1) process with SV) and suggest a feasible estimation method. ii) We utilise these new features in the HARL framework (SHARP and SHARP-sv) to improve the forecast of $RV$. iii) We compare the proposed models against recent proposals in modelling the $logRV$ series with time-varying coefficients.

The remainder of this paper is organised as follows. Firstly, in Section 3.2, we describe the realised volatility measure and provide an overview of existing dynamic and time-varying parameters extensions to the HARL model. We introduce the proposed models in Section 3.3 (refer to the Appendix for details on the estimation method). Section 3.4 analyses and evaluates the proposed extensions under the HARL framework in forecasting stock volatility using real data. Section 3.5 concludes.

## 3.2 Volatility Measure and HARL Family of Models

Consider an asset whose log-price, $\log(P_s)$, process is given by the stochastic differential equation:

$$d\log(P_s) = \mu_s + \sigma_s dW_s \tag{3.1}$$

where $\mu_s$ denotes the drift, $\sigma_s$ is the instantaneous volatility and $W_t$ a standard Brownian motion. The latent integrated variance for day $t$ is defined as:

$$IV_t = \int_{t-1}^{t} \sigma_s^2 ds. \tag{3.2}$$

$RV$ is a nonparametric ex-post estimate of the return variation, where the daily $RV$ measure is the sum of high-frequency squared returns of sub-intervals of the day. Specifically, the daily $RV$ is calculated by averaging intra-daily squared returns over a one-day horizon, $t$, using $M$ sub-intervals.

$$RV_t^{(d)} := \Sigma_{j=1}^{M} r_{j,t}^2, \tag{3.3}$$

where $r_{j,t} = log(P_{(t-1)M+j}) - log(P_{(t-1)M+(j-1)})$ is the intra-day return of the $jth$ sub-interval within the $t^{th}$ day. $P$ is the asset price at the start of the $j^{th}$ interval computed as the average of the closing and opening prices of intervals $j-1$ and $j$, respectively. Zhang, Mykland, and Aït-Sahalia (2005) provide a discussion on optimising the sampling frequency for the estimation of $RV$. However, Buccheri and Corsi (2021) show that the relative forecast performance of models with time-varying coefficients, such as the ones discussed in this paper, is independent of the sampling frequency. Therefore, for conciseness, we use sub-intervals of length 300 seconds in constructing the daily $RV$ series. The latter defines 78 intraday sub-intervals and combines balanced information from high-frequency data and microstructure effects (Andersen et al., 2001).

In the aim of forecasting the daily $RV$, Corsi (2009) introduced the Heterogeneous Autoregressive Realised Volatility (HARL) model defined as:

$$logRV_t^{(d)} = \beta_1 + \beta_2 logRV_{t-1}^{(d)} + \beta_3 logRV_{t-1}^{(w)} + \beta_4 logRV_{t-1}^{(m)} + v_t \ ; \quad v_t \sim N(0, \sigma_v^2) \quad (3.4)$$

where $logRV^{(d)}$ denotes the logarithmic transformation of daily $RV$. To account for the bias generated by the logarithmic transformation, one-step-ahead forecasts of $RV$ are estimated using the moment generating function of the normal distribution.

Respectively, $logRV^{(w)}$ and $logRV^{(m)}$ are the weekly and monthly $logRV$ realised at time $t$. These are computed over a recursive rolling window of fixed length (week or month) as follows:

$$logRV_t^{(w)} = \frac{1}{5}\Sigma_{i=1}^{5} logRV_{t-i}^{(d)} \quad ; \qquad logRV_t^{(m)} = \frac{1}{22}\Sigma_{i=1}^{22} logRV_{t-i}^{(d)}$$

Henceforth, we refer to $\beta_1$, $\beta_2$, $\beta_3$, and $\beta_4$ as the intercept, daily, weekly, and monthly coefficient, respectively.

Inspired by Barndorff-Nielsen and Shephard (2002) and Bollerslev, Patton, and Quaedvlieg (2016), Bekierman and Manner (2018) explain that the measurement error, though less severe, also exists in $logRV$ series, where:

$$logRV_t = logIV_t + \epsilon_t, \quad \epsilon_t \sim MN(0, 2\Delta \frac{IQ_t}{IV_t^2}). \tag{3.5}$$

and $V_t = \frac{\Sigma_{j=1}^{M} r_{j,t}^4}{(\Sigma_{j=1}^{M} r_{j,t}^2)^2}$ is a consistent estimate of the variance of $\epsilon_t$ (Buccheri and Corsi, 2021).

Hence, Bekierman and Manner (2018) propose the HARSL model, a state-space HARL model for the $logRV$ series that specifies the daily coefficient as a latent Gaussian AR(1) process. The model is estimated using maximum likelihood with a standard Kalman filter. Their model consistently outperforms the HAR model in forecasting the $RV$. Its empirical success lies in the specification of capturing other sources of temporal variation in addition to the variance of the measurement error. However, they note that the model's maximum likelihood estimator is inefficient. Further, allowing all the coefficients to follow an AR process is computationally challenging to perform using their employed estimation method.

In parallel work, Chen et al. (2018) suggest a HARL model with time-varying coefficients of an unknown functional form (TVCHAR). They apply a local linear smoothing method to estimate the model[1]. While their proposed model is flexible by allowing all the coefficients to be time-varying, they show that it only outperforms the HARL model for long forecasting horizons. Within the same line of research, Buccheri and Corsi (2021) propose the SHARK model that features time-varying coefficients and heteroskedastic error terms and handles measurement errors. They show that their SHARK model yields moderate improvements in the forecasts for the one day

[1] We refer the reader to their paper for more details about model estimation and to Casas and Fernandez-Casal (2019) for their R code.

ahead but is more prominent for long-term forecasting.

The above studies are primarily concerned with estimation techniques with specific underlying assumptions that better reflect the underlying process of $logRV$ series. They suggest remedies for particular miss-specifications and successfully improve long-term forecasting of $RV$. In comparison, the employed estimation method presented in the next section makes it feasible to relax some of the assumptions. Consequently, our proposed model of time-varying coefficients with SV is general and, by nature, accounts for all forms of the HARL miss-specifications. It captures the dynamic statistical characteristics of $logRV$ and its relationship with the lagged observations.

## 3.3 Methodology

In this section we describe the two variations of the introduced model. The general framework is:

$$y_t = x_t'\beta_t + v_t, \quad v_t \sim \mathscr{N}(0, \sigma_v^2), \quad t = 1, \ldots, n, \tag{3.6}$$

$$\beta_{t,j} = \alpha_j + \rho_j \beta_{t-1,j} + \varepsilon_{tj}, \quad j = 1, \ldots, k, \quad t = 1, \ldots, n, \tag{3.7}$$

In the first specification, we assume:

$$\varepsilon_{tj} \sim \mathscr{N}\left(0, \sigma_{\varepsilon_j}^2\right), \quad j = 1, \ldots, k, \quad t = 1, \ldots, n, \tag{3.8}$$

In the second one, we assume stochastic volatility (SV) :

$$\varepsilon_{tj} | h_{tj} \sim \mathscr{N}\left(0, h_{tj}\right), \quad j = 1, \ldots, k, \quad t = 1, \ldots, n, \tag{3.9}$$

where

$$\log h_{tj} = \gamma_j + \delta_j \log h_{t-1,j} + u_{tj}, \quad j = 1, \ldots, k, \quad t = 1, \ldots, n, \tag{3.10}$$

$$u_{tj} \sim \mathcal{N}(0, \sigma_{uj}^2), \quad j = 1, \ldots, k, \quad t = 1, \ldots, n. \tag{3.11}$$

We treat $h_0$ and $\beta_{0j}$ as unknown parameters. Note that the model can be extended to include autoregressive structure, deterministics such as a constant, (linear or nonlinear) trends, or seasonal dummies. Also, other specifications for time-varying heteroskedasticity are possible.

Utilising the introduced models in forecasting volatility, we set $y_t = log(RV_t^{(d)})$ and $x_t = (log(RV_{t-1}^{(d)}), log(RV_{t-1}^{(w)}), log(RV_{t-1}^{(m)}))'$. Hence, the first variation, denoted by SHARP, is a state space HARL model where the coefficients are time-varying that follow an AR(1) process. The second variation, denoted by SHARP-sv, is the same as the SHARP model with the additional feature of SV in the AR process of the coefficients.

The *measurement equation* (3.6) describes the vector of observations, $log(RV_t^{(d)})$, in terms of the independent variables, the state vector, $\beta_t$, and the disturbances, $v_t$. The *"coefficients" transition equation* (3.7) describes the evolution of the coefficients over time. Indeed, the AR process of the intercept also captures the autocorrelation of the residuals in the measurement equation. Under stationarity, while the unconditional first and second moments of $\beta_{jt}$ are constant, its conditional second moment, $\text{var}(\beta_{jt}|h_{jt})$, can change over time. The *"variance" transition equation* (3.10) describes the change over time of the logarithmic transformation of the conditional variance of the innovations, $\varepsilon_{tj}$, in the *"coefficients" transition equation*. The latter feature allows not only for heteroskedasticity in the main equation (through the intercept) but for heteroskedasticity in the coefficients' process as well. Note that, under stationarity, the first and second moments of $h_{jt}$ are constant (see Appendix 3.A).

To estimate the model, we use a modification of the sequential Monte Carlo method known as the particle Gibbs (PG) sampler, see Andrieu, Doucet, and Holenstein (2010). The latent variables in our model are $\lambda_t = [\beta_t', h_t']'$ where $\beta_t = [\beta_{t1}, \ldots, \beta_{tk}]$ and $h_t = [h_{t1}, \ldots, h_{tk}]$ whose prior can be described by $p(\lambda_t|\lambda_{t-1}, \theta)$. The joint posterior is $p(\theta, \lambda_{1:T}|\mathbf{y}_{1:T})$. In the PG sampler, we can draw the structural parameters as usual, from their posterior conditional distributions $\theta|\lambda_{1:T}, \mathbf{y}_{1:T}$ (see Appendix 3.B). This is important because, in this way, we can avoid mixture approximations or other Monte Carlo procedures that need considerable tuning and may not have good convergence properties. The latent variables can be integrated out of the joint posterior using the Creal and Tsay (2015) procedure (see Appendix 3.C).

Our choice of conjugate prior for this particular empirical exercise is as follows:

$$
\begin{gathered}
\alpha_j, \gamma_j \sim \mathscr{N}(0,1), \quad j = 1, \ldots, k, \\
\rho_j, \delta_j \sim \mathscr{N}(0.5,1)\mathbb{I}_{\rho_j \in (0,1)}\mathbb{I}_{\delta_j \in (0,1)}, \quad j = 1, \ldots, k, \\
\sigma_{uj}^2 \sim \Gamma(6.5, 0.005), \quad j = 1, \ldots, k \\
\sigma_v^2 \sim \Gamma(6.5, 0.005)
\end{gathered}
$$

Using the above prior, we restrict the estimation range of $\rho$ and $\delta$ between $(0,1)$ to guarantee stationarity and reflect the belief that coefficients are positively autocorrelated. The prior of $\gamma$ and $\delta$ can be left more flexible in this exercise with a prior mean of 0. The conjugate prior specification of the variances, $\sigma_u^2$ and $\sigma_v^2$, has a low mean.

**Algorithm of Forecasting $RV$ using SHARP or SHARP-sv**

Let $n$ be the total number of observations, $is$ be the in-sample estimation window size (approximately four years), $nsim = 15000$ be the total number of MCMC iterations, and $nburn = 5000$ be the number burned iterations.

- For $T = is, .., n-1$

  A. For $i = 1, .., nsim$

  i Draw $\lambda_t$ for $t = is - T + 1, .., T$ as illustrated in Appendix (3.C)

  ii Sample the parameters of (3.7) and (3.10) using their posterior distributions in Appendix (3.B)

  iii Forecast $\widehat{logRV}_{T+1,i}$ for $i = nburn + 1, ..., nsim$

  B. Estimate $\widehat{logRV}_{T+1} = \Sigma_{i=nburn+1}^{isim} \frac{\widehat{logRV}_{T+1,i}}{isim-nburn}$

Forecasts of the realised volatility are then computed based on the expectation of a log-normal distribution, as follows:

$$\hat{RV}_{t+1} = \exp(\widehat{log(RV_{t+1})} + \frac{\hat{\omega}^2_{t+1|t}}{2}) \tag{3.12}$$

where:

$$\begin{aligned} \hat{\omega}^2_{t+1|t} = \ & \hat{\sigma}^2_v + \widehat{Var}(\beta_{1,t+1|t}) + (log(RV^d_t))^2 \times \widehat{Var}(\beta_{2,t+1|t}) \\ & + (log(RV^w_t))^2 \times \widehat{Var}(\beta_{3,t+1|t}) + (log(RV^m_t))^2 \times \widehat{Var}(\beta_{4,t+1|t}) \end{aligned}$$

Here, the first term in the expression of $\hat{\omega}^2_{t+1|t}$ is the variance of the measurement equation whereas the subsequent terms represent the variance of each of the state equations entering through the coefficients in the measurement equation. In the SHARP model, $\widehat{Var}(\beta_{j,t+1|t}) = \widehat{\sigma}^2_{\epsilon,j}$ whereas in the SHARP-sv model, $\widehat{Var}(\beta_{j,t+1|t}) = exp(\widehat{log(h_{j,t+1}}) + \hat{\sigma}^2_{u_j}/2)$, for $j = 1, ..., 4$.

## 3.4 Empirical Study

### 3.4.1 Data

The data relates to 4277 trading days of SPY-ETF and twenty NYSE individual stocks RV from 03/01/2000 to 31/12/2016 computed from tick level price observations obtained from TickWrite[2]. The cleaned data makes our results easier to authenticate and replicate. The selected assets allow examining the model performance to predict the price volatility of the market and diverse individual stocks from multiple economic sectors. Also, the sample range enables us to explore the forecasting performance of the models across different market regimes, including the 2008 global financial crisis (GFC). Table (3.1) presents the descriptive statistics of $RV$ of SPY-ETF and the twenty NYSE individual stocks.

[2]TickWrite is a database that provides data on a commercial basis for futures, Index, and equity markets. Tick Data is sourced from NYSE's TAQ (Trade and Quote) database. Tick adjusts the TAQ database for ticker mapping, code filtering, price splits, and dividend payments https://www.tickdata.com/.

Table 3.1: Descriptive statistics for $RV$ of SPY and selected twenty NYSE individual stocks.

| **Stock names** | **Ticker** | **Sector** | **Mean** | **Median** | **St. Dev.** | **Skewness** | **Kurtosis** | **Min** | **Max** |
|---|---|---|---|---|---|---|---|---|---|
| SPDR S&P 500 ETF Trust | SPY | | 1.0372 | 0.4853 | 2.2589 | 10.3174 | 172.6643 | 0.0128 | 59.8630 |
| Apple Inc. | AAPL | Information Technology | 5.2923 | 2.5776 | 7.8061 | 4.4878 | 37.8812 | 0.0791 | 126.1716 |
| Constellation Energy Group | AEE | Utilities | 1.8275 | 1.0614 | 3.4532 | 14.5795 | 362.0346 | 0.1089 | 113.4878 |
| Brown-Forman Corp. | BFB | Consumer Staples | 1.9203 | 1.1517 | 4.8608 | 32.2490 | 1449.2850 | 0.0742 | 240.4141 |
| BT Group plc (ADR) | BT | Communications Services | 2.3113 | 1.1621 | 3.2277 | 4.7116 | 47.6503 | 0.1004 | 59.5677 |
| Exelon Corp. | EXC | Utilities | 2.6354 | 1.4288 | 4.7911 | 9.4223 | 161.6486 | 0.1585 | 130.8746 |
| Freeport-McMoran | FCX | Materials | 8.0344 | 4.3266 | 12.2786 | 5.6708 | 50.4463 | 0.3168 | 188.5795 |
| General Dynamics | GD | Industrials | 2.2370 | 1.2810 | 3.2592 | 6.2650 | 67.3871 | 0.0807 | 63.2822 |
| General Electric | GE | Industrials | 3.0201 | 1.3030 | 6.9818 | 10.3833 | 172.3605 | 0.1077 | 180.3886 |
| The Home Depot | HD | Consumer Discretionary | 3.1214 | 1.5733 | 4.9381 | 6.7444 | 83.0200 | 0.1557 | 103.4768 |
| TECO Energy | HES | Energy | 4.4738 | 2.5702 | 8.6746 | 12.7967 | 280.6983 | 0.2109 | 271.5113 |
| Humana Inc. | HUM | Health Care | 6.6787 | 2.6090 | 11.3665 | 4.4673 | 33.5169 | 0.2404 | 157.5287 |
| IBM | IBM | Information Technology | 2.0255 | 0.9862 | 3.5274 | 7.4383 | 92.6942 | 0.1019 | 71.2926 |
| Coca-Cola | KO | Consumer Staples | 1.5608 | 0.8355 | 2.5353 | 8.6219 | 138.4686 | 0.0456 | 58.8085 |
| Marriott Int'l. | MAR | Consumer Discretionary | 3.5370 | 1.7819 | 5.3635 | 5.4536 | 55.7000 | 0.1543 | 104.5781 |
| Nucor Corp. | NUE | Materials | 4.9098 | 2.7544 | 10.5817 | 13.8696 | 279.0002 | 0.3337 | 266.8244 |
| Pfizer | PFE | Health Care | 2.3324 | 1.3819 | 3.2242 | 6.4660 | 77.5016 | 0.1498 | 62.6970 |
| AT&T | T | Communications Services | 2.6549 | 1.1840 | 4.7673 | 9.4248 | 195.7955 | 0.1082 | 141.8456 |
| Travelers -Travelers Group Inc | TRV | Financials | 2.9683 | 1.1863 | 7.8664 | 15.2588 | 379.9201 | 0.1020 | 263.9287 |
| Wells Fargo | WFC | Financials | 4.3023 | 1.3299 | 12.1389 | 8.2667 | 94.9689 | 0.1036 | 226.6092 |
| ExxonMobil | XOM | Energy | 1.9866 | 1.1409 | 3.9555 | 15.8314 | 430.8596 | 0.1067 | 141.1297 |

### 3.4.2 Empirical Estimation of the SHARP and SHARP-sv models

This section reports the estimation results of the SHARP and SHARP-sv models. While the flexible specification of the coefficients as an AR process (eq. 3.7) is possible, we report results using $\rho = 0.99$, $\delta = 0.99$, $\gamma = 0$, and $\delta = 0$. The general AR process does not yield significantly better out-of-sample results in this exercise. The time-varying estimates of the coefficients using the SHARP and SHARP-sv models are plotted in Figure (3.1) along with horizontal lines representing the full-sample OLS estimates of the simple HARL model (eq. 3.4). Further, the correlation matrix in Table (3.2) depicts the relationship between the estimated time-varying coefficients. The SHARP and SHARP-sv models result in a comparative estimation of the coefficients.

Firstly, the plot of the intercept coefficient, $\hat{\beta}_{1,t}$, has a wide range, implying that there may be time-varying factors beyond the lagged $logRV$ terms that explain the variation of daily $logRV$. Also, $\hat{\beta}_{1,t}$ is remarkably below the corresponding constant-coefficient during the tranquil periods. Secondly, the daily and weekly coefficients are negatively correlated. For example, the daily coefficient is more prominent during periods of uncertainty, such as the GFC period, while the opposite effects emerge in the weekly coefficient, which declines during such periods. On the other hand, the monthly coefficient is (weekly) positively correlated with the daily coefficient, increasing during the GFC period. During times of uncertainty in the financial market, the observed phenomenon can be attributed to the primacy and recency effects. The current long-term conditions, reflected in the monthly average of $logRV^d$, can be regarded as the primary information. The recent information is depicted in the daily $logRV^d$.

When predicting short-term (daily) volatility during times of uncertainty, primary and recent information become more pertinent than information in the middle, such as the weakly average volatility.

Figure 3.1: Smoothed posterior means of $\beta_{tj}$ parameters of SHARP and SHARP-sv models along with the corresponding constant-coefficient by the HARL model estimated on SPY-ETF realised variance in the period January 1st, 2000 to December 31st, 2016.

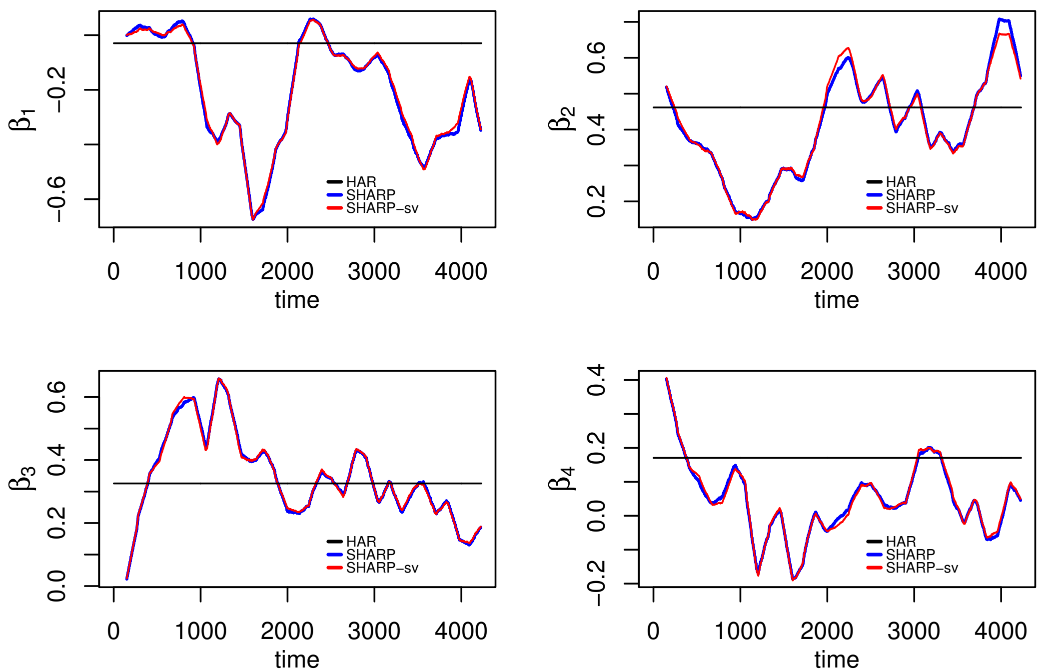
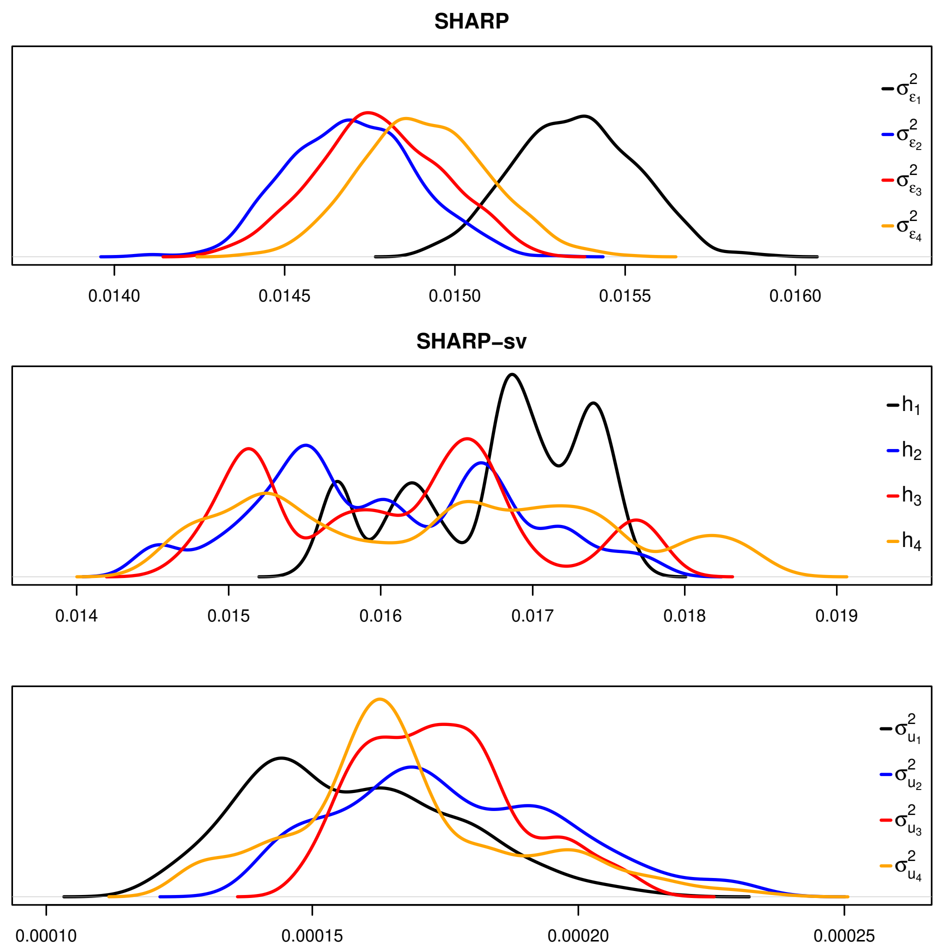



Table 3.2: Correlation matrix of $\beta_{tj}$ parameters obtained by estimating the SHARP and SHARP-sv models on SPY-ETF realised variance in the period January 1st, 2000 to December 31st, 2016.

| | SHARP | | | | | SHARP-sv | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ |
| $\hat{\beta}_1$ | 1 | 0.1924 | -0.0780 | 0.5014 | $\hat{\beta}_1$ | 1 | 0.2144 | -0.0951 | 0.5088 |
| $\hat{\beta}_2$ | 0.1924 | 1 | -0.6284 | 0.2382 | $\hat{\beta}_2$ | 0.2144 | 1 | -0.6353 | 0.2563 |
| $\hat{\beta}_3$ | -0.0780 | -0.6284 | 1 | -0.2734 | $\hat{\beta}_3$ | -0.0951 | -0.6353 | 1 | -0.2909 |
| $\hat{\beta}_4$ | 0.5014 | 0.2382 | -0.2734 | 1 | $\hat{\beta}_4$ | 0.5088 | 0.2563 | -0.2909 | 1 |

Figure 3.2: Marginal posterior densities of $\hat{\sigma}^2_{\epsilon_j}$ in SHARP as well as $\hat{h}_{t,j}$ and $\hat{\sigma}^2_{u_j}$ in SHARP-sv obtained by estimating the models on SPY-ETF realised variance in the period January 1st, 2000 to December 31st, 2016.




We also plot the marginal posterior densities of the variance of the residuals in each state equation (3.7) and (3.10) in Figure (3.2). The density plots of $\hat{h}_{t,j}$ have slightly higher means and variances than that of the $\hat{\sigma}^2_{\epsilon_j}$. Also, it is evident from the plots that the SHARP-sv model depicts the presence of heteroskedasticity in the variances of the coefficients' state equations. In both models, we observe that the mean of the variance of the state equation of the intercept coefficient ($\beta_1$) is higher than that of the other coefficients ($\beta_2,\ \beta_3,\ \beta_4$).

Lastly, we use the Bayes factor to assess if sufficient evidence in the SPY-ETF $logRV$ data series supports the introduced features represented by the statistical models: SHARP and SHARP-sv. Firstly, we compute the Bayes factor to compare the two hypotheses "$H_1$ : model's coefficients ($\beta$'s) are constant", represented by the HARL model, and "$H_2$ : model's coefficients ($\beta$'s) are time-varying", represented by the SHARP model. Secondly, we compare the hypothesis "$H_3$ : model's coefficients are time-varying with SV", represented by the SHARP-sv model, and "$H_2$". Assigning equal priors to the two models under consideration, the Bayes factor can then be computed as $BF_{i,j} = \frac{P(Data|H_i)}{P(Data|H_j)}$, where $P(Data|H_k)$ is the probability of observing the data under the model representing hypothesis $H_k$. A higher value of the Bayes factor means greater evidence for $H_i$ against $H_j$. We follow Berger and Pericchi (1996)'s interpretation of twice the natural logarithm of $BF_{i,j}$ in Table (3.3).

According to the reported results, we conclude that there is evidence favouring time-varying parameters with SV. However, the evidence of $H_2$ against $H_1$ is stronger than $H_3$ against $H_2$. Also, the evidence of each of the time-varying and SV features is stronger in the second sub-sample period, which coincides with the GFC period. Hence, all the preliminary findings are promising for both the SHARP and SHARP-sv models, where, in line with Chen et al. (2018) and Buccheri and Corsi (2021), it shows supporting evidence that the coefficients of the HARL model are time-varying with SV.

Table 3.3: Bayes Factor obtained by estimating the HARL, SHARP, and SHARP-sv models on SPY-ETF realised variance in the period January 1st, 2000 to December 31st, 2016

| **2 log($\mathbf{B_{2,1}}$)** | **Decision** | **2 log($\mathbf{B_{3,2}}$)** | **Decision** |
|---|---|---|---|
| | January 2004 to December 2006 | | |
| $8.1258 > 6$ | Strong evidence for H2 against H1 | $2.1789 > 2$ | Positive evidence for H3 against H2 |
| | January 2007 to December 2010 | | |
| $12.7844 > 10$ | Very Strong evidence for H2 against H1 | $2.9059 > 2$ | Positive evidence for H3 against H2 |
| | January 2011 to December 2016 | | |
| $6.5801 > 6$ | Strong evidence for H2 against H1 | $2.7126 > 2$ | Positive evidence for H3 against H2 |

### 3.4.3 Out-of-sample Comparative analyses

This section presents an out-of-sample analysis of the introduced SHARP and SHARP-sv models and a set of comparable models outlined below in forecasting the daily $RV$ of SPY-ETF and twenty NYSE individual stocks. We report the out-of-sample forecasts using the rolling window approach with an estimation window of 1000 daily observations (approximately four years). Hence, the in-sample period is 1000 observations coinciding roughly from January 2000 to December 2003. The out-of-sample period starts from January 2004 to December 2016. We perform the out-of-sample (OOS) analyses over three sub-periods relative to the GFC period. The first sub-sample spans from January 2004 to December 2006; the second sub-sample covers the GFC period, the most volatile period in our data sample, starting from January 2007 to December 2010; the last sub-sample is from January 2011 until December 2016.

There is extensive evidence in the literature showing the HARL, i.e. modelling the $logRV$, yields better $RV$ forecasts than the HAR model estimated using the $RV$ series. The HARL and all the other models estimated on the log series are less affected by the huge peaks of $RV$. Therefore, we consider the HARL as our benchmark model. This paper's set of competing methods consists of 1) the HARL model (Corsi, 2009), eq. (3.4), is estimated by OLS. 2) HARLQ (inspired from Bollerslev, Patton, and Quaedvlieg (2016) and Buccheri and Corsi (2021)) is a HARL model with time-varying daily coefficient where $\beta_{2,t} = \beta_2 + \gamma \frac{\sqrt{RQ_{t-1}}}{RV_{t-1}}$, estimated by OLS. 3) HARSL (Bekierman and Manner, 2018), which is a HARL model with a time-varying daily coefficient that follows an AR(1) process, estimated by maximum likelihood using the Kalman filter. 4) TVCHAR (Chen et al., 2018) is a HARL model with time-varying coefficients of unspecified functional form, estimated by the local linear method. 5) SHARK (Buccheri and Corsi, 2021), a HARL model with heteroskedastic disturbances and score-driven parameters, estimated by maximum likelihood using the Kalman filter. 6) SHARP model, proposed in this paper, is a HARL model with time-varying coefficients that follow an AR(1) process, estimated by particle Gibbs sampling. 7) SHARP-sv model is similar to the SHARP model with the additional feature of SV in the process of the coefficients.

The selected competing set allows us to evaluate the proposed SHARP and SHARP-sv models and their recent rivals to highlight new comparisons across these models. For example, i) HARL vs HARLQ: knowing that the measurement error of $logRV$ series still exists, we explore whether accounting for this measurement error would improve the forecasts by HARL as evident in the case of HARQ vs HAR in Bollerslev, Patton, and Quaedvlieg (2016). ii) HARSL vs HARLQ: outlines whether modelling the daily coefficient as an AR(1) process is better in forecasting than the more restrictive yet straightforward approach of the HARLQ model. iii) TVCHAR vs SHARK vs SHARP: demonstrates which coefficients' time-varying specification yields better forecasts.

Finally, iv) SHARP vs SHARP-sv: reveals whether the additional SV feature in the AR process of the coefficients improves the forecasts.

In Table 3.4, we report the out-of-sample average of the estimated coefficients used at each forecasting point by each model. The $logRV$ series has lower measurement errors, implying a minor impact on the magnitude of the coefficients than when modelling the $RV$ series. The average values of the coefficients across the models are within a close range. The variances of the coefficients are the lowest for the HARL model and the highest for the SHARK model. As we pointed out in the previous subsection 3.4.2, the daily and monthly coefficients' estimates increase during the GFC period. In contrast, the weekly coefficient estimate decreases. Also, the coefficients' variances increase during that period.

Next, we compare the models' accuracy in forecasting the one-day-ahead $RV$. We use two popular unit-free loss functions (see, Wang, Wu, and Xu (2015) and Zhang, Ma, and Liao (2020), among others): (i) the heteroskedasticity-adjusted version of the mean squared error, $HMSE = \frac{1}{n-k}\Sigma_{t=k+1}^{n}(1 - \frac{\hat{y_t}}{y_t})^2$, (iii) heteroskedasticity-adjusted version of the mean absolute error, $HMAE = \frac{1}{n-k}\Sigma_{t=k+1}^{n}|1 - \frac{\hat{y_t}}{y_t}|$. Both HMSE and HMAE quantify the relative forecast error with respect to the true objective value, $RV$.

Table 3.4: Average and (standard deviation) of coefficients' out-of-sample estimates used at each forecasting point obtained by estimating the HAR, HARL, HARQ, HARLQ, HARSL, TVCHAR, SHARK, SHARP, and SHARP-sv models on SPY-ETF realised variance over a rolling window of 1000 observations.

January 2004 to December 2016

| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
|---|---|---|---|---|---|---|---|
| $\bar{\hat{\beta}}_1$ | -0.0674 | -0.0672 | -0.0518 | -0.0699 | -0.0808 | -0.0695 | -0.0693 |
| | (0.0691) | (0.0701) | (0.0393) | (0.0722) | (0.0858) | (0.0721) | (0.0733) |
| $\bar{\hat{\beta}}_2$ | 0.4225 | 0.4257 | 0.4332 | 0.4225 | 0.5404 | 0.4222 | 0.4220 |
| | (0.1010) | (0.0988) | (0.1114) | (0.1021) | (0.1191) | (0.1024) | (0.1033) |
| $\bar{\hat{\beta}}_3$ | 0.3633 | 0.3579 | 0.3500 | 0.3632 | 0.3043 | 0.3724 | 0.3731 |
| | (0.0886) | (0.0852) | (0.0830) | (0.0896) | (0.0877) | (0.0938) | (0.0948) |
| $\bar{\hat{\beta}}_4$ | 0.1393 | 0.1410 | 0.1572 | 0.1377 | 0.0572 | 0.1308 | 0.1311 |
| | (0.0386) | (0.0394) | (0.0427) | (0.0391) | (0.0731) | (0.0346) | (0.0380) |

January 2004 to December 2006

| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
|---|---|---|---|---|---|---|---|
| $\bar{\hat{\beta}}_1$ | -0.0255 | -0.0247 | -0.0278 | -0.0271 | -0.1085 | -0.0278 | -0.0279 |
| | (0.0205) | (0.0217) | (0.0222) | (0.0224) | (0.0680) | (0.0227) | (0.0270) |
| $\bar{\hat{\beta}}_2$ | 0.2703 | 0.2849 | 0.2891 | 0.2678 | 0.4064 | 0.2668 | 0.2670 |
| | (0.0260) | (0.0450) | (0.0929) | (0.0246) | (0.0527) | (0.0246) | (0.0296) |
| $\bar{\hat{\beta}}_3$ | 0.4951 | 0.4794 | 0.4708 | 0.4978 | 0.3942 | 0.5171 | 0.5173 |
| | (0.0389) | (0.0539) | (0.0537) | (0.0364) | (0.0503) | (0.0342) | (0.0379) |
| $\bar{\hat{\beta}}_4$ | 0.1871 | 0.1887 | 0.1990 | 0.1858 | 0.1075 | 0.1678 | 0.1683 |
| | (0.0136) | (0.0139) | (0.0326) | (0.0124) | (0.0364) | (0.0117) | (0.0200) |

January 2007 to December 2010

| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
|---|---|---|---|---|---|---|---|
| $\bar{\hat{\beta}}_1$ | -0.0668 | -0.0675 | -0.0486 | -0.0697 | -0.0277 | -0.0680 | -0.0673 |
| | (0.0708) | (0.0722) | (0.0416) | (0.0759) | (0.0698) | (0.0754) | (0.0769) |
| $\bar{\hat{\beta}}_2$ | 0.4343 | 0.4354 | 0.4735 | 0.4367 | 0.5447 | 0.4370 | 0.4363 |
| | (0.0835) | (0.0840) | (0.0933) | (0.0847) | (0.1360) | (0.0845) | (0.0852) |
| $\bar{\hat{\beta}}_3$ | 0.3669 | 0.3641 | 0.3394 | 0.3650 | 0.3252 | 0.3697 | 0.3700 |
| | (0.0446) | (0.0402) | (0.0516) | (0.0459) | (0.0754) | (0.0471) | (0.0494) |
| $\bar{\hat{\beta}}_4$ | 0.1241 | 0.1238 | 0.1329 | 0.1211 | 0.0128 | 0.1194 | 0.1199 |
| | (0.0286) | (0.0288) | (0.0221) | (0.0281) | (0.0979) | (0.0275) | (0.0322) |

January 2011 to December 2016

| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
|---|---|---|---|---|---|---|---|
| $\bar{\hat{\beta}}_1$ | -0.0882 | -0.0878 | -0.0656 | -0.0909 | -0.1027 | -0.0911 | -0.0911 |
| | (0.0739) | (0.0746) | (0.0382) | (0.0762) | (0.0874) | (0.0764) | (0.0774) |
| $\bar{\hat{\beta}}_2$ | 0.4894 | 0.4882 | 0.4769 | 0.4888 | 0.6033 | 0.4888 | 0.4885 |
| | (0.0332) | (0.0408) | (0.0597) | (0.0336) | (0.0641) | (0.0338) | (0.0377) |
| $\bar{\hat{\beta}}_3$ | 0.2963 | 0.2942 | 0.2978 | 0.2961 | 0.2462 | 0.3030 | 0.3044 |
| | (0.0428) | (0.0428) | (0.0412) | (0.0432) | (0.0628) | (0.0432) | (0.0460) |
| $\bar{\hat{\beta}}_4$ | 0.1260 | 0.1292 | 0.1529 | 0.1251 | 0.0623 | 0.1202 | 0.1204 |
| | (0.0336) | (0.0354) | (0.0426) | (0.0349) | (0.0436) | (0.0338) | (0.0369) |

The relative loss measure of a model, M, over the benchmark model, HARL, using a loss function, LF (i.e. HMSE or HMAE), is defined as:

$$RL_M := \frac{LF_M}{LF_{HAR}} \tag{3.13}$$

$RL_M < 1$ indicates that Model M outperforms the benchmark model. We report the $RL_M$ for each model, $M$, using SPY-ETF data in Table (3.5) and their average over the twenty NYSE individual stocks in Table (3.6). We also present the box plot of the $RL_M$ of each model, M, using the twenty NYSE individual stocks data in Figure (3.3).

To test the significance of the forecasting gains of a model, we employ the Model Confidence Set (MCS) developed by Hansen, Lunde, and Nason (2011). We test the null hypothesis that all the models are equally good against the alternative that there is a smaller subset of superior models. We follow Hansen, Lunde, and Nason (2011) and choose 0.1 as the critical p-value based on the range statistics. Hence, models with p-values below the significance level, $\alpha = 0.1$, are excluded from the superior subset, denoted by $\hat{M}_{90\%}$. Using the SPY-ETF RV data, in Table (3.5), an asterisk indicates the model is included in the $\hat{M}_{90\%}$. Table (3.6) reports the number of individual financial stocks (out of 20) where a model is included in the $\hat{M}_{90\%}$.

Table 3.5: Out-of-sample relative loss measure (3.13) of HARL, HARLQ, HARSL, TVCHAR, SHARK, SHARP, and SHARP-sv models obtained by estimating the models on SPY-ETF realised variance over a rolling window of 1000 observations in the period January 1st, 2000 to December 31st, 2016.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| January 2004 to December 2006 | | | | | | | |
| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| HMSE | 1 | 1.0134 | 0.8994* | 0.9547 | 0.8918* | 0.8608* | **0.8520***|
| HMAE | 1 | 1.0044 | 0.9468* | 0.9777 | 0.9399* | 0.9403* | **0.9349*** |
| January 2007 to December 2010 | | | | | | | |
| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| HMSE | 1 | 1.0060 | 0.9867 | 0.8832 | 1.1296 | 0.7454* | **0.7062*** |
| HMAE | 1 | 1.0005 | 0.9905 | 0.9428 | 1.0444 | 0.8672* | **0.8638*** |
| January 2011 to December 2016 | | | | | | | |
| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| HMSE | 1 | 1.0072 | 0.9525 | 0.9140 | 0.9176 | 0.8827 | **0.8236*** |
| HMAE | 1 | 1.0004 | 0.9768 | 0.9552 | 0.9636 | 0.9397 | **0.9187*** |

* indicates that the model is included in $\hat{M}_{90\%}$.

Table 3.6: Average of the out-of-sample relative loss measure (3.13) of HARL, HARLQ, HARSL, TVCHAR, SHARK, SHARP, and SHARP-sv models obtained by estimating the models on twenty individual NYSE stocks' realised variance over a rolling window of 1000 observations in the period January 1st, 2000 to December 31st, 2016.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| January 2004 to December 2006 | | | | | | | |
| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| HMSE | $1^{(0)}$ | $0.9779^{(0)}$ | $0.9732^{(0)}$ | $0.9351^{(0)}$ | $0.9252^{(0)}$ | $0.5923^{(20)}$ | $\mathbf{0.5631}^{(20)}$ |
| HMAE | $1^{(0)}$ | $0.9941^{(0)}$ | $0.9871^{(0)}$ | $0.9673^{(0)}$ | $0.9640^{(0)}$ | $0.7935^{(20)}$ | $\mathbf{0.7812}^{(19)}$ |
| January 2007 to December 2010 | | | | | | | |
| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| HMSE | $1^{(0)}$ | $0.9648^{(0)}$ | $1.0167^{(1)}$ | $0.8873^{(5)}$ | $1.0832^{(0)}$ | $0.7302^{(18)}$ | $\mathbf{0.5622}^{(20)}$ |
| HMAE | $1^{(0)}$ | $0.9877^{(0)}$ | $0.9952^{(0)}$ | $0.9459^{(3)}$ | $1.0356^{(0)}$ | $0.8478^{(19)}$ | $\mathbf{0.7915}^{(20)}$ |
| January 2011 to December 2016 | | | | | | | |
| | HARL | HARLQ | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| HMSE | $1^{(0)}$ | $0.9748^{(0)}$ | $0.9623^{(1)}$ | $0.9452^{(0)}$ | $0.9523^{(1)}$ | $0.6046^{(20)}$ | $\mathbf{0.5652}^{(20)}$ |
| HMAE | $1^{(0)}$ | $0.9946^{(0)}$ | $0.9807^{(1)}$ | $0.9748^{(0)}$ | $0.9807^{(1)}$ | $0.7978^{(20)}$ | $\mathbf{0.7788}^{(20)}$ |

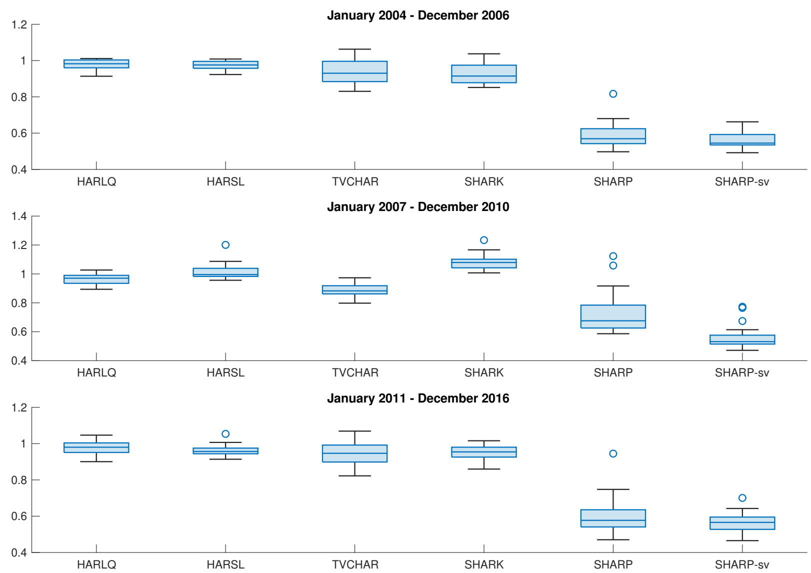
We show in parenthesis the number of times each model is included in $\hat{M}_{90\%}$

Generally, the results obtained using the $RV$ of SPY-ETF and individual stocks are consistent. Correcting for the measurement error in the HARL model relatively improves the forecasts, on average, since the average RL of the HARLQ is less than one. However, the HARLQ does not consistently outperform the HARL model across the twenty individual stocks and the SPY-ETF. The HARSL model, on the other hand, yields moderate forecast improvements compared to the HARLQ, meaning a general time-varying specification such as the AR process for the daily coefficient is better than the restrictive correction in the daily coefficient based on the measurement error variance. It indicates that other factors beyond the measurement error lead to changes in the daily coefficient. Hence, an AR process allows for such flexibility.

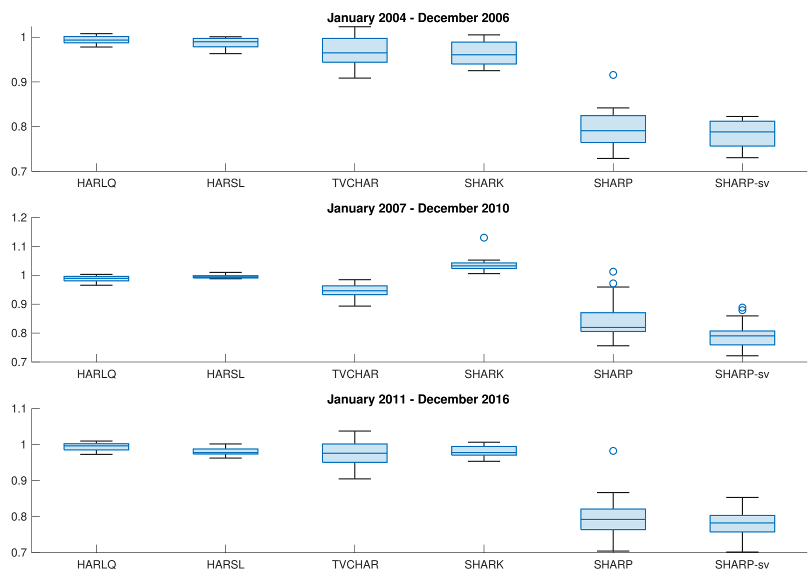
Further, the TVCHAR model is more promising than the HARSL model. The former has all its coefficients time-varying, not just the daily coefficient. However, as pointed out by their respective authors, both the TVCHAR and HARSL models do not consistently outperform the HARL model for the one-day ahead forecasts. The SHARK model is better than the TVCHAR model in the first and second subsample periods. Still, the forecasting accuracy of the SHARK model deteriorates while that of TVCHAR improves during the GFC period.

Figure 3.3: Box plot of the out-of-sample relative loss measure (3.13) of HARL, HARLQ, HARSL, TVCHAR, SHARK, SHARP, and SHARP-sv models obtained by estimating the models on twenty individual NYSE stocks' realised variance over a rolling window of 1000 observations in the period January 1st, 2000 to December 31st, 2016.

(a) HMSE




(b) HMAE



For most individual stocks, some form of time-varying specification of the coefficients generally improves the model's forecasting. However, only the SHARP-sv model's forecasting performance is in all cases superior to the HARL model. The SHARP and SHARP-sv models have the lowest RL using the SPY-ETF data and average RL using the twenty individual stocks data. Adding the SV feature to the AR process of the coefficients yields a further moderate improvement compared to the forecasts by the SHARP model. Hence, we find that the general specification of the time-varying coefficients leads to the most significant improvement in the forecasts of $RV$.

The findings based on the model confidence set reveal that our new extensions, SHARP and SHARP-sv, are always included in the confidence set. It excludes the remaining models with few exceptions. For example, between January 2004 and December 2006, the HARSL and SHARK models are included in the confidence set using the SPY-ETF data but not the individual stocks data. Also, the TVCHAR model is in the superior set during the GFC period for a few individual stocks.

Lastly, we use the *unconditional* predictive ability (uCPA) test by Giacomini and White (2006) to evaluate the out-of-sample predictions produced by the models. It builds on Diebold and Mariano (1995) test and provides a framework for an unconditional forecast evaluation criterion robust to misspecified forecasting models. Our forecast evaluation involves seven models with various specifications and estimation methods. Giacomini and White (2006) test does not impose restrictions on the estimation methods used and allows the models to be nested or non-nested. In Tables (3.7) and (3.9), we report the results for pairwise tests of uncodnitional predictive ability over three OOS periods with significance level of $\alpha = 0.05$. Based on the results, we rank the models from best to worse in Tables (3.8) and (3.10).

Table 3.7: Results of the uCPA test for out-of-sample forecasting performance between the models with the selected loss function (HMSE) obtained by estimating the models on SPY-ETF realised variance over a rolling window of 1000 observations in the period January 1st, 2000 to December 31st, 2016. The model that is significantly better than the other is stated inside the table unless both models are equally good.

| | | | | | | |
|---|---|---|---|---|---|---|
| January 2004 to December 2006 | | | | | | |
| **vs** | **HARLQ** | **HARSL** | **TVCHAR** | **SHARK** | **SHARP** | **SHARP-sv** |
| **HARL** | Equally good | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| **HARLQ** | | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| **HARSL** | | | HARSL | Equally good | Equally good | Equally good |
| **TVCHAR** | | | | SHARK | SHARP | SHARP-sv |
| **SHARK** | | | | | Equally good | Equally good |
| **SHARP** | | | | | | Equally good |
| January 2007 to December 2010 | | | | | | |
| **vs** | **HARLQ** | **HARSL** | **TVCHAR** | **SHARK** | **SHARP** | **SHARP-sv** |
| **HARL** | Equally good | Equally good | TVCHAR | HARL | SHARP | SHARP-sv |
| **HARLQ** | | Equally good | TVCHAR | HARLQ | SHARP | SHARP-sv |
| **HARSL** | | | TVCHAR | HARSL | SHARP | SHARP-sv |
| **TVCHAR** | | | | TVCHAR | SHARP | SHARP-sv |
| **SHARK** | | | | | SHARP | SHARP-sv |
| **SHARP** | | | | | | SHARP-sv |
| January 2011 to December 2016 | | | | | | |
| **vs** | **HARLQ** | **HARSL** | **TVCHAR** | **SHARK** | **SHARP** | **SHARP-sv** |
| **HARL** | Equally good | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| **HARLQ** | | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| **HARSL** | | | TVCHAR | Equally good | SHARP | SHARP-sv |
| **TVCHAR** | | | | Equally good | Equally good | SHARP-sv |
| **SHARK** | | | | | Equally good | SHARP-sv |
| **SHARP** | | | | | | SHARP-sv |

Table 3.8: Ranking of models from best to worse in every subsample period according to the uCPA results in Table (3.7). Models appearing on the same row in a given period of time are equally good.

| January 2004 to December 2006 | January 2007 to December 2010 | January 2011 to December 2016 |
|---|---|---|
| SHARP-sv, SHARP, SHARK, HARSL | SHARP-sv | SHARP-sv |
| TVCHAR | SHARP | SHARP, TVCHAR, SHARK |
| HARL, HARLQ | TVCHAR | SHARK, HARSL |
| | HARL, HARLQ, HARSL | HARL, HARLQ |
| | SHARK | |

Table 3.9: Results of the uCPA test for out-of-sample forecasting performance between the models with the selected loss function (HMAE) obtained by estimating the models on SPY-ETF realised variance over a rolling window of 1000 observations in the period January 1st, 2000 to December 31st, 2016. The model that is significantly better than the other is stated inside the table unless both models are equally good.

| January 2004 to December 2006 | | | | | | |
|---|---|---|---|---|---|---|
| **vs** | **HARLQ** | **HARSL** | **TVCHAR** | **SHARK** | **SHARP** | **SHARP-sv** |
| **HARL** | Equally good | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| **HARLQ** | | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| **HARSL** | | | HARSL | Equally good | Equally good | Equally good |
| **TVCHAR** | | | | SHARK | SHARP | SHARP-sv |
| **SHARK** | | | | | Equally good | Equally good |
| **SHARP** | | | | | | Equally good |
| January 2007 to December 2010 | | | | | | |
| **vs** | **HARLQ** | **HARSL** | **TVCHAR** | **SHARK** | **SHARP** | **SHARP-sv** |
| **HARL** | Equally good | Equally good | TVCHAR | HARL | SHARP | SHARP-sv |
| **HARLQ** | | Equally good | TVCHAR | HARLQ | SHARP | SHARP-sv |
| **HARSL** | | | TVCHAR | HARSL | SHARP | SHARP-sv |
| **TVCHAR** | | | | TVCHAR | SHARP | SHARP-sv |
| **SHARK** | | | | | SHARP | SHARP-sv |
| **SHARP** | | | | | | Equally good |
| January 2011 to December 2016 | | | | | | |
| **vs** | **HARLQ** | **HARSL** | **TVCHAR** | **SHARK** | **SHARP** | **SHARP-sv** |
| **HARL** | Equally good | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| **HARLQ** | | HARSL | TVCHAR | SHARK | SHARP | SHARP-sv |
| **HARSL** | | | TVCHAR | SHARK | SHARP | SHARP-sv |
| **TVCHAR** | | | | Equally good | SHARP | SHARP-sv |
| **SHARK** | | | | | SHARP | SHARP-sv |
| **SHARP** | | | | | | SHARP-sv |

Table 3.10: Ranking of models from best to worse in every subsample period according to the CPA results in Table 3.9. Models appearing on the same row in a given period of time are equally good.

| January 2004 to December 2006 | January 2007 to December 2010 | January 2011 to December 2016 |
|---|---|---|
| SHARP-sv, SHARP, SHARK, HARSL | SHARP-sv, SHARP | SHARP-sv |
| TVCHAR | TVCHAR | SHARP |
| HARL, HARLQ | HARSL,HARLQ | TVCHAR, SHARK |
| | HARLQ, HARL | HARSL |
| | SHARK | HARLQ, HARL |

The SHARP-sv consistently ranks among the best models. The SHARP model ranks as good as the SHARP-sv or the second best. During the second and third subsample periods, the SHARP-sv is first exclusively. While the SHARK model is among the top in the first and third sub-sample periods, its performance deteriorates during the GFC period. The TVCHAR ranks better than the SHARK except for the first subsample period and is always superior to the HARL and HARLQ models. Also, the HARSL ranks better than the HARL and HARLQ models except during the GFC period. Lastly, the HARLQ and HARL models are equally good.

## 3.5 Conclusion

We propose a dynamic state-space model that defines time-varying parameters as an AR process with heteroskedastic innovations: the log-transformation of the conditional variance of those innovations is an AR process. Thereby, we allow for time-varying heteroscedasticity in the measurement equation and the state process of coefficients. We follow Creal and Tsay (2015) in using particle filtering as a computationally efficient approach (Andrieu, Doucet, and Holenstein, 2010). In building upon three recent analyses of the dynamic features of the HARL (HAR with the log of RV) model coefficients (Bekierman and Manner, 2018; Buccheri and Corsi, 2021; Chen et al., 2018), the SHARP and SHARP-sv models incorporate additional innovations with a feasible estimation method.

With our empirical study, we apply the SHARP and SHARP-sv models to forecast the volatility of financial stock returns. Comparisons are made between the out-of-sample forecasting performance of the models, across three sub-sample periods, including the financial crisis of 2008. For a robustness check, we evaluate forecasts using two loss functions. We also use the Model Confidence Set by Hansen, Lunde, and Nason (2011) and conditional predictive ability test by Giacomini and White (2006) to assess the significance of our results.

First: While the literature shows that accounting for the measurement error variance improves the forecast by the HAR model, its failure to do likewise by the HARL may be attributed to reduced measurement errors when using the logarithmic form. Second: Permitting the coefficients to be time-varying, with time-varying conditional variance, improves the forecasts of $RV$. By their overall improved performance, the proposed SHARP and SHARP-sv models significantly outperform other extensions of the HARL model in forecasting the realised volatility of financial stocks.

# Appendices

## 3.A Derivations of the first and second moments

For conciseness we will assume for what follows that $k = 1$ so the transition equations are defined as:

$$\beta_t = \alpha + \rho\beta_{t-1} + \varepsilon_t, \quad \varepsilon_t|h_t \sim \mathcal{N}(0, h_t) \tag{3.14}$$

and

$$\ln h_t = \gamma + \delta \ln h_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2) \tag{3.15}$$

With $\delta < 1$, one can show that:

$$\mu := E(\ln h_t) = \frac{\gamma}{1-\delta^2} \text{ and } \omega^2 := var(\ln h_t) = \frac{\sigma_u^2}{1-\delta^2} \tag{3.16}$$

$$E(h_t) = \exp(\mu)\exp(\omega^2/2) \text{ and } var(h_t) = \exp(2\mu)\left(\exp(2\omega^2) - \exp(\omega^2)\right) \tag{3.17}$$

With $\rho < 1$, one can show that:

$$E(\beta_t) = \frac{\alpha}{1-\rho^2} \tag{3.18}$$

The conditional second moment (variance) of $\beta_t$ could change over time:

$$\begin{aligned} var(\beta_t|h_t) &= \sum_{i=0}^{\infty} \rho^{2i} h_{t-i} \\ &= \exp(\mu) \sum_{i=0}^{\infty} \rho^{2i} \exp(S_{t-i}) \end{aligned}$$

Where $S_t = \sum_{i=0}^{\infty} \delta^i u_{t-i} \sim N(0, \omega^2)$. Note that, since $|\delta| < 1$, $S_t$ converges in mean square to some finite number as $T \to \infty$ provided that $\sum_{i=0}^{\infty} (\delta^i)^2 < \infty$. Similarly, since $|\rho^2| < 1$, $\sum_{i=0}^{T} \rho^{2i} \exp(S_{t-i})$ converges in mean square to some finite number as $T \to \infty$ provided that $\sum_{i=0}^{\infty} (\rho^{2i})^2 < \infty$.

Finally, one can show that the unconditional variance of $\beta_t$ is constant:

$$\begin{aligned} var(\beta_t) &= E(var(\beta_t|h_t)) + var(E(\beta_t|h_t)) \\ &= \frac{\exp(\mu + \omega^2/2)}{1-\rho^2} \end{aligned}$$

## 3.B Posterior Derivation

**Model:**

$$y_t = x_t'\beta_t + v_t; \quad v_t \sim \mathcal{N}(0, \sigma_v^2), \quad t = 1, \ldots, n$$

$$\beta_{tj} = \alpha_j + \rho_j \beta_{t-1,j} + \varepsilon_{tj}; \quad \varepsilon_{tj}|h_{tj} \sim \mathcal{N}\left(0, h_{tj}\right), \quad j = 1, \ldots, k,\ t = 1, \ldots, n$$

$$\log h_{tj} = \gamma_j + \delta_j \log h_{t-1,j} + u_{tj}; \quad u_{tj} \sim \mathcal{N}(0, \sigma_{uj}^2), \quad j = 1, \ldots, k,\ t = 1, \ldots, n$$

$$\textbf{Let} \quad \lambda_t = [\beta_t', logh_t']'$$

$$\textbf{where} \quad \beta_t = [\beta_{t1}, \ldots, \beta_{tk}] \quad \textbf{and} \quad logh_t = [logh_{t1}, \ldots, logh_{tk}]$$

$$\textbf{Let} \quad \theta = \{\theta_1, ..., \theta_k\}, \quad \textbf{where} \quad \theta_j = \{\alpha_j, \rho_j, \gamma_j, \delta_j, \sigma_{uj}\}, \quad j = 1, \ldots, k,$$

**Prior**

$$\alpha_j, \gamma_j \sim \mathcal{N}(0, 1), \quad j = 1, \ldots, k,$$

$$\rho_j, \delta_j \sim \mathcal{N}(0.5, 1)\mathbb{I}_{\rho_j \in (0,1)}\mathbb{I}_{\delta_j \in (0,1)}, \quad j = 1, \ldots, k,$$

$$\sigma_{uj}^2 \sim \Gamma(6.5, 0.005), \quad j = 1, \ldots, k$$

$$\sigma_v^2 \sim \Gamma(6.5, 0.005)$$

Equivalently:

$$p(\alpha_j) \propto exp(-\tfrac{n}{2}\alpha_j^2), \quad j = 1, \ldots, k$$

$$p(\gamma_j) \propto exp(-\tfrac{n}{2}\gamma_j^2), \quad j = 1, \ldots, k$$

$$p(\rho_j) \propto exp(-\tfrac{n}{2}(\rho_j - 0.5)^2)\mathbb{I}_{\rho_j \in (0,1)}, \quad j = 1, \ldots, k$$

$$p(\delta_j) \propto exp(-\tfrac{n}{2}(\delta_j - 0.5)^2)\mathbb{I}_{\delta_j \in (0,1)}, \quad j = 1, \ldots, k$$

$$p(\sigma_{uj}) \propto \sigma_{uj}^{-(\underline{n}+1)} exp\left(-\frac{\underline{q}}{2\sigma_{uj}^2}\right), \quad j = 1, \ldots, k, \quad \underline{n} = 10 \quad \textbf{and} \quad \underline{q} = 0.01$$

$$p(\sigma_v) \propto \sigma_{uj}^{-(\underline{n}+1)} exp\left(-\frac{\underline{q}}{2\sigma_{uj}^2}\right), \quad \underline{n} = 10 \quad \textbf{and} \quad \underline{q} = 0.01$$

**Posterior**

$$P\left(\theta, \sigma_v, \{\lambda\}_{t=1}^{T}|Y\right) \propto (\sigma_v^{-n})exp(-\frac{1}{2\sigma_v^2}\sum_{t=1}^{n}(y_t - x_t'\beta_t)^2)$$
$$(\prod_{i=1}^{k}\prod_{t=1}^{n} h_{it}^{-1})exp(-\sum_{i=1}^{k}\sum_{t=1}^{n}\frac{1}{2h_{it}^2}(\beta_{it} - \alpha_i - \rho_i\beta_{i,t-1})^2)$$
$$(\prod_{i=1}^{k}\sigma_{u_i}^{-n})exp(-\sum_{i=1}^{k}\frac{1}{2\sigma_{u_i}^2}\sum_{t=1}^{n}(logh_{it} - \gamma_{it} - \delta_{it}logh_{i,t-1})^2)$$
$$\prod_{j=1}^{k}p(\sigma_{u_i})\prod_{i=1}^{k}p(\alpha_i)\prod_{i=1}^{k}p(\gamma_j)\prod_{i=1}^{k}p(\rho_i)\prod_{i=1}^{k}p(\delta_i)$$

1. For $P(\theta|\{\lambda\}_{t=1}^{T}, Y)$, for $i = 1, .., k$ :

2. For $\gamma_i\Big|\left\{\{\theta\backslash\{\gamma_i\}\}, \sigma_v, \{\lambda\}_{t=1}^{T}, Y\right\}$ we use mixed estimator:

   let $Z = \begin{bmatrix} logh_{i,2} - \delta_i logh_{i,1} \\ . \\ . \\ logh_{i,T} - \delta_i logh_{i,T-1} \end{bmatrix}$ and $u_i = \begin{bmatrix} u_{i,2} \\ . \\ . \\ u_{i,T} \end{bmatrix}$

   then, $\begin{bmatrix} Z \\ 0 \end{bmatrix} = \begin{bmatrix} 1_{(n\times 1)} \\ 1 \end{bmatrix}\gamma_i + \begin{bmatrix} u_i \\ \eta \end{bmatrix}$ ; $\begin{bmatrix} u_i \\ \eta \end{bmatrix} \sim N(0, \begin{bmatrix} \sigma_i I_n & \\ & 1 \end{bmatrix})$

   where $n = T - 1$

   $\Longrightarrow \gamma_i\Big|\left\{\{\theta\backslash\{\gamma_i\}\}, \{\lambda\}_{t=1}^{T}, Y\right\} \sim N(\frac{(logh_{i,T} - logh_{i,1})}{1+\sigma_i^2}, \frac{\sigma_i^2}{1+\sigma_i^2})$

3. For $\delta_i\Big|\left\{\{\theta\backslash\{\delta_i\}\}, \sigma_v, \{\lambda\}_{t=1}^{T}, Y\right\}$, we use mixed estimator:

let $Z = \begin{bmatrix} logh_{i,2} - \gamma_i \\ . \\ . \\ logh_{i,T} - \gamma_i \end{bmatrix}$, $\kappa = \begin{bmatrix} logh_{i,1} \\ . \\ . \\ logh_{i,T-1} \end{bmatrix}$ and $u_i = \begin{bmatrix} u_{i,2} \\ . \\ . \\ u_{i,T} \end{bmatrix}$

then, $\begin{bmatrix} Z \\ 0.5 \end{bmatrix} = \begin{bmatrix} \kappa \\ 1 \end{bmatrix} \delta_i + \begin{bmatrix} u_i \\ \eta \end{bmatrix}$ ; $\begin{bmatrix} u_i \\ \eta \end{bmatrix} \sim N(0, \begin{bmatrix} \sigma_i I_n & \\ & 1 \end{bmatrix})$

where $n = T - 1$

$$\Longrightarrow \delta_i \Big| \Big\{\{\theta \backslash \{\delta_i\}\}, \{\lambda\}_{t=1}^T, Y\Big\} \sim N\left(\frac{\kappa' Z + 0.5\sigma_i^2}{\kappa'\kappa + \sigma_i^2}, \frac{\sigma_i^2}{\kappa'\kappa + \sigma_i^2}\right)$$

4.

$$P\left(\sigma_{u_i} \Big| \{\theta \backslash \{\sigma_{u_i}\}\}, \sigma_v, \{\lambda\}_{t=1}^T, Y\right)$$

$$\propto$$

$$\sigma_{u_i}^{-n} exp\left(-\frac{1}{2\sigma_{u_i}^2} \sum_{t=1}^n (logh_{it} - \gamma_i - \delta_i logh_{i,t-1})^2\right) \quad \sigma_{u_i}^{-(\underline{n}+1)} exp\left(-\frac{\underline{q}}{2\sigma_{u_i}^2}\right)$$

$$\propto$$

$$\sigma_{u_i}^{-(n+\underline{n}+1)} \quad exp\left(-\frac{1}{2\sigma_{u_i}^2} \left(\sum_{t=1}^n (logh_{it} - \gamma_i - \delta_i logh_{i,t-1})^2 + \underline{q}\right)\right)$$

$$\Longrightarrow \frac{\sum_{t=1}^n (logh_{it} - \gamma_i - \delta_i logh_{i,t-1})^2 + \underline{q}}{\sigma_{u_i}^2} \Big| \Big\{\{\theta \backslash \{\sigma_i\}\}, \{\lambda\}_{t=1}^T, Y\Big\} \sim \chi^2(n + \underline{n} + 3)$$

5. For $\alpha_i \Big| \Big\{\{\theta \backslash \{\alpha_i\}\}, \{\lambda\}_{t=1}^T, Y\Big\}$ we use mixed estimator:

let $Z = \begin{bmatrix} \beta_{i,2} - \rho_i \beta_{i,1} \\ . \\ . \\ \beta_{i,T} - \rho_i \beta_{i,T-1} \end{bmatrix}$, $H_i = \begin{bmatrix} h_{i,1} & & & \\ & . & & \\ & & . & \\ & & & h_{i,T} \end{bmatrix}$, and $\epsilon_i = \begin{bmatrix} \epsilon_{i,2} \\ . \\ . \\ \epsilon_{i,T} \end{bmatrix}$

then, $\begin{bmatrix} Z \\ 0 \end{bmatrix} = \begin{bmatrix} 1_{(n\times 1)} \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} \epsilon_i \\ \eta \end{bmatrix}$ ; $\begin{bmatrix} u_i \\ \eta \end{bmatrix} \sim N(0, \begin{bmatrix} H_i & \\ & 1 \end{bmatrix})$
where $n = T - 1$

$$\Longrightarrow \alpha_i \Big| \Big\{\{\theta \backslash \{\alpha_i\}\}, \{\lambda\}_{t=1}^{T}, Y\Big\} \sim N(\frac{\Sigma_{t=2}^{T} \frac{Z_{i,t}}{h_{i,t}}}{\Sigma_{t=2}^{T} \frac{1}{h_{i,t}} + 1}, \frac{1}{\Sigma_{t=2}^{T} \frac{1}{h_{i,t}} + 1})$$

6. For $\rho_i \Big| \Big\{\{\theta \backslash \{\rho_i\}\}, \sigma_v, \{\lambda\}_{t=1}^{T}, Y\Big\}$ we use mixed estimator:

let $Z = \begin{bmatrix} \beta_{i,2} - \alpha_i \\ . \\ . \\ \beta_{i,T} - \alpha_i \end{bmatrix}$, $\kappa = \begin{bmatrix} \beta_{i,1} \\ . \\ . \\ \beta_{i,T-1} \end{bmatrix}$, $H_i = \begin{bmatrix} h_{i,1} & & & \\ & . & & \\ & & . & \\ & & & h_{i,T} \end{bmatrix}$,

and $\epsilon_i = \begin{bmatrix} \epsilon_{i,2} \\ . \\ . \\ \epsilon_{i,T} \end{bmatrix}$

then, $\begin{bmatrix} Z \\ 0.5 \end{bmatrix} = \begin{bmatrix} \kappa \\ 1 \end{bmatrix} \rho_i + \begin{bmatrix} \epsilon_i \\ \eta \end{bmatrix}$ ; $\begin{bmatrix} u_i \\ \eta \end{bmatrix} \sim N(0, \begin{bmatrix} H_i & \\ & 1 \end{bmatrix})$
where $n = T - 1$

$$\Longrightarrow \rho_i \Big| \Big\{\{\theta \backslash \{\rho_i\}\}, \{\lambda\}_{t=1}^{T}, Y\Big\} \sim N(\frac{\Sigma_{t=2}^{T} \frac{Z_{i,t}}{h_{i,t}} + 0.5}{\Sigma_{t=2}^{T} \frac{1}{h_{i,t}} + 1}, \frac{1}{\Sigma_{t=2}^{T} \frac{1}{h_{i,t}} + 1})$$

7.

$$P\left(\sigma_v \middle| \{\theta\}, \{\lambda\}_{t=1}^{T}, Y\right)$$

$$\propto$$

$$\sigma_v^{-n} exp(-\frac{1}{2\sigma_v^2}\sum_{t=1}^{n}(y_t - x_t'\beta_t)^2) \quad \sigma_v^{-(\underline{n}+1)} exp\left(-\frac{\underline{q}}{2\sigma_v^2}\right)$$

$$\propto$$

$$\sigma_v^{-(n+\underline{n}+1)} \quad exp(-\frac{1}{2\sigma_i^2}\left(\sum_{t=1}^{n}(y_t - x_t'\beta_t)^2 + \underline{q}\right))$$

$$\Longrightarrow \frac{\sum_{t=1}^{n}(y_t - x_t'\beta_t)^2 + \underline{q}}{\sigma_v^2}\Bigg|\left\{\{\theta\}, \{\lambda\}_{t=1}^{T}, Y\right\} \sim \chi^2(n + \underline{n} + 3)$$

## 3.C Particle filtering within MCMC

Particle filtering is a simulation-based algorithm that sequentially approximates continuous marginal distributions using discrete distributions. This is performed by using a set of support points called "particles" and probability masses; see Creal (2012) for a review. The PG sampler draws a single path of the latent or state variables from this discrete approximation. As the number of particles M goes to infinity, the PG sampler draws from the exact full conditional distribution. The advantage of the algorithm is that it allows for drawing paths of the state variables in large blocks.

As mentioned in Creal and Tsay (2015), the PG sampler is a standard Gibbs sampler but defined on an extended probability space where a particle filter generates all the random variables. Chopin and Singh (2015) analysed the theoretical properties of the PG sampler, and showed that the sampler is uniformly ergodic. Unlike the standard particle filter, the PG sampler involves a "conditional" resampling algorithm in the last step. Namely, for draws from the particle filter to be a valid Markov transition kernel on the extended probability space, the state variables drawn at the previous iteration must have a positive sampling probability (Andrieu, Doucet, and Holenstein, 2010). The conditional resampling step within the PG forces the pre-existing path to survive the particle filter's resampling steps. We use the conditional multinomial resampling algorithm from Andrieu, Doucet, and Holenstein (2010), although other resampling algorithms exist (see, for example, Fearnhead et al. (2010) and Chopin and Singh (2015)).

Suppose the posterior is $p(\theta, \lambda_{1:T}|\mathbf{y}_{1:T})$ where $\lambda_{1:T}$ denotes the latent variables whose prior can be described by $p(\lambda_t|\lambda_{t-1}, \theta)$. In the PG sampler, we can draw the structural parameters $\theta|\lambda_{1:T}, \mathbf{y}_{1:T}$ as usual, from their posterior conditional distributions. This is important because, in this way, we can avoid mixture approximations or other Monte Carlo procedures that need considerable tuning and may not have good convergence properties. Suppose we have $\lambda_{1:T}^{(1)}$ from the previous iteration. The particle filtering procedure consists of two phases, forward and backward filtering.

Phase I: Forward filtering (Andrieu, Doucet, and Holenstein, 2010).

*For* $t = 1, \ldots, T$

- Draw a proposal $\lambda_t^{(m)}$ from an importance density $q(\lambda_t|\lambda_{t-1}^{(m)}, \theta), \quad m = 2, ..., M.$
- Compute the importance weights:

$$w_t^{(m)} = \frac{p(y_t|\lambda_t^{(m)}, \theta)p(\lambda_t^{(m)}|\lambda_{t-1}^{(m)}, \theta)}{q(\lambda_{it}|\lambda_{t-1}^{(m)}, \theta)}, \quad m = 1, ..., M. \tag{3.19}$$

- Normalise the weights: $\tilde{w}_t^{(m)} = \frac{w_t^{(m)}}{\sum_{m=1}^{M} w_t^{(m)}}, \quad m = 1, ..., M.$
- Re-sample, conditionally, the particles $\{\lambda_t^{(m)}, \; m = 1, ..., M\}$ with probabilities

$\{\tilde{w}_t^{(m)},\ m = 1, ..., M\}$.

In the original PG sampler, the particles are stored for $t = 1, ..., T$ and a single trajectory is sampled using the probabilities from the last iteration. An improvement upon the original PG by drawing the path of the latent variables from the particle approximation is using the backwards sampling algorithm of Godsill, Doucet, and West (2004). In the forward pass, we store the normalised weights and particles then we draw a path of the latent variables as we detail below (the draws are from a discrete distribution).

Phase II: Backward filtering (Chopin and Singh, 2015; Godsill, Doucet, and West, 2004).

- At time $t = T$ draw a particle $\lambda_T^* = \lambda_T^{(m)}$.
- Compute the backward weights: $w_{t|T}^{(m)} \propto \tilde{w}_t^{(m)} p(\lambda_{t+1}^* | \lambda_t^{(m)}, \theta)$.
- Normalise the weights: $\tilde{w}_{t|T}^{(m)} = \frac{w_{t|T}^{(m)}}{\sum_{m=1}^{M} w_{t|T}^{(m)}}, m = 1, ..., M$.
- Draw a particle $\lambda_{it}^* = \lambda_t^{(m)}$ with probability $\tilde{w}_{t|T}^{(m)}$.

Therefore, $\lambda_{1:T}^* = \{\lambda_1^*, ..., \lambda_T^*\}$ is a draw from the full conditional distribution. When the state vector dimension is large, we can draw $\lambda_{i,1:T}$, conditional on all other paths $\lambda_{-i,1:T}$ that are not path $i$. Therefore, we can draw from the full conditional distribution $p(\lambda_{i,1:T} | \lambda_{-i,1:T}, \mathbf{y}_{1:T}, \theta)$. The backwards step often results in dramatic improvements in computational efficiency and strictly dominates the original PG (Chopin and Singh, 2015). For example, Creal and Tsay (2015) find that $M = 100$ particles is enough.

There remains the problem of selecting an importance density $q(\lambda_t | \lambda_{t-1}, \theta)$. We use an importance density implicitly defined by $\lambda_{it} = a_i + \sum_{p=1}^{P} b_{i,p}\lambda_{i,t-1}^{p} + \Omega_i\xi_{it}$ where $\xi_{it}$ follows a standard (zero location and unit scale) Student-$t$ distribution with $\nu = 5$ degrees of freedom. That is, we use polynomials in $\lambda_{i,t-1}$ of order $P$. We select the parameters $a_i, b_i$ and $\Omega_i$ during the burn-in phase (using $P = 1$ and $P = 2$) so that the weights $\{\tilde{w}_{it}^{(m)}, m = 1, ..., M\}$ and $\{\tilde{w}_{t|T}^{(m)}, m = 1, ..., M\}$ are approximately not too far from a uniform distribution.

## 3.D HARSL estimation using Kalman Filter

The HARSL equation is:

$$\begin{aligned} lnRV_t &= \beta_1 + (\beta_2 + \lambda_t)lnRV_{t-1}^{(d)} + \beta_3 lnRV_{t-1}^{(w)} + \beta_4 lnRV_{t-1}^{(m)} + \epsilon_t \\ \lambda_{t+1} &= \phi\lambda_t + \eta_t \end{aligned} \tag{3.20}$$

It can be rewritten as:

$$\begin{aligned} lnRV_t &= \beta_1 + \beta_2 lnRV_{t-1}^{(d)} + \beta_3 lnRV_{t-1}^{(w)} + \beta_4 lnRV_{t-1}^{(m)} + \lambda_t lnRV_{t-1}^{(d)} + \epsilon_t \\ \lambda_{t+1} &= \phi\lambda_t + \eta_t \end{aligned} \tag{3.21}$$

We use the fkf package on R to estimate the model. The above model can be rewritten as the general form of the transition and measurement equation for Kalman filter is given by:

$$\begin{aligned} y_t &= c_t + Z_t\alpha_t + G_t.\epsilon_t \\ \alpha_{t+1} &= d_t + T_t\alpha_t + H_t.\eta_t \end{aligned} \tag{3.22}$$

where:

$$y_t = lnRV_t$$

$$Z_t = (1,\ lnRV^{(d)}_{t-1},\ lnRV^{(w)}_{t-1},\ lnRV^{(m)}_{t-1},\ lnRV^{(d)}_{t-1})$$

$$c_t = 0$$

$$\alpha_t = (\beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \lambda_t)'$$

$$G_t = G = \sigma^2_\epsilon$$

$$d_t = 0$$

$$T_t = (1, 1, 1, 1, \phi)'$$

$$H_t = \begin{bmatrix} 0_{(4\times 4)} & \\ & \sigma^2_{\eta_t} \end{bmatrix}$$

# References


*ALFRED* (n.d.). Accessed: 2019-01-01. URL: https://alfred.stlouisfed.org/category?cid=1.

Andersen, T. G. and T. Bollerslev (1998). “Answering the skeptics: Yes, standard volatility models do provide accurate forecasts”. In: *International economic review*, pp. 885–905.

Andersen, T. G., T. Bollerslev, and J. Cai (2000). “Intraday and interday volatility in the Japanese stock market”. In: *Journal of international Financial Markets, institutions and Money* 10.2, pp. 107–130.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). “The distribution of realized exchange rate volatility”. In: *Journal of the American statistical association* 96.453, pp. 42–55.

— (2003a). “Modeling and forecasting realized volatility”. In: *Econometrica* 71.2, pp. 579–625.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003b). “Micro effects of macro announcements: Real-time price discovery in foreign exchange”. In: *American Economic Review* 93.1, pp. 38–62.

Andersen, T. G., T. Bollerslev, and S. Lange (1999). "Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon". In: *Journal of empirical finance* 6.5, pp. 457–477.

Andersen, T. G. and O. Bondarenko (2007). *Construction and interpretation of model-free implied volatility*. Tech. rep. National Bureau of Economic Research.

Andrieu, C., A. Doucet, and R. Holenstein (2010). "Particle markov chain monte carlo methods". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3, pp. 269–342.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009). "Real-time measurement of business conditions". In: *Journal of Business & Economic Statistics* 27.4, pp. 417–427.

Audrino, F. and Y. Hu (2016). "Volatility forecasting: Downside risk, jumps and leverage effect". In: *Econometrics* 4.1, p. 8.

Audrino, F., F. Sigrist, and D. Ballinari (2019). "The impact of sentiment and attention measures on stock market volatility". In: *International Journal of Forecasting*.

Barndorff-Nielsen, O. E. and N. Shephard (2002). "Estimating quadratic variation using realized variance". In: *Journal of Applied econometrics* 17.5, pp. 457–477.

Becker, R., A. E. Clements, and A. McClelland (2009). "The jump component of S&P 500 volatility and the VIX index". In: *Journal of Banking & Finance* 33.6, pp. 1033–1038.

Bekaert, G. and M. Hoerova (2014). "The VIX, the variance premium and stock market volatility". In: *Journal of Econometrics* 183.2, pp. 181–192.

Bekierman, J. and H. Manner (2018). “Forecasting realized variance measures using time-varying coefficient models”. In: *International Journal of Forecasting* 34.2, pp. 276–287.

Beltratti, A. and C. Morana (2006). “Breaks and persistency: macroeconomic causes of stock market volatility”. In: *Journal of econometrics* 131.1-2, pp. 151–177.

Berge, T. J. and Ò. Jordà (2011). “Evaluating the classification of economic activity into recessions and expansions”. In: *American Economic Journal: Macroeconomics* 3.2, pp. 246–77.

Berger, J. O. and L. R. Pericchi (1996). “The intrinsic Bayes factor for model selection and prediction”. In: *Journal of the American Statistical Association* 91.433, pp. 109–122.

Bollerslev, T. (1986). “Generalized autoregressive conditional heteroskedasticity”. In: *Journal of Econometrics* 31.3, pp. 307–327.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016). “Exploiting the errors: A simple approach for improved volatility forecasting”. In: *Journal of Econometrics* 192.1, pp. 1–18.

Bollerslev, T. and V. Todorov (2011). “Tails, fears, and risk premia”. In: *The Journal of Finance* 66.6, pp. 2165–2211.

Bollerslev, T. and H. Zhou (2006). “Volatility puzzles: a simple framework for gauging return-volatility regressions”. In: *Journal of Econometrics* 131.1-2, pp. 123–150.

Bournay, J. and G. Laroque (1979). “Réflexions sur la méthode d’elaboration des comptes trimestriels”. In: *Annales de l’INSEE.* JSTOR, pp. 3–30.

Buccheri, G. and F. Corsi (2021). “HARK the SHARK: Realized volatility modeling with measurement errors and nonlinear dependencies”. In: *Journal of Financial Econometrics*.

Caporin, M., E. Rossi, and P. Santucci de Magistris (2015). “Volatility jumps and their economic determinants”. In: *Journal of Financial Econometrics* 14.1, pp. 29–80.

Casas, I. and R. Fernandez-Casal (2019). “tvReg: Time-varying Coe fficient Linear Regression for Single and Multi-Equations in R”. In: *Available at SSRN*.

Chen, X. B., J. Gao, D. Li, and P. Silvapulle (2018). “Nonparametric estimation and forecasting for time-varying coefficient realized volatility models”. In: *Journal of Business & Economic Statistics* 36.1, pp. 88–100.

Chernov, M. (2007). “On the role of risk premia in volatility forecasting”. In: *Journal of Business & Economic Statistics* 25.4, pp. 411–426.

Chopin, N. and S. S. Singh (2015). “On particle Gibbs sampling”. In: *Bernoulli* 21.3, pp. 1855–1883.

Chow, G. C. and A.-l. Lin (1971). “Best linear unbiased interpolation, distribution, and extrapolation of time series by related series”. In: *The review of Economics and Statistics*, pp. 372–375.

Christiansen, C., M. Schmeling, and A. Schrimpf (2012). “A comprehensive look at financial volatility prediction by economic variables”. In: *Journal of Applied Econometrics* 27.6, pp. 956–977.

Chun, H. and S. Keleş (2010). “Sparse partial least squares regression for simultaneous dimension reduction and variable selection”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.1, pp. 3–25.

Clark, P. K. (1973). “A subordinated stochastic process model with finite variance for speculative prices”. In: *Econometrica: journal of the Econometric Society*, pp. 135–155.

Clark, T. E. and F. Ravazzolo (2015). “Macroeconomic forecasting performance under alternative specifications of time-varying volatility”. In: *Journal of Applied Econometrics* 30.4, pp. 551–575.

Clark, T. E. and K. D. West (2007). “Approximately normal tests for equal predictive accuracy in nested models”. In: *Journal of econometrics* 138.1, pp. 291–311.

Conrad, C. and K. Loch (2015). “Anticipating long-term stock market volatility”. In: *Journal of Applied Econometrics* 30.7, pp. 1090–1114.

Copeland, T. E. (1976). “A model of asset trading under the assumption of sequential information arrival”. In: *The Journal of Finance* 31.4, pp. 1149–1168.

Corradi, V., W. Distaso, and A. Mele (2012). “Macroeconomic determinants of stock market volatility and volatility risk-premiums”. In: *Swiss Finance Institute Research Paper* 12-18.

Corsi, F. (2009). “A simple approximate long-memory model of realized volatility”. In: *Journal of Financial Econometrics* 7.2, pp. 174–196.

Corsi, F. and R. Reno (2009). “HAR volatility modelling with heterogeneous leverage and jumps”. In: *Available at SSRN 1316953*.

Creal, D. (2012). “A survey of sequential Monte Carlo methods for economics and finance”. In: *Econometric reviews* 31.3, pp. 245–296.

Creal, D. and R. S. Tsay (2015). “High dimensional dynamic stochastic copula models”. In: *Journal of Econometrics* 189.2, pp. 335–345.

D'Agostino, A., L. Gambetti, and D. Giannone (2013). "Macroeconomic forecasting and structural change". In: *Journal of applied econometrics* 28.1, pp. 82–101.

Dagum, E. B. and P. A. Cholette (2006). *Benchmarking, temporal distribution, and reconciliation methods for time series*. Vol. 186. Springer Science & Business Media.

Darrat, A. F., M. Zhong, and L. T. Cheng (2007). "Intraday volume and volatility relations with and without public news". In: *Journal of Banking & Finance* 31.9, pp. 2711–2729.

Diebold, F. X. and R. S. Mariano (1995). "Comparing Predictive Accuracy". In: *Journal of Business  Economic Statistics* 13.3, pp. 253–263. ISSN: 07350015. URL: http://www.jstor.org/stable/1392185 (visited on 06/07/2022).

Engle, R. F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

Engle, R. F., E. Ghysels, and B. Sohn (2013). "Stock market volatility and macroeconomic fundamentals". In: *Review of Economics and Statistics* 95.3, pp. 776–797.

Fearnhead, P., O. Papaspiliopoulos, G. O. Roberts, and A. Stuart (2010). "Random-weight particle filtering of continuous time processes". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.4, pp. 497–512.

Fernández, R. B. (1981). "A methodological note on the estimation of time series". In: *The Review of Economics and Statistics* 63.3, pp. 471–476.

Foroni, C., P. Guérin, and M. Marcellino (2018). “Using low frequency information for predicting high frequency variables”. In: *International Journal of Forecasting* 34.4, pp. 774–787.

Fuentes, J., P. Poncela, and J. Rodríguez (2015). “Sparse partial least squares in time series for macroeconomic forecasting”. In: *Journal of Applied Econometrics* 30.4, pp. 576–595.

Ghysels, E. (2016). “Macroeconomics and the reality of mixed frequency data”. In: *Journal of Econometrics* 193.2, pp. 294–314.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). “The MIDAS touch: Mixed data sampling regression models”. In: *UCLA: Finance*.

— (2006). “Predicting volatility: getting the most out of return data sampled at different frequencies”. In: *Journal of Econometrics* 131.1-2, pp. 59–95.

Giacomini, R. and B. Rossi (2009). “Detecting and predicting forecast breakdowns”. In: *The Review of Economic Studies* 76.2, pp. 669–705.

— (2010). “Forecast comparisons in unstable environments”. In: *Journal of Applied Econometrics* 25.4, pp. 595–620.

Giacomini, R. and H. White (2006). “Tests of conditional predictive ability”. In: *Econometrica* 74.6, pp. 1545–1578.

Girard, E. and R. Biswas (2007). “Trading volume and market volatility: Developed versus emerging stock markets”. In: *Financial Review* 42.3, pp. 429–459.

Godsill, S. J., A. Doucet, and M. West (2004). “Monte Carlo smoothing for nonlinear time series”. In: *Journal of the american statistical association* 99.465, pp. 156–168.

Green, T. C. (2004). “Economic news and the impact of trading on bond prices”. In: *The Journal of Finance* 59.3, pp. 1201–1233.

Hamilton, J. D. and G. Lin (1996). "Stock market volatility and the business cycle". In: *Journal of applied econometrics* 11.5, pp. 573–593.

Hamilton, J. D. and R. Susmel (1994). "Autoregressive conditional heteroskedasticity and changes in regime". In: *Journal of econometrics* 64.1-2, pp. 307–333.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). "The model confidence set". In: *Econometrica* 79.2, pp. 453–497.

Harris, L. (1987). "Transaction data tests of the mixture of distributions hypothesis". In: *Journal of Financial and Quantitative Analysis* 22.2, pp. 127–141.

Inoue, A., L. Jin, and B. Rossi (2017). "Rolling window selection for out-of-sample forecasting with time-varying parameters". In: *Journal of Econometrics* 196.1, pp. 55–67.

Kydland, F. E., P. Rupert, and R. Šustek (2016). "Housing dynamics over the business cycle". In: *International Economic Review* 57.4, pp. 1149–1177.

Litterman, R. B. (1983). "A random walk, Markov model for the distribution of time series". In: *Journal of Business & Economic Statistics* 1.2, pp. 169–173.

Liu, L. Y., A. J. Patton, and K. Sheppard (2015). "Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes". In: *Journal of Econometrics* 187.1, pp. 293–311.

McAleer, M. and M. C. Medeiros (2008). "A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries". In: *Journal of Econometrics* 147.1, pp. 104–119.

Müller, U. A., M. M. Dacorogna, R. D. Davé, R. B. Olsen, O. V. Pictet, and J. E. Von Weizsäcker (1997). "Volatilities of different time resolutions—analyzing the dynamics of market components". In: *Journal of Empirical Finance* 4.2-3, pp. 213–239.

*OECD* (n.d.). Accessed: 2019-01-01. URL: https://www.oecd.org/unitedkingdom/.

Officer, R. R. (1973). "The variability of the market factor of the New York Stock Exchange". In: *the Journal of Business* 46.3, pp. 434–453.

Pesaran, M. H. and A. Timmermann (2005). "Small sample properties of forecasts from autoregressive models under structural breaks". In: *Journal of Econometrics* 129.1-2, pp. 183–217.

Poissonnier, A. (2018). "The Chow-Lin method extended to dynamic models with autocorrelated residuals". In: *Journal of Time Series Econometrics* 10.1.

*Policy Uncertainty* (n.d.). Accessed: 2020-01-01. URL: https://www.policyuncertainty.com.

Poon, S.-H. and C. W. J. Granger (2003). "Forecasting volatility in financial markets: A review". In: *Journal of economic literature* 41.2, pp. 478–539.

*QUANDL* (n.d.). Accessed: 2019-01-01. URL: https://www.quandl.com.

Sax, C. and P. Steiner (2013). "Temporal disaggregation of time series". In: *The R Journal* 5.2, pp. 80–87.

Schwert, G. W. (1989). "Why does stock market volatility change over time?" In: *The journal of finance* 44.5, pp. 1115–1153.

Smirlock, M. and L. Starks (1988). "An empirical analysis of the stock price-volume relationship". In: *Journal of Banking & Finance* 12.1, pp. 31–41.

So, M. E. P., K. Lam, and W. K. Li (1998). "A stochastic volatility model with Markov switching". In: *Journal of Business & Economic Statistics* 16.2, pp. 244–253.

Stark, T. (2010). "Realistic evaluation of real-time forecasts in the Survey of Professional Forecasters". In: *Federal Reserve Bank of Philadelphia Research Rap, Special Report* 1.

Swanson, N. R. and H. White (1995). "A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks". In: *Journal of Business & Economic Statistics* 13.3, pp. 265–275.

*TickWrite* (n.d.). Accessed: 2019-01-01. URL: https://www.tickdata.com/.

Wang, X., C. Wu, and W. Xu (2015). "Volatility forecasting: The role of lunch-break returns, overnight returns, trading volume and leverage effects". In: *International Journal of Forecasting* 31.3, pp. 609–619.

Wang, Y., F. Ma, Y. Wei, and C. Wu (2016). "Forecasting realized volatility in a changing world: A dynamic model averaging approach". In: *Journal of Banking & Finance* 64, pp. 136–149.

Wold, H. (1966). "Estimation of principal components and related models by iterative least squares". In: *Multivariate analysis*, pp. 391–420.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). "A tale of two time scales: Determining integrated volatility with noisy high-frequency data". In: *Journal of the American Statistical Association* 100.472, pp. 1394–1411.

Zhang, Y., F. Ma, and Y. Liao (2020). "Forecasting global equity market volatilities". In: *International Journal of Forecasting* 36.4, pp. 1454–1475.