Using Asymmetric Loss Functions in Time Series Econometrics

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Master of Science

Anna Titova
aus Ivanovo

Kiel, 2019

Dekan:
Prof. Dr. Till Requate

Erstberichterstattender:
Prof. Dr. Matei Demetrescu

Zweitberichterstattender:
Prof. Dr. Kai Carstensen

Tag der Abgabe der Arbeit:
16. April 2019

Tag der mündlichen Prüfung:
22. Mai 2019

Für die Frauen der Wissenschaft

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftliche Mitarbeiterin am Institut für Statistik und Ökonometrie der Christian-Albrechts-Universität zu Kiel. Mein besonderer Dank gilt meinem Doktorvater Professor Dr. Matei Demetrescu, dessen erstklassige Betreuung entscheidend zum Gelingen dieser Arbeit beigetragen hat. Herrn Professor Dr. Kai Carstensen danke ich herzlich für die Übernahme des Zweitgutachtens und stets hilfreiche Vorschläge und Bemerkungen.

Abgesehen von meiner Person haben Herr Professor Dr. Matei Demetrescu, Herr Professor Dr. Vasyl Golosnoy und Herr Dr. Christoph Roling ebenfalls an den in dieser Arbeit enthaltenen Kapiteln mitgewirkt.

Des Weiteren möchte ich allen Kollegen am Institut für hilfreiche Kommentare und eine angenehme Zusammenarbeit in freundschaftlich-herzlicher Atmosphäre danken, insbesondere jedoch Herrn Professor Dr. Uwe Jensen und Herrn Benjamin Hillmann.

Kiel, am 15. April 2019 Anna Titova

Contents

List of Abbreviations

ACF autocorrelation function

AIC Akaike information criterion

apARCH asymmetric power autoregressive conditionally heteroskedastic model

AR(p) autoregression of order p

CAViaR conditional autoregressive value-at-risk

DGP data generating process

EU European Union

GAS generalized autoregressive score model

GMM generalized method of moments

HAR heterogeneous autoregression

iff if and only if

IV implied volatility

MCS model confidence set

md martingale difference

MLE maximum-likelihood estimator

MSE mean squared error

NMD Normal mixture distribution

OLS ordinary least-squares

OU Ornstein-Uhlenbeck process

SND skew-Normal distribution

SSM superior set of models

UK United Kingdom

VaR value-at-risk

w.p. 1 with probability 1

w.r.t. with respect to

List of Tables

List of Figures

Chapter 1

Introduction

This thesis serves the purpose of examining the role of asymmetric loss functions in time series analysis. Asymmetric loss functions can sometimes be interpreted as mathematical representations of risk-averse or risk-seeking behaviour of economic agents. This thesis shows how these functions may be used for forecasting certain economic variables. It contains methodological work, statistical simulations as well as empirical studies. It is based on four articles, one of which is currently under review. The thesis is structured as follows.

Chapter 2 deals with building an optimal forecast of a long autoregression under an asymmetric loss. The properties of autoregressive models whose order grows with the sample size make such long autoregressions a popular tool in time series analysis. Among others, this semiparametric approach consistently estimates the linear MSE-optimal point forecast. Should a loss function other than the squared error loss be relevant, e.g. an asymmetric one, it has been argued that estimation should be conducted under the relevant loss. The chapter offers a loss-specific Wold-type decomposition motivating the use of long autoregressions under the relevant loss. It also shows that fitting under relevant loss consistently delivers the linear loss-optimal point forecast. The semiparametric approach is compared to a model-based one in terms of efficiency. We find in Monte Carlo simulations that the semiparametric approach is generally preferable, except for cases where the series to be forecast exhibits strong GARCH effects.

Chapter 3 represents a replication study of Christodoulakis and Mamatzakis (2009, Journal of Applied Econometrics 24, pp. 583-606). In their article, the authors estimate the EU Commission loss preferences for selected economic forecasts of 12 EU Member States. They employ the GMM estimation procedure proposed by Elliott et al. (2005, Review of Economic Studies 72, pp. 1107-1125) and find the forecasts to be somewhat optimistic on average. This chapter shows, however, the GMM estimator to possess non-standard limiting distributions when some of the instruments are highly persistent, which is the case with one of the instruments employed by Christodoulakis and Mamatzakis. Standard distributions are recovered in some interesting particular cases which are relevant in practice. A re-examination of the EU Commission loss preferences using methods robust to persistence and an extended dataset reveals that, while the conclusions of the original study are, by and large, still justified, the EU Commission loss preferences have become more symmetric over the whole studied period.

Chapter 4 addresses the issue of predicting value-at-risk. This quantity is widely used in practice for risk management purposes. In the majority of related literature validation of a VaR, forecasting method is performed using only a few criteria and is often not discriminatory enough. With this part of my research, I propose a class of models that often shows to be superior in terms of forecasting performance. I compare the proposed models to several prominent competitors. Moreover, I combine various validation methods in a manner that hasn't been used before.

Chapter 5 examines certain bias correction techniques for a widely used log transformation of data. In many economic applications, it is convenient to model and forecast the logs rather than the levels of a variable of interest. However, the reverse transformation from log forecasts to levels introduces a bias. This chapter compares different bias correction methods for the reverse transformation of log series which follow a linear process with various types of error distributions. Based on Monte Carlo simulations and an empirical study of realized volatilities, we find that there is no uniformly best choice of a correction method. By and large, a variance-based correction appears to be preferable, but bias corrections may even increase the forecast MSE when the log series exhibits high persistence.

Finally, the last chapter offers a short summary and gives an outlook for further research on the matter.

Chapter 2

Long Autoregressions under Asymmetric Loss

Coauthored by: Matei Demetrescu

2.1 Motivation

Least-squares long autoregressions have been successfully used in time series analysis and forecasting since the seminal work of Berk (1974) and Bhansali (1978); see also Gonçalves and Kilian (2007) for conditional heteroskedasticity, Poskitt (2007) for models with long memory, and Demetrescu and Hassler (2016) for the case with changes in the mean. At the same time, forecasting under asymmetric loss is a relevant task in applied work. E.g. Artis and Marcellino (2001) find IMF and OECD forecasts of the deficit of G7 countries to be systematically biased, which they explain by asymmetric loss preferences of the IMF and the OECD. Clements et al. (2007) and Capistrán (2008) analyze the loss function of the Federal Reserve to find asymmetries in its forecast preferences – and even some time variation thereof. Christodoulakis and Mamatzakis (2008, 2009) find asymmetric preferences of EU institutional forecasts, while Pierdzioch et al. (2011) does the same for the Bank of Canada. See also Wang and Lee (2014) and Tsuchiya (2016) for additional evidence. Individual forecasters are not immune to asymmetric forecast preferences either; see e.g. Elliott et al. (2008), Boero et al. (2008), Aretz et al. (2011), Clatworthy et al. (2012) or Fritsche et al. (2015). A natural question is then, how can one deploy long autoregressions for forecasting under asymmetric loss?

At least since the work of Weiss and Andersen (1984); Weiss (1996), it has been argued that estimation should be conducted using the relevant forecast optimality criterion. Naturally, this suggests estimation of long autoregressions under the relevant loss function. We therefore address the question, what properties do such long autoregressions under generic loss functions have.

Our contributions are as follows. We first derive the theoretical properties of innovations in infinite-order linear autoregressive forecasts under a general loss function. In doing so, we focus on the class of asymmetric power loss functions proposed by Elliott et al. (2005), of which the asymmetric linear or asymmetric quadratic are particular cases. This derivation provides the theoretical underpinning of the use of linear autoregressions under the relevant loss as well as a

Wold-type decomposition which is specific to the loss function considered. Then we address the issue of fitting long autoregressions under the relevant loss. Since the quantile check function is a particular case of the loss functions we consider, we extend in this respect the work of Zernov et al. (2009) who discuss quantile long autoregressions, but without any justification for this class of semiparametric forecast models. Imposing an external loss function has the disadvantage of potential estimation inefficiency; here, model-based approaches may perform better in practice, since, in a more parametric perspective, parameter estimation can be adjusted to take relevant data features into account. See Dumitrescu and Hansen (2016) for a precise discussion of bias vs. variance when estimation is conducted under another criterion than the evaluation. To complete the discussion, we provide a Monte Carlo based comparison of the semiparametric approach based on long autoregressions under the relevant loss with alternative approaches, in particular a location-scale model-based procedure.

Let us set some notation before proceeding. By y_t, $t \in \mathbb{Z}$, we denote the process to be forecast and by $y_t(1)$ the optimal one-step ahead (linear) forecast conditional on the information set $\mathcal{F}_t = \{y_t, y_{t-1}, \ldots\}$ under the relevant loss function, i.e. the forecast minimizing the expected loss of forecasting y_{t+1} given $\mathcal{F}_t$. Forecasts at higher horizon may be generated by direct forecasts; while we do not pursue this topic here, it seems plausible that the main findings remain valid. The loss function evaluating the forecast error is denoted by $\mathcal{L}(\cdot)$, and we take it to be in difference form. The L_r norm of a random variable is given by $\|\cdot\|_p = \sqrt[r]{\mathrm{E}(|\cdot|^r)}$. Moreover, $\|\cdot\|_p$ also denotes the ℓ_p vector norm and the corresponding induced matrix norm. We use $\|\cdot\|$ and $\|\cdot\|_1$ to denote Euclidean and city-block norms whenever no confusion is possible. The probabilistic Landau symbols O_p and o_p have their usual meaning.

2.2 Autoregressive modelling under the relevant loss

We focus on the class of asymmetric loss functions proposed by Elliott et al. (2005): they are quite flexible but do not place strict requirements on existence of moments of the forecast errors, unlike the Linex loss which essentially requires finiteness of moments of any order.

Assumption 2.1 *Let $\mathcal{L} : \mathbb{R} \mapsto \mathbb{R}^+$ be given by*

$$\mathcal{L}(u) = (\alpha + (1 - 2\alpha) \cdot \mathbf{1}(u < 0)) |u|^p ,$$

where $\alpha \in (0,1)$ and $p \in \{1, 2, \ldots\}$, and $\mathbf{1}(\cdot)$ is the usual indicator function.

The assumption covers the popular asymmetric linear (lin-lin) and asymmetric quadratic (quad-quad) losses, and has derivative $\mathcal{L}'(u) = p(\alpha - \mathbf{1}(u < 0)) |u|^{p-1}$ which is continuous for $p > 1$. The parameter α controls the degree of asymmetry of the loss function; $\alpha = 0.5$ recovers a symmetric loss function. The parameter p on the other hand controls the tail behavior of $\mathcal{L}(\cdot)$. The case $p = 1$ leads to the asymmetric linear loss, which is convex and continuous, but not differentiable at 0; this is nothing else than the check function used in quantile regression. For

$p > 1$, $\mathcal{L}$ is strictly convex and piecewise smooth; the second-order derivative is only continuous for $p > 2$, or $p = 2$ and $\alpha = 0.5$.

Let us now examine the task of linearly forecasting y_{t+1} under $\mathcal{L}$ given its infinite past, $y_t, y_{t-1}, \ldots$, i.e. finding

$$y_t(1) = \sum_{j\geq 1} a_j y_{t+1-j} + b, \tag{2.1}$$

for a suitable parameter b and a sequence of parameters $\{a_j\}_{j\in\mathbb{N}\setminus\{0\}}$. A linear modelling approach is quite common for the conditional mean, and we only take the idea one step further to forecasting under asymmetric loss.

By the desired optimality of $y_t(1)$, the coefficients minimize the forecast risk,

$$a_j, b = \underset{a_j^*, b^*}{\arg\min}\ \mathrm{E}\left[\mathcal{L}\left(y_{t+1} - \sum_{j\geq 1} a_j^* y_{t+1-j} - b^*\right)\right]. \tag{2.2}$$

Denote by ε_{t+1} the corresponding forecast error,

$$\varepsilon_{t+1} = y_{t+1} - y_t(1) = y_{t+1} - \sum_{j\geq 1} a_j y_{t+1-j} - b. \tag{2.3}$$

For the case of squared-error loss, one can draw on functional analytic results in Hilbert spaces of weakly stationary processes to analyze the optimum problem in (2.2), taking e.g. advantage of the fact that the covariance may be used to define an inner product. The Projection Theorem then ensures existence and uniqueness of the optimal forecast under quadratic loss, as well as orthogonality of the forecast errors and the predictors y_{t+1-j}, $j \geq 1$. An immediate consequence is lack of serial correlation of the forecast errors. But if $\mathcal{L}(\cdot)$ is not quadratic (or not even symmetric in general), one is not able to use the approach anymore.

An analogous result may however be proved by elementary methods:

Proposition 2.1 *Given a loss function satisfying Assumption 2.1 and a strictly stationary process y_t for which $\mathrm{E}[|y_t|^p] < \infty$, the following statements hold true.*

a) For $p > 1$, the forecast risk $\mathcal{Q}\left(a_j^*, b^*\right) = \mathrm{E}\left[\mathcal{L}\left(y_{t+1} - \sum_{j\geq 1} a_j^* y_{t+1-j} - b^*\right)\right]$ has a unique minimum that satisfies the following set of first-order conditions

$$\mathrm{E}\left[\mathcal{L}'\left(y_{t+1} - \sum_{j\geq 1} a_j y_{t+1-j} - b\right)\right] = \mathrm{E}\left[\mathcal{L}'\left(\varepsilon_{t+1}\right)\right] = 0 \qquad \text{and}$$

$$\mathrm{E}\left[y_{t+1-j}\mathcal{L}'\left(y_{t+1} - \sum_{j\geq 1} a_j y_{t+1-j} - b\right)\right] = \mathrm{E}\left[y_{t+1-j}\mathcal{L}'\left(\varepsilon_{t+1}\right)\right] = 0 \quad \text{for all } j \geq 1.$$

b) For $p = 1$, item a) holds if all finite-dimensional distributions of y_t are absolutely continuous.

Proof: *See the Appendix.*

The so-called generalized forecast error, $\mathcal{L}'(\varepsilon_{t+1})$, is therefore unbiased and uncorrelated with the predictors, in other words it is linearly unpredictable given past levels of the series of interest. Moreover, past forecast errors are linearly noninformative as well, as shown by the following corollary.

Corollary 2.1 *The generalized forecast error, $\mathcal{L}'(\varepsilon_{t+1})$, is uncorrelated with past forecast errors, $\mathrm{E}\left[\varepsilon_{t+1-j}\mathcal{L}'(\varepsilon_{t+1})\right]$ for all $j \geq 1$.*

Proof: *See the Appendix.*

A slightly stricter martingale difference condition (given a set of forecast-relevant information $\mathcal{F}_t$) has often been used in the literature to characterize optimal forecasts (see e.g. Granger, 1999). Such conditions have also been used to set up GMM estimation of the parameters of an unknown loss function on the basis of observed forecast errors assumed to stem from rational forecasts (Elliott et al., 2005). The novelties are here the entirely semiparametric approach to constructing the optimal forecast and the idea of constructing innovations specific to the relevant loss.

Remark 2.1 *The linear forecast can in principle be improved upon in nonlinear setups. Since $\mathcal{L}$ is homogenous of degree p, the optimal conditional point forecast under $\mathcal{L}$ is given as*

$$\bar{y}_t(1) = \mathrm{E}\left[y_{t+1}|\mathcal{F}_t\right] + \bar{b}\sqrt{\mathrm{Var}\left[y_{t+1}|\mathcal{F}_t\right]}; \tag{2.4}$$

see Patton and Timmermann (2007a). Moreover, $\bar{b}$ only depends on the shape of the forecast distribution of y_{t+1} and on $\mathcal{L}$,

$$\bar{b} = \arg\min_{\bar{b}^*} \mathrm{E}\left[\mathcal{L}\left(\frac{y_{t+1} - \mathrm{E}\left[y_{t+1}|\mathcal{F}_t\right]}{\sqrt{\mathrm{Var}\left[y_{t+1}|\mathcal{F}_t\right]}} - \bar{b}^*\right)\right],$$

i.e. $\bar{b}$ is the optimal forecast of the conditionally standardized series. This suggests that forecasts based on location-scale models may be an alternative to long autoregressions under the relevant loss. In practice, the question arises, as to which method should be preferred. In fact, this is just the bias vs. variance discussion in a slightly modified form: if estimation of the model in Equation (2.4) *is noisy, then a linear (mis-)specification may perform better in terms of forecasting performance. At the same time, a linear fit ignores conditional heteroskedasticity so it may be estimated in an inefficient manner. We compare the two approaches in Section 2.4.*

It should be emphasized that different loss functions lead to essentially different linear autoregressive representations. To understand the mechanism, let us examine the following example.

Example 2.1 *Let y_t be a bilinear process, given as*

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-1} u_t,$$

where $u_t \sim iid\,(0,1)$ *and is* L_p*-bounded, and* $\theta_{1,2}$ *are such that* y_t *is strictly stationary (see e.g. Douc et al., 2014, Section 4.3.1, for suitable conditions).*

Take the squared-error loss first, $\mathcal{L} = u^2$*. Now,* y_t *is a conditionally heteroskedastic AR(1) process with martingale difference [md] innovations* $\varepsilon_t = \theta_2 y_{t-1} u_t$*. Under squared-error loss, the optimal point forecast is the conditional mean given by*

$$\mathrm{E}\left[y_{t+1}|y_t, \ldots\right] = \theta_1 y_t,$$

and the optimal conditional forecast under squared error loss is, in the notation of Proposition 2.1, $y_t\,(1) = a_1 y_t$ *with* $a_1 = \theta_1$*.*

Under a loss function $\mathcal{L}$ *with* $p \neq 2$ *or* $\alpha \neq 0.5$*, let* $\bar{b}$ *denote the optimal forecast of* u_t*; cf. Eq. (2.4) (for* $p = 1$*, assume for simplicity that* $\bar{b}$ *is unique). With* $\mathcal{L}'\left(u_t - \bar{b}\right)$ *being zero-mean iid and thus independent of* $y_{t-1}, \ldots$*, rewrite the model as*

$$y_t = a_1 y_{t-1} + \varepsilon_t$$

with $a_1 = \theta_1 + \bar{b}\theta_2$ *and* $\varepsilon_t = \theta_2 y_{t-1}\left(u_t - \bar{b}\right)$*. Notice further that, since* $\bar{b}$ *is the optimal forecast of* u_t*, it holds that*

$$\mathrm{E}\left[\mathcal{L}'\left(\varepsilon_{t+1}\right)|y_t, \ldots\right] = |\theta_2 y_t|^{p-1}\,\mathrm{E}\left[\mathcal{L}'\left(u_{t+1} - \bar{b}\right)|y_t, \ldots\right] = 0,$$

such that ε_{t+1} *satisfies the properties derived in Proposition 2.1 and the optimal forecast is given, under this second choice for* $\mathcal{L}$*, by*

$$y_t\,(1) = a_1 y_t \quad \textit{with} \quad a_1 = \theta_1 + \bar{b}\theta_2.$$

Unless $\bar{b} = 0$ *(or, trivially,* $\theta_2 = 0$*), it holds that* $\theta_1 \neq a_1$ *and the optimal forecast under* $\mathcal{L}$ *is quite different from the forecast under squared-error loss.*

Hence, under any (non-quadratic) loss function for which $\bar{b} \neq 0$*, we still have an AR representation for the process* y_t*, but with* different *coefficients depending on the loss function. Conversely, for* $\theta_1 + \bar{b}\theta_2 = 0$*,* y_t *is linearly unpredictable under* $\mathcal{L}$*, although still serially dependent (in both the conditional mean and the conditional variance).*

There is an obvious exception to this dependence of the AR representation on the loss function: should y_t be an invertible general linear process driven by *iid* innovations, it is straightforward to show that the sequence of autoregressive parameters (but not b) is the same for all strictly convex loss functions.[1] One is tempted to conjecture that the converse holds as well; we leave this question for further research.

In a nutshell, each loss function ultimately leads to a specific understanding of what the innovations (forecast errors) should behave like. In lack of a better notation, we may call the innovations sequence ε_t from Proposition 2.1 $\mathcal{L}$-innovations.

[1] In fact, Granger (1969) exploits this to optimally forecast under $\mathcal{L}$ linear processes with iid innovations; see his so-called two-step procedure.

Note that $\mathcal{L}$-innovations which are not linearly predictable under some loss function need not be unpredictable under another – for the same process y_t. To underscore this, consider the following re-telling of the previous example from the point of view of the innovations sequence.

Example 2.2 *Let ε_t be a GARCH-in-mean process,*

$$\varepsilon_t = \theta_1 \sigma_t + \sigma_t u_t$$

where $\sigma_t = f(\varepsilon_{t-1}, \ldots, u_{t-1}, \ldots)$ such that ε_t is strictly stationary, and $u_t \sim iid\,(0,1)$ is independent of past ε's.

Clearly, ε_t is not uncorrelated (unless σ_t is constant a.s.) and is predictable under squared error loss with one-step ahead optimal forecast given by $y_t(1) = \theta_1 \sigma_t$.

One may however find a suitable loss function under which $\mathcal{L}'(\varepsilon_t)$ is a md sequence and as such ε_t is an $\mathcal{L}$-innovation, concretely when $\theta_1 = -\bar{b}$ with $\bar{b}$ the optimal forecast of u_t under $\mathcal{L}$; still, ε_t is linearly predictable under squared error loss.

Some further remarks are in order.

Remark 2.2 *Unlike the MSE case, the first-order conditions given in Proposition 2.1 cannot be given in terms of autocovariances of y_t, since $\mathcal{L}'$ is nonlinear in general. In the MSE case, weak stationarity of y_t ensures time invariance of this equations system. Here, it is strict stationarity of y_t which guarantees that the system of equations characterizing the solution does not depend on the time index t. Strict stationarity is a sufficient condition for any loss function considered here; for a particular loss, however, time invariance of the first-order conditions may replace strict stationarity. One might call such a property weak $\mathcal{L}$-stationarity. We shall stick however to strict stationarity as it is more convenient not to tie the properties of the data generating process [DGP] to the loss function, as the loss is – exogenously – imposed by the forecaster.*

Remark 2.3 *Examining the proof of Proposition 2.1, we note that the result is actually valid for strictly convex loss functions without additional conditions; for convex (but not strictly convex) loss functions, additional conditions on the distribution of y_t may be required; see the proof for details. Moreover, the analogous result holds for h-step ahead forecasts if using direct forecasts.*

Before moving on to discuss estimation of the autoregressive representation, we may build on the above characterization of $\mathcal{L}$-innovations to provide a linear representation of the process analogous to the Wold decomposition. To this end, let us call a process y_t $\mathcal{L}$-predictable iff its $\mathcal{L}$-innovations ε_t have zero L_p norm, and regular if it is not predictable.

Proposition 2.2 *Define $\mathcal{S}_t$ as the span of $\{1, y_t, y_{t-1} \ldots\}$. Under the conditions of Proposition 2.1, it holds that*

$$y_t = m_t + e_t \quad \forall t \in \mathbb{Z},$$

where $e_t = \sum_{j\geq 0} b_j \varepsilon_{t-j}$ *with for* ε_{t-j} *being the* $\mathcal{L}$*-innovations process of* y_t *and the coefficients are taken as*

$$b_j = \frac{\mathrm{E}\left[y_t \mathcal{L}'\left(\varepsilon_{t-j}\right)\right] - \sum_{k=0}^{j-1} b_k \,\mathrm{E}\left[\varepsilon_{t-k}\mathcal{L}'\left(\varepsilon_{t-j}\right)\right]}{\mathrm{E}\left[\varepsilon_t \mathcal{L}'\left(\varepsilon_t\right)\right]},$$

such that

1. m_{t+h} *and* $\mathcal{L}'\left(\varepsilon_t\right)$ *are uncorrelated* $\forall h \in \mathbb{Z}$*;*
2. $m_t \in \mathcal{S}_{-\infty}$*;*
3. e_t *is a regular process;*
4. m_t *is a predictable process.*

Proof: *See the Appendix.*

It should be emphasized again that the decomposition is loss-function specific, as illustrated in the following:

Example 2.3 *Let* $y_t = s\varepsilon_t$ *with* ε_t *a nondegenerate zero-mean iid sequence, where* s *is random and independent of* ε_t $\forall t$*. Then,* y_t *is white noise provided that* s *and* ε_t *are* L_2*-bounded and the Wold decomposition of* y_t *has no predictable component. Under an asymmetric loss function, let* b *be the conditional optimal point forecast of* ε_t *(assuming that* ε_t *is* L_p*-bounded). We may then write*

$$y_t = m_t + e_t \qquad \text{with} \quad e_t = \left(\varepsilon_t - b\right) s \quad \text{and} \quad m_t = bs.$$

The conditional optimal point predictor of e_t *is easily seen to be zero; therefore,* e_t *is its own sequence of* $\mathcal{L}$*-innovations and thus a regular process under* $\mathcal{L}$*. At the same time,* $m_t = bm$ *is predictable.*

Remark 2.4 *Comparing the statement of Proposition 2.2 with the "classical" Wold decomposition, where one sets* $b_j = \frac{\mathrm{E}(y_t\varepsilon_{t-j})}{\mathrm{E}(\varepsilon_t^2)}$*, the additional term* $-\sum_{k=0}^{j-1} b_k \,\mathrm{E}\left(\varepsilon_{t-k}\mathcal{L}'\left(\varepsilon_{t-j}\right)\right)$ *adjusts for the fact that* ε_{t-k} *may be correlated, even if* $\mathcal{L}'\left(\varepsilon_{t-j}\right)$ *is orthogonal to past* ε_t*. Of course, this term is zero when setting* $\mathcal{L} = u^2$*.*

Remark 2.5 *Unlike for the Wold decomposition, it cannot be stated that the coefficients* b_j *are square summable. To understand why square summability is not available in general, recall the GARCH-in-mean example. Then, it could well be that* σ_t *has long memory (and thus a linear representation without absolutely summable coefficients). Now, the convolution of two filters with square summable coefficients does not exist in general, so in order to still have strict stationarity of* y_t*, the coefficients* b_j *must be restricted beyond square summability.*

To sum up, any strictly stationary, L_p-bounded process possesses an infinite-order linear representation under asymmetric power loss functions, with uniqueness given for $p = 1$ only under additional conditions. This parallels the situation under squared-error loss. But the parameters of the representation depend on the loss function, and the innovations are tailored to the respective loss. We shall now exploit this characterization to provide the theoretical motivation for prediction using long autoregressions under asymmetric loss.

2.3 Fitting long autoregressions under the relevant loss

2.3.1 Model and assumptions

We move on to setting up forecasts given a sample, $y_1, \ldots, y_T$. Concretely, we would like to estimate the coefficients of *the relevant* AR representation given a specific $\mathcal{L}$. Perhaps not surprisingly, we will prove in Section 2.3.2 that this task is accomplished by estimation under the relevant loss; we shall make some regularity assumptions beyond strict stationarity to achieve our goal. Also, we shall not address the case $p = 1$ as it has already been discussed by Zernov et al. (2009). In exchange, this allows us to drop continuity restrictions on the distribution of y_t.

Assumption 2.2 *Let y_t be given by*

$$y_t = \varepsilon_t + \sum_{j\geq 1} b_j \varepsilon_{t-j}, \forall t \in \mathbb{Z},$$

where $B(L) = 1 + \sum_{j\geq 1} b_j L^j$ is an invertible lag polynomial whose coefficients satisfy $\sum_{j\geq 1} |b_j| < \infty$ and ε_t a sequence of $\mathcal{L}$-innovations specified below.

Absolute summability is stronger that the square summability often required for linear processes with martingale difference innovations. The reason is that ε_t are not predictable under $\mathcal{L}$ (see Assumption 2.3 below), but this does not imply lack of serial correlation; see Remark 2.5. Therefore, absolute summability simply ensures that y_t exists irrespective of the serial correlation of ε_t.

Assumption 2.2 effectively describes the relevant dependence structure that can be used to set up forecasts, with ε_t being linearly unpredictable as specified in

Assumption 2.3 *Let the innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a nondegenerate strictly stationary and ergodic process, L_p-bounded. Further, let b exist uniquely such that*

$$\mathrm{E}\left[\varepsilon_{t-j}\mathcal{L}'\left(\varepsilon_t - b\right)\right] = 0 \quad \forall j \geq 1.$$

In the case of squared-error loss, it is common to require $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ to possess the martingale difference property. Under general loss, however, this is not the best way to model innovations; quite naturally in light of Section 2.2, we require them to behave as outlined by Corollary 2.1. This analogue of the white noise property is sufficient for consistency, as shall be seen in Proposition 2.3 below. However, in order to derive convergence rates of the $\mathcal{L}$-specific long AR coefficients, we shall strengthen the requirement to an analogue of the md property; see Assumption 2.4 below. Because of the forecast bias under nonquadratic losses, we do not specify the expectation of ε_t to be zero; clearly, $\mathrm{E}\left[y_t\right] = \sum_{j\geq 1} b_j \,\mathrm{E}\left[\varepsilon_t\right]$, which gives a further reason to require absolute summability.

Essentially, we require here ε_t to be linearly unforecastable given its past; all information relevant for (linearly) forecasting y_t under $\mathcal{L}$ is thus captured by the linear structure of Assumption 2.2. This is not uncommon in the literature. For $p = 1$, one recovers the linear infinite-order model with zero conditional-quantile innovations of Zernov et al. (2009); cf. also the earlier CAViaR model of Engle and Manganelli (2004). For $p = 2$ and $\alpha = 0.5$, one recovers the classical case to be estimated by means of least squares.

A preliminary question of interest concerns the stochastic properties of y_t from Assumptions 2.2 and 2.3. They are summarized in the following

Lemma 2.1 *Under Assumptions 2.2 and 2.3, the process y_t exists almost surely, is uniformly L_p-bounded, strictly stationary, and ergodic.*

Proof: *See the Appendix.*

Ergodicity, for instance, eliminates predictable components. Given the assumed invertibility and absolute summability of its MA(∞) representation, the process y_t also has an AR(∞) representation in terms of innovations ε_t,

$$y_t = \sum_{j\geq 1} a_j y_{t-j} + \varepsilon_t.$$

It is known from Brillinger (1975, p. 79) that the coefficients a_j are absolutely summable as well. Under Assumption 2.3, it is then straightforward to derive the optimal linear one-step ahead forecast, which is simply given by the autoregression

$$y_t\,(1) = \sum_{j\geq 1} a_j y_{t+1-j} + b.$$

Given uniqueness of the decomposition from Proposition 2.1, the above AR coefficients are the same as the coefficients from Equation (2.2). (Recall, they are $\mathcal{L}$-specific.)

The long autoregression is given by

$$y_t = \sum_{j=1}^{h_T} a_j y_{t-j} + \varepsilon_{t,h_T}, \quad t = h_T + 1, \ldots, T,$$

where $h_T \to \infty$ at a suitable rate, and the disturbances ε_{t,h_T} are easily seen to satisfy, like for the OLS long autoregression,

$$\sup_{t\in\mathbb{Z}} \|\varepsilon_{t,h_T} - \varepsilon_t\|_p \leq \sup_{t\in\mathbb{Z}} \|y_{t-j}\|_p \sum_{j\geq h_T+1} |a_j| \to 0. \tag{2.5}$$

The long autoregressive approximation leads to a truncated forecast function:

$$y_t\,(1) \approx \sum_{j=1}^{h_T} a_j y_{t+1-j} + b.$$

For a given sample, one requires estimators to plug in, together with restrictions on h_T. To obtain coefficient estimators, one minimizes the average observed (in-sample) loss:[2]

$$\left(\hat{\boldsymbol{a}}'_{h_T}, \hat{b}\right)' = \underset{\left(\boldsymbol{a}^{*\prime}_{h_T}, b^*\right)' \in \Theta}{\arg\min} \frac{1}{T} \sum_{t=p+1}^{T} \mathcal{L}\left(y_t - \sum_{j=1}^{p} a_j^* y_{t-j} - b^*\right), \tag{2.6}$$

where $\hat{\boldsymbol{a}}_{h_T} = (\hat{a}_1, \ldots, \hat{a}_{h_T})'$ and $\boldsymbol{a}^*_{h_T} = \left(a_1^*, \ldots, a_{h_T}^*\right)'$. (They actually form a triangular array but we drop the extra notation to avoid notational overflow.) As is common in nonlinear optimization, we discuss optimization over a compact set Θ, where $\Theta = \left\{\left\|\boldsymbol{a}^*_{h_T} - \boldsymbol{a}_{h_T}\right\|_1 < C \; \forall T\right\} \subset \ell_1$, the space of absolutely summable sequences (filling in zeros for $j > h_T$).

Minimizing the in-sample risk is chosen to ensure consistency of the estimators for the correct pseudo-true values. (Beyond the usual pseudo-ML interpretation of the term, we also call them pseudo-true since their true value depends on $\mathcal{L}$ in our setup.) Dumitrescu and Hansen (2016) point out that such estimation is not efficient in general, even if the true model is linear. However, efficiency is not the main concern here; rather, in their terminology, we need to ensure *robustness* of the estimators – i.e. that they deliver the right forecast functional, at least in the limit. For this minimization to deliver consistent results, we require additional assumptions.

Further, in order to derive convergence rates, we strengthen the no-linear-predictability condition on the innovations ε_t:

Assumption 2.4 *Let ε_t satisfy Assumption 2.3 with the stronger requirement*

$$\mathrm{E}\left[\mathcal{L}'\left(\varepsilon_t - b\right)|\mathcal{F}_{t-1}\right] = 0$$

replacing

$$\mathrm{E}\left[\varepsilon_{t-j}\mathcal{L}'\left(\varepsilon_t - b\right)\right] = 0 \quad \forall j \geq 1.$$

We also need to ensure that the process is not overdifferenced in a certain sense. Therefore,

Assumption 2.5 *Let smallest eigenvalue of the sample autocovariance matrix of order h_T of y_t be bounded away from zero w.p. 1.*

It may be surprising that, unlike for the characterization of the $\mathcal{L}$-innovations, the (sample) autocovariances play a role, but it should be reminded that the memory of the process is closely related to this behavior of the sample autocovariances in a linear setup. In fact, one may impose low-level conditions on the dependence of y_t such that the above assumption is fulfilled, but we find the assumption on autocovariances to be more informative.

[2] To this end, one can use the numerical method proposed by Demetrescu (2006), which is tailored for this kind of loss minimization problems.

2.3.2 Asymptotic results

We first discuss consistency.

For OLS estimation of the coefficients $\{a_j\}_{j\geq 1}$, Berk (1974) suggested that the ℓ_2 (Euclidean) vector norm of the difference between the vector of autoregressive estimates and $\boldsymbol{a}_{h_T}$, the vector containing the first h_T elements of the sequence of (pseudo-)true parameter values, should vanish as $T \to \infty$ and $h_T \to \infty$. Elementwise convergence is not sufficient for a number of applications, in particular forecasting using long autoregressions.

The following proposition gives the first asymptotic result. It establishes consistency in ℓ_1 vector norm (which, taken alone, is stronger than in ℓ_2 norm).

Proposition 2.3 *Under Assumptions 2.1 with $p > 1$, 2.2 and 2.3, it holds as $h_T, T \to \infty$ such that $h_T/T \to 0$ that*

$$\left\| \left(\hat{\boldsymbol{a}}'_{h_T}, \hat{b}\right)' - \left(\boldsymbol{a}'_{h_T}, b\right)' \right\|_1 \xrightarrow{p} 0.$$

Proof: *See the Appendix.*

Clearly, this also implies convergence in L_2 vector norm, which has been discussed before; see e.g. Berk (1974) for the OLS case. The assumptions on h_T are, however, less strict here than in the literature on OLS long autoregressions. This is because the latter results usually also allow one to establish limiting distributions and $\sqrt{T}$ consistency, which are not of interest (yet).

To obtain an (asymptotically) optimal forecast, the difference between the theoretical forecast, $y_t(1) = \sum_{j\geq 1} a_j y_{t+1-j} + b$, and its sample counterpart $y_t(1) = \sum_{j\geq 1} \hat{a}_j y_{t+1-j} + \hat{b}$ should vanish asymptotically,

$$\sum_{j=1}^{h_T} (\hat{a}_j - a_j)\, y_{t+1-j} + \sum_{j\geq h_T+1} a_j y_{t+1-j} + \left(b - \hat{b}\right) \xrightarrow{p} 0. \tag{2.7}$$

The sum $\sum_{j\geq h_T+1} a_j y_{t-j}$, and hence $\sum_{j\geq h_T+1} a_j y_{t+1-j}$, vanishes as $h_T \to \infty$, see Equation (2.6).

Unless y_t is a.s. bounded, the above consistency of the estimators is not sufficient for consistency of the forecast function, since $h_T \to \infty$. So we are left with showing that $\hat{a}_j$ converge fast enough. Let us now examine the convergence rates required for setting up a forecast. We provide a result for the L_2 vector norm as it is more convenient for later use.

Proposition 2.4 *Let $\exists r \geq 2p$ such that ε_t is uniformly L_r-bounded and $\exists s > 1/2$ such that $\sum_{j\geq 1} j^s |b_j| < \infty$. Moreover, if $p = 2$, let ε_t have absolutely continuous conditional distribution. Then, under Assumptions 2.1 – 2.5 with $p > 1$ and $h_T/T^{1/2} \to 0$ as $T \to \infty$, it holds that*

$$\left\| \left(\hat{\boldsymbol{a}}'_{h_T}, \hat{b}\right)' - \left(\boldsymbol{a}'_{h_T}, b\right)' \right\|_2 = O_p\left(\max\left\{h_T^{1/2-s}; \frac{h_T^{1/2}}{T^{1/4}}\right\}\right).$$

Proof: *See the Appendix.*

Remark 2.6 *The convergence rates depend on sample covariance matrix of y_t, and not of non-linear transformations thereof. Comparing with Proposition 2.3, we note that the requirements here are stronger, since Proposition 2.4 requires in any case $h_T = o\left(\sqrt{T}\right)$. This is because of the different proof technique, here the effect of the bias terms is stronger; see the proof for details.*

Given the convergence rates, we may then show that the linear forecast function is estimated consistently, considering further restrictions on h_T.

Corollary 2.2 *If $s > 1$ and $h_T = o(\sqrt[4]{T})$ then $y_t(1) - y_t(1) \xrightarrow{p} 0$.*

Proof: *Obvious and omitted.*

The following section provides an evaluation of the finite-sample predictive performance of the semiparametric long autoregressive approach under asymmetric loss functions.

2.4 Finite sample evidence

2.4.1 Forecast methods

First, we fit a long autoregression of increasing order h_T under asymmetric loss. The plug-in estimates are obtained according to (2.6). The choice of the model order h_T of the long autoregression influences the quality of the forecasts. To select an autoregressive model order, we work with information criteria [IC] in the spirit of Weiss (1996). Since we must choose a lag order under a given loss function, we use the modified information criterion proposed by Demetrescu and Hoke (2019) for the family of asymmetric power loss functions, given by

$$IC_{\mathcal{L}}(k) = \frac{2}{p} \log\left(\sum \mathcal{L}\left(\hat{\varepsilon}_{t,k}\right)\right) + \frac{2k}{T},$$

where $\hat{\varepsilon}_{t,k}$ are the residuals fitted for an autoregressive model of order k estimated under $\mathcal{L}$. This version of the loss information criterion is based on the Akaike criterion. We hold it for obvious that choosing h_T by minimizing $IC_{\mathcal{L}}(k)$ over $k \in \{1, 2, \ldots, h^{\max}\}$ with $h^{\max} \to \infty$ ensures that $h_T \to \infty$ if the true model order (under the relevant $\mathcal{L}$) is not finite. We simulate with $h^{\max} = \lfloor 4(T/100)^{0.25} \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function.

We compare the long autoregression against two alternatives. The first is the two-step procedure proposed by Granger (1969), while the second one is based on a standard AR-GARCH model with QML estimation. The two-step procedure described by Granger (1969) consists of first fitting an AR(h_T) process with intercept by OLS. Here we choose h_T using the standard AIC. One thus obtains estimators $\hat{\varphi}_j$ for the autoregressive parameters, $\hat{c}$ for the (OLS-specific) intercept, followed by computation of the OLS residuals,

$$\hat{e}_{t,h_T} = y_t - \sum_{j=1}^{h_T} \hat{\varphi}_j y_{t-j} - \hat{c}, \quad t = h_T + 1, \ldots, T. \tag{2.8}$$

Since the OLS residuals $\hat{e}_{t,h_T}$ are demeaned by construction, one only needs to estimate the so-called bias factor in the second step (see Granger, 1969),

$$\hat{b} = \underset{b^* \in \mathbb{R}}{\arg\min} \frac{1}{T} \sum_{t=h_T+1}^{T} \mathcal{L}\left(\hat{e}_{t,h_T} - b^*\right), \tag{2.9}$$

and the two-step forecast for y_{t+1} is given under the loss function $\mathcal{L}$ by

$$\tilde{y}_t\left(1\right) = \hat{m} + \hat{b} + \sum_{j=1}^{h_T} \hat{a}_j y_{t+1-j}.$$

In the presence of conditional heteroskedasticity, this forecast is suboptimal in the terminology of Christoffersen and Diebold (1997) as it averages the volatility dynamics; also, the first-step least-squares estimation is inefficient.

The second alternative approach is more parametric in nature and consists of fitting an AR-GARCH model to capture the dynamics in both the conditional mean and the conditional variance. Here we estimate an AR(h_T)-GARCH(1,1) model by means of quasi-maximum-likelihood assuming conditionally Gaussian innovations. We use the "classical" AIC for selecting the lag order of the autoregressive part of the AR-GARCH model, while the GARCH(1,1) component is fixed. Based on this approximate model, we forecast the conditional mean and the conditional variance, which are then used to scale and shift the estimated standardized innovations to obtain an estimate of the forecast distribution. Based on this forecast distribution, we minimize in the last step the forecast loss to obtain the needed optimal forecast. An early paper proposing such an approach for forecasting under general loss functions is McCullough (2000);[3] see also Dumitrescu and Hansen (2016). Such model-based approaches have the advantage that it applies too in situations where $\mathcal{L}$ is not homogenous, or even not in difference form. Moreover, parameter estimation is more efficient when accounting for conditional heteroskedasticity whenever this is present. However, if no conditional heteroskedasticity is present, then an AR-GARCH based forecast may underperform, as will be seen in the following.

2.4.2 Data generating processes

We generate a variety of ARMA(1,1)-GARCH(1,1) series with different degrees of serial correlation and heteroskedasticity for several sample sizes. Moreover, we allow the standardized shocks η_t to exhibit nonzero skewness and excess kurtosis. The first data generating process is as follows:

$$y_t = \phi\, y_{t-1} + \theta\, e_{t-1} + e_t \tag{2.10}$$

$$e_t = \sigma_t\, \eta_t \tag{2.11}$$

$$\sigma_t^2 = \omega + \gamma\, e_{t-1}^2 + \beta\, \sigma_{t-1}^2, \tag{2.12}$$

[3] McCullough (2000) employs a bootstrap scheme, which allows him to take the influence of estimation risk on the optimal point forecast into account.

where $\eta_t \sim \text{iid}(0,1)$. The simulated data can be split into two main groups, namely

a) strong serial correlation with weak conditional heteroskedasticity ($\phi = 0.9$, $\theta = 0.6$; $\omega = 1$, $\gamma = 0.1$, $\beta = 0.1$);

b) weak serial correlation with strong conditional heteroskedasticity ($\phi = 0.1$, $\theta = 0.5$; $\omega = 1$, $\gamma = 0.3$, $\beta = 0.6$).

This design delivers efficient estimation, since the AR-GARCH model is correctly specified (taking for granted that the AR part reasonably approximates the ARMA DGP). To allow for some misspecification, we also consider the following DGP building on Example 2.1

$$y_t = \theta_1 \, y_{t-1} + e_t \tag{2.13}$$

$$e_t = (\theta_0 + \theta_2 \, y_{t-1}) \, u_t \tag{2.14}$$

where $u_t \sim \text{iid}(0,1)$. As before, we control for the intensity of the mean and variance dynamics by changing the respective parameters:

a) strong serial correlation with weak conditional heteroskedasticity ($\theta_0 = 1$, $\theta_1 = 0.7$, $\theta_2 = 0.2$);

b) weak serial correlation with strong conditional heteroskedasticity ($\theta_0 = 1$, $\theta_1 = 0.2$, $\theta_2 = 0.4$).

The innovations in both scenarios were generated to follow skewed t distributions as in Fernández and Steel (1998) with shape parameters $\nu \in \{5, 50\}$, the number of degrees of freedom, and $\xi \in \{0.5, 2\}$, representing left and right skewness, respectively. Each parameterization was repeated with increasing sample size $T \in \{100, 150, \ldots, 450, 500\}$, so that convergent behavior would become more evident.[4] The shapes of the respective loss functions were controlled by the degree of asymmetry $\alpha \in \{0.2, 0.8\}$[5] and the tail parameter $p \in \{2, 3\}$. The number of Monte Carlo replications was set to $MC = 25,000$. All simulations were conducted in R (R-Core-Team, 2014; Ghalanos, 2019).

We report p-roots of average forecast losses normalized to the losses of the theoretical one-step ahead forecast

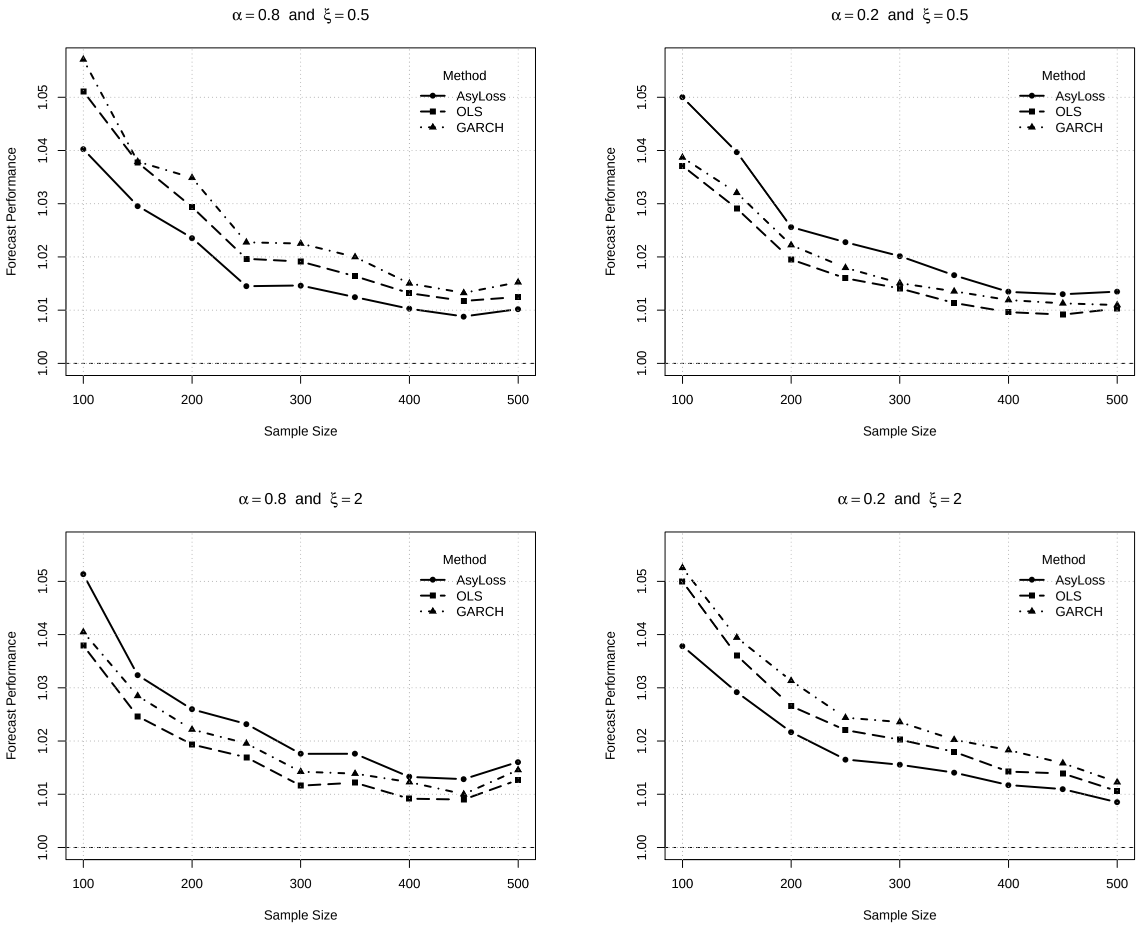
$$\sqrt[p]{\frac{\frac{1}{MC}\sum_{i=1}^{MC} \mathcal{L}\Big(y_{T+1} - \hat{y}_T(1)\Big)}{\frac{1}{MC}\sum_{i=1}^{MC} \mathcal{L}\Big(y_{T+1} - \bar{y}_T(1)\Big)}}$$

with $\bar{y}_T(1)$ defined in (2.4) and $\bar{b}$ obtained numerically. The ratio takes values larger than unity, since the true model parameters are used to compute the theoretical optimal forecast $\bar{y}_T(1)$, which is therefore not plagued by any estimation risk. Hence, the smallest figures give the best relative forecasting performance.

[4] We also performed estimation on sample sizes up to 1000. The results did not change significantly after $T = 500$ and we do not report them to save space.

[5] The results for estimation under symmetric loss were also left out, since the setup parallels minimizing the MSE.

Figure 2.1: Relative forecast performance for highly serially correlated ARMA-GARCH processes



***Notes*:** Loss function with tail parameter $p = 2$ and asymmetry parameter $\alpha \in \{0.2, 0.8\}$, skewed Student $t(50)$ innovations with left and right skewness $\xi \in \{0.5, 2\}$

2.4.3 Results

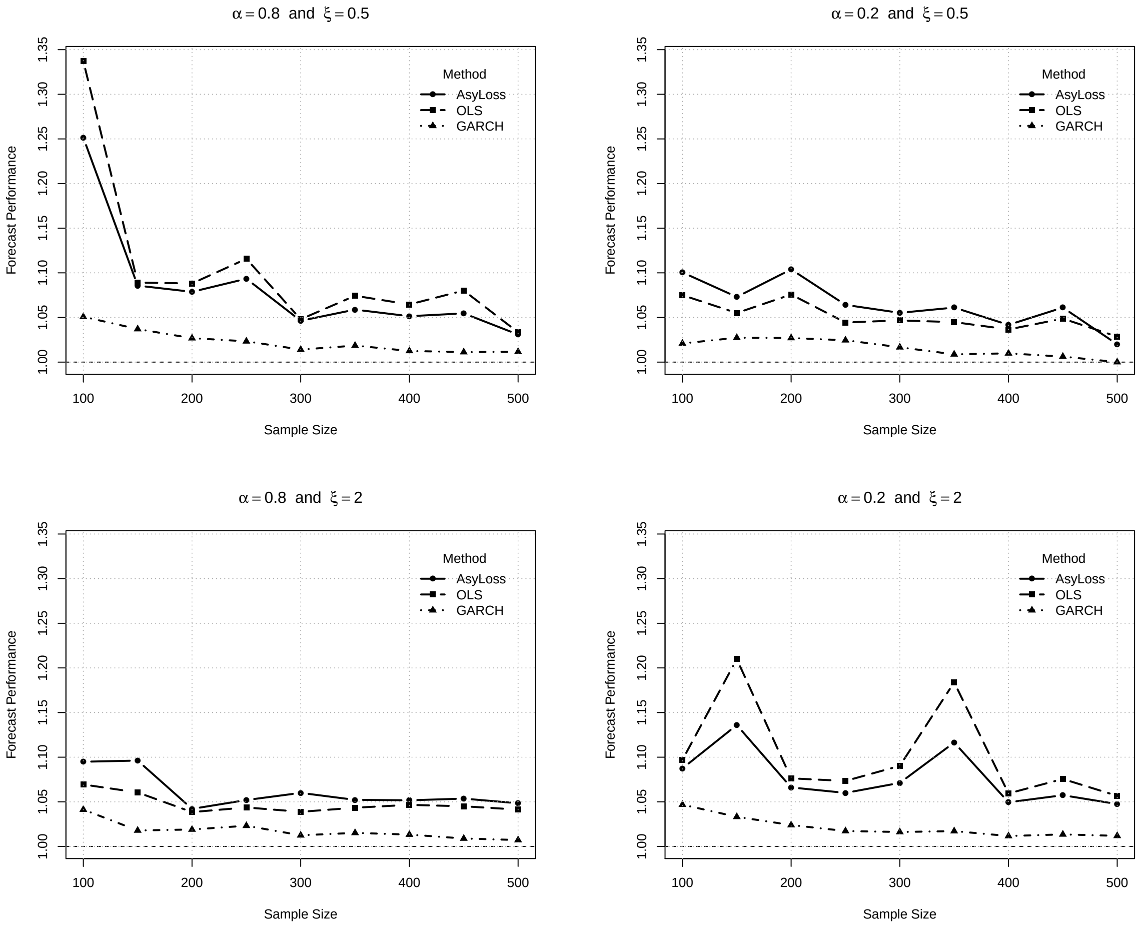
First, we discuss the results for ARMA-GARCH DGPs as defined by Equations (2.10)–(2.12) with linear dynamics as dominating feature.

Figure 2.1 displays the relative performances for different combinations of the loss function asymmetry and skewness of the innovations in presence of high serial correlation. As T grows, the difference to the theoretical optimal predictor decreases, illustrating the convergence behavior of the involved estimators. The upper left set of curves represents a case when a forecaster puts higher weight on positive forecast errors while dealing with a process driven by left-skewed innovations. Here fitting a long autoregression under asymmetric loss yields the best results compared to the alternatives. The same can be observed in a reversed situation (lower right plot of Figure 2.1). It is interesting that the long autoregression under the relevant loss delivers the best results whenever the asymmetry of the loss function “compensates” for the skewness of the innovations (of course, in these cases, negative of the log-likelihood is closest to the observed loss and estimators are efficient). In the cases where the negative quasi log-likelihood is at odds with the loss function used in estimation (upper right and lower left panels), estimation under the relevant loss is inefficient, and the two-step OLS-based procedure delivers best forecasts.

The cases in Figure 2.1 have linearity as the main data generation feature, since the conditional heteroskedasticity is being held rather tame. This explains why fitting a full AR-GARCH model and building forecasts based on this model is dominated by either estimation under the relevant loss or by Granger's two-step procedure. Among the latter two there is no clear winner, as the ranking depends on the estimation efficiency as pinned down here by the match or mismatch in innovations skewness and loss function asymmetry.

With strong well-specified GARCH effects, the picture changes in favor of a modelling approach.

Figure 2.2: Relative forecast performance for weakly serially correlated ARMA-GARCH processes

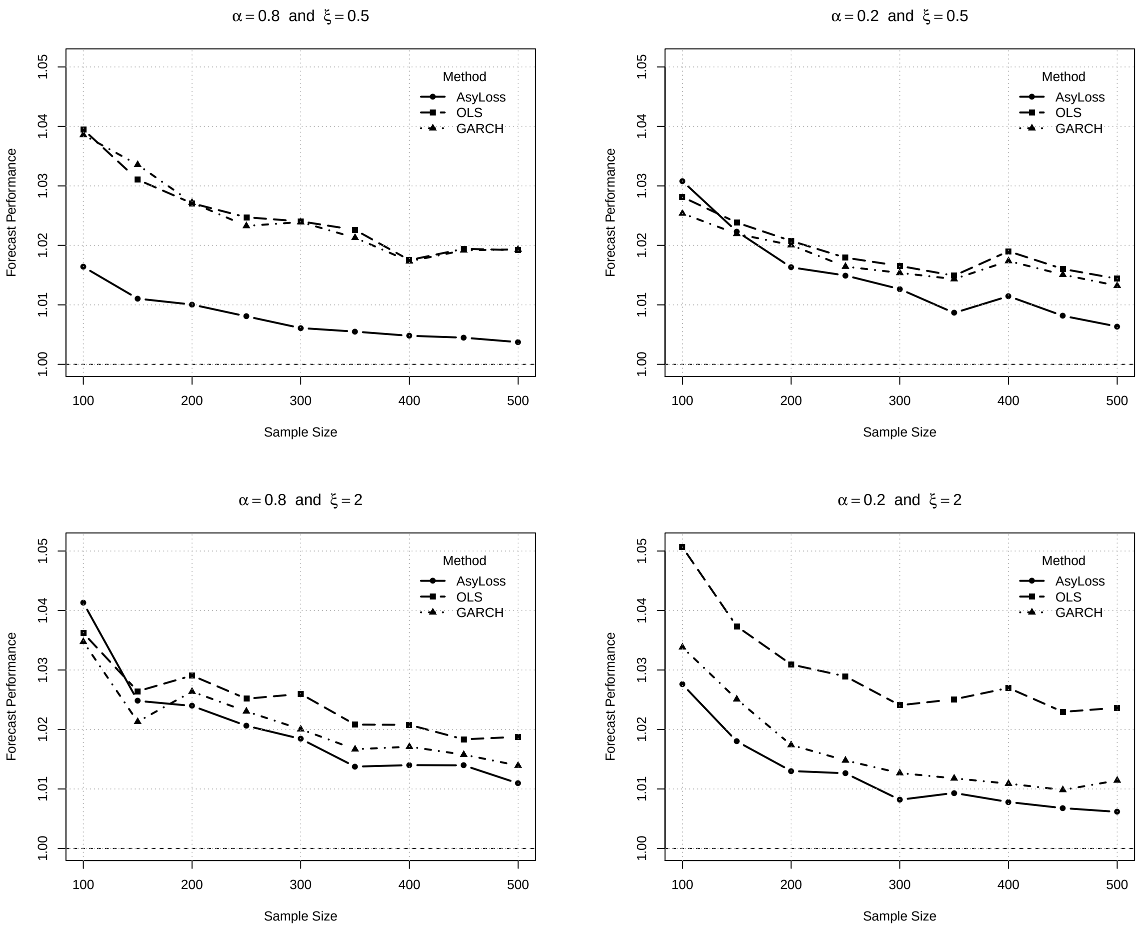


Notes: See Figure 2.1 for details.

Figure 2.2 depicts the simulation results for weakly serially correlated processes with high degree of volatility clustering. Here, using an AR-GARCH model for forecasting delivers the better forecasts. Fitting the AR-GARCH model is thus more beneficial in cases, when conditional heteroskedasticity is the main feature of the data. The differences between the long autoregression under the relevant loss and Granger's two-step procedures are minimal, with some advantage of the former when estimation under the relevant loss is efficient.

It may seem that one should in principle use a model-based forecast, at least whenever there is strong volatility clustering. Part of the good performance of the AR-GARCH forecast is however due to the fact that the volatility model is the correct one. Figures 2.3 and 2.3 presenting the results for the second DGP defined by equations (2.13) and (2.14) show that knowing the true volatility model indeed gives a boost in forecasting performance.

Figure 2.3: Relative forecast performance for strongly serially correlated bilinear processes

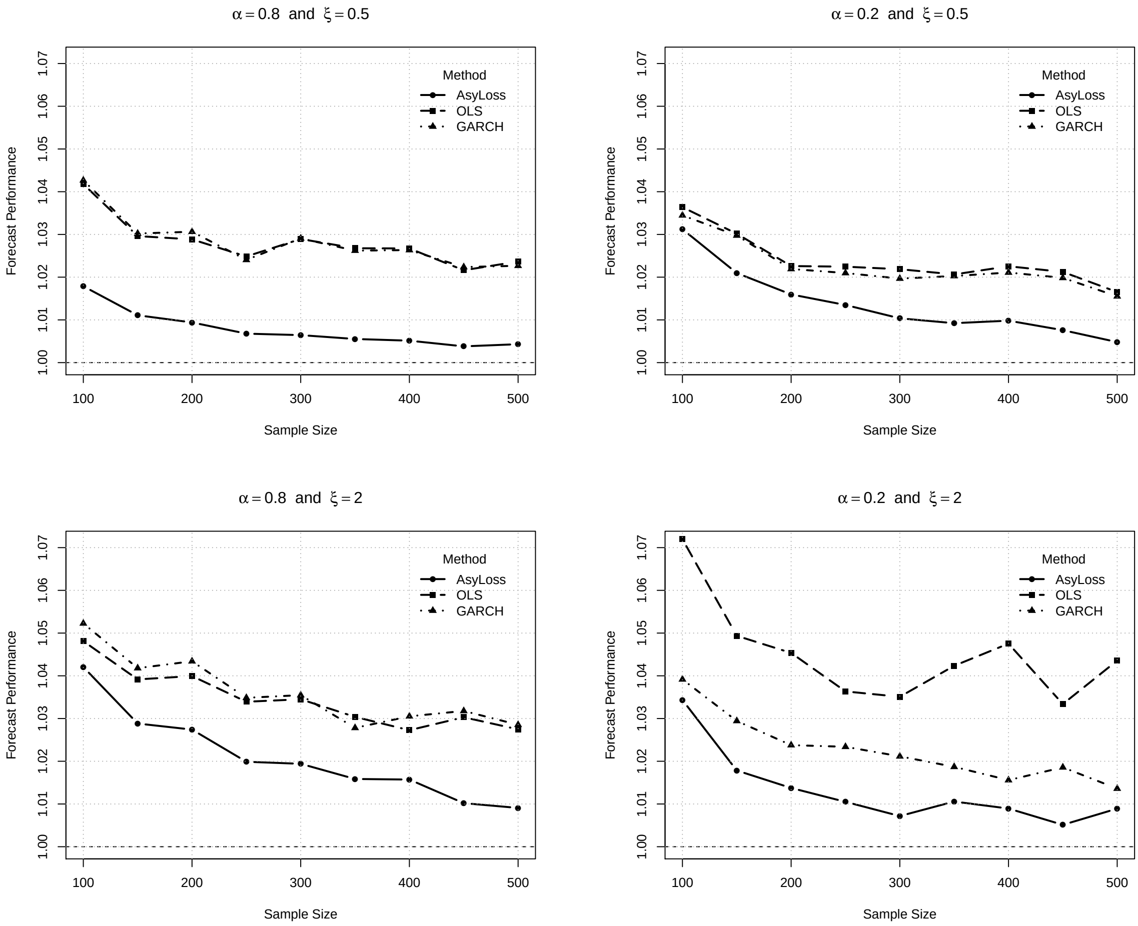


Notes: See Figure 2.1 for details.

As shown in Figure 2.3, the long autoregression under the relevant loss dominates for all of the skewness/asymmetry combinations, now that the AR-GARCH based forecast only approximates the true volatility dynamics. The improvement in the relative performance of one-step forecasting is rather impressive, but not surprising in the light of Example 2.1. The two-step OLS-based approach is not competitive at all. Figure 2.4 confirms this conclusion.

Further results for different loss functions ($p = 3$) and kurtosis ($\nu = 5$) can be found in the Appendix; they largely confirm the above findings. AR-GARCH models do well in certain high-GARCH cases, but lose edge when volatility model not well specified; long autoregressions under

Figure 2.4: Relative forecast performance for weakly serially correlated bilinear processes



Notes: See Figure 2.1 for details.

asymmetric loss appear to be somewhat more robust; moreover, when they dominate, they do so by a larger margin that in the cases where estimation is inefficient. All in all, we may recommend the use of long autoregressions under asymmetric loss, except for the cases where GARCH effects are strong.

2.5 Summary

Long autoregressions have proved to be indispensable in theoretical and applied time series analysis. This paper provides arguments in favor of using long autoregressions for forecasts in conjunction with generic loss functions.

Concretely, we characterized a loss-specific autoregressive representation of strictly stationary processes which provides a theoretical justification for long autoregressions under the relevant loss. Consistency and convergence rates of the relevant coefficient estimators is established under mild regularity conditions.

A finite-sample evaluation of the forecasting performance of long autoregressions under the relevant loss functions shows that they deliver reliable forecasts for a variety of data generating processes. One exception is represented by strong GARCH effects, where AR-GARCH models have the potential to provide better forecasts.

Appendix

The following Lemma is required for the proofs of Proposition 2.3.

Lemma 2.2 *Let ε_t and u_{t-1} be two possibly dependent, non-degenerate, uniformly L_p-bounded random variables and $\mathcal{L}(\cdot)$ a loss function obeying Assumption 2.1. Let b be the optimal predictor (with respect to $\mathcal{L}$) for ε_t conditional on u_{t-1}. It then holds*

$$\mathrm{E}\left[\mathcal{L}\left(\varepsilon_t + u_{t-1} - b^*\right)\right] > \mathrm{E}\left[\mathcal{L}\left(\varepsilon_t - b\right)\right] \quad \forall b^* \in \mathbb{R}.$$

Proof of Lemma 2.1

By using uniform L_p-boundedness of ε_t and Minkowski's inequality, it follows that y_t itself is uniformly L_p-bounded, and thus L_1-bounded. The a.s. existence follows e.g. from Torres (1986), and, given the existence, strict stationarity and ergodicity follow; see e.g. White (2001, Theorem 3.35).

Proof of Lemma 2.2

By assumption we have

$$\mathrm{E}\left[\mathcal{L}\left(\varepsilon_t - b^* + u_{t-1}\right)|u_{t-1}\right] > \mathrm{E}\left[\mathcal{L}\left(\varepsilon_t - b\right)|u_{t-1}\right], \ \forall b^* \neq b + u_{t-1}.$$

The result follows with the Law of Iterated Expectations since u_{t-1} is nondegenerate and the probability that $b^* = b + u_{t-1}$ is strictly smaller than one.

Proof of Proposition 2.1

Since $\mathcal{L}$ is nonnegative with $\mathcal{L}(u) \neq 0 \ \forall u \neq 0$, $\mathcal{Q}$ is nonnegative and a minimum exists. For characterizing the minimum, it suffices to focus on sequences of parameters a_j for which the linear combinations $y_{t+1} - \sum_{j\geq 1} a_j^* y_{t+1-j} - b$ have finite expected loss.

Examine $y_{t+1} - \sum_{j\geq 1} a_j^* y_{t+1-j}$; should this have a degenerate distribution for some values of $a_j^* = a_j$, one may obviously choose $b = \mathrm{E}\left[y_{t+1} - \sum_{j\geq 1} a_j y_{t+1-j}\right]$ such that the forecast loss is zero w.p. 1 and the minimum is found. The f.o.c. are obviously fulfilled for $p \geq 1$ since $y_{t+1} - \sum_{j\geq 1} a_j y_{t+1-j} - b = 0$ and $\mathcal{L}'(0) = 0$ w.p. 1. This is the case of a purely predictable process y_t.

Let us then examine the nondegenerate case; we discuss $p > 1$ first, where the function

$$\mathcal{Q}(a_j^*, b^*) = \mathrm{E}\left[\mathcal{L}\left(y_{t+1} - \sum_{j\geq 1} a_j^* y_{t+1-j} - b^*\right)\right]$$

is differentiable. Also, we show $\mathcal{Q}$ to be a strictly convex function in the parameters (a_j, b), such that $\mathcal{Q}$ is coercive, implying that the infimum is attained and a global minimum of $\mathcal{Q}$ exists in

ℓ_1, the space of absolutely summable real sequences. To establish the desired convexity, define $\boldsymbol{y}_{t+1}^{*} = (y_{t+1}, -y_t, -y_{t-1}, ..., -1)$ and $\boldsymbol{u} = (1, a_1, a_2, \ldots, b)$, and consider the function

$$Q_t(\boldsymbol{u}) = \mathcal{L}\left(\boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u}\right),$$

which we show to be strictly convex for $p > 1$ as follows: for any $\lambda \in (0,1)$ and $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^{\mathbb{N}}$, we have

$$Q_t(\lambda \boldsymbol{u}_1 + (1-\lambda)\boldsymbol{u}_2) = \mathcal{L}\left(\boldsymbol{y}_{t+1}^{*}(\lambda \boldsymbol{u}_1 + (1-\lambda)\boldsymbol{u}_2)\right) = \mathcal{L}\left(\lambda \boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u}_1 + (1-\lambda)\boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u}_2\right);$$

since $\mathcal{L}$ is strictly convex for $p > 1$[6], it holds for any real $u_1 \neq u_2$ that

$$\mathcal{L}(\lambda u_1 + (1-\lambda)u_2) < \lambda \mathcal{L}(u_1) + (1-\lambda)\mathcal{L}(u_2),$$

so let $u_{1,2} = \boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u}_{1,2}$ to obtain

$$\mathcal{L}\left(\lambda \boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u}_1 + (1-\lambda)\boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u}_2\right) < \lambda \mathcal{L}\left(\boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u}_1\right) + (1-\lambda)\mathcal{L}\left(\boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u}_2\right)$$

leading to

$$Q_t(\lambda \boldsymbol{u}_1 + (1-\lambda)\boldsymbol{u}_2) < \lambda Q_t(\boldsymbol{u}_1) + (1-\lambda)Q_t(\boldsymbol{u}_2),$$

i.e. strict convexity of Q_t. Strict convexity of $\mathcal{Q}$ is established by taking expectations and recalling that we discuss the nondegenerate case, $\boldsymbol{y}_{t+1}^{\prime *}\boldsymbol{u} \neq 0$ w.p. 1. To complete the case $p > 1$, recall that, if a strictly convex function has a global minimum, the minimum is unique. Moreover, the optimum has to be a stationary point due to differentiability of $\mathcal{Q}$, which is the case here since $\mathcal{L}'$ is continuous for $p > 1$. The f.o.c. are time-invariant due to strict stationarity of y_t.

The case $p = 1$ of an asymmetric linear loss function is not essentially different. Convexity is established analogously to the case $p > 1$, such that a minimum exists. To establish the uniqueness, assume that the minimum of $\mathcal{Q}(a_j^*, b^*)$ is not unique. Note however that the set of optimum points must be a star domain, otherwise $\mathcal{Q}(a_j^*, b^*)$ would not be convex. We may hence examine the effect on $\mathcal{Q}\left(a_j^*, b^*\right)$ of arbitrarily small deviations from some optimum a_j, b within this domain, and show that continuity of the distributions suffices for an increase in Q, hence nonuniqueness is contradicted. We have that

$$\begin{aligned}\mathcal{Q}(a_j + \xi_j, b + \xi) &= \mathrm{E}\left[\mathcal{L}\left(y_{t+1} - \sum_{j\geq 1} a_j y_{t+1-j} - b - \left(\sum_{j\geq 1}\xi_j y_{t+1-j} + \xi\right)\right)\right] \\ &= \mathrm{E}\left[\mathcal{L}(v_{t+1} - \psi_t)\right]\end{aligned}$$

with $v_{t+1} = y_{t+1} - \sum_{j\geq 1} a_j y_{t+1-j} - b$ (which is not degenerate) and $\psi_t = \sum_{j\geq 1}\xi_j y_{t+1-j} + \xi$. Note that $\mathrm{E}\left[\mathcal{L}(v_{t+1} - c)\right]$ is minimized at $c = 0$ by the construction of v_{t+1} since a_j, b

[6]Note that the proof for $p > 1$ holds for smooth, strictly convex loss functions in general and not just asymmetric power loss; a necessary condition is however finiteness of the expected loss.

characterize a minimum. Then, if all finite-dimensional distributions of y_t are continuous, so are the distributions of v_{t+1} and ψ_t (with the exception of the trivial case $\xi_j = 0$). With $\mathcal{L}$ the asymmetric linear loss function, $\mathrm{E}\left[\mathcal{L}\left(v_t - c\right)\right]$ is minimized uniquely at some quantile of the distribution of v_{t+1}, which, given continuity of the distribution of v_{t+1}, must be unique. Since ψ_t is not a degenerate random variable, Lemma 2.2 implies that $\mathrm{E}\left[\mathcal{L}\left(v_{t+1} - \psi_t\right)\right] > \mathrm{E}\left[\mathcal{L}\left(v_{t+1}\right)\right]$ for all a_j^*, b^* different from a_j, b, as required for the uniqueness.

Finally, given the continuity of the finite-dimensional distributions of y_t, the discontinuity of $\mathcal{L}'$ occurs on a set of measure zero and the same characterization of the stationary point emerges for $p = 1$ as for $p > 1$.

Proof of Corollary 2.1

Recall that the generalized forecast error at time $t+1$ is uncorrelated with y at all times $t+1-i$ for $i \geq 1$. Then, having thus for all i

$$\varepsilon_{t+1-i} = y_{t+1-i} - \sum_{j \geq 1} a_j y_{t+1-i-j} - b,$$

we have that

$$\mathcal{L}'\left(\varepsilon_{t+1}\right) \varepsilon_{t+1-i} = \mathcal{L}'\left(\varepsilon_{t+1}\right) y_{t+1-i} - \sum_{j \geq 1} a_j \mathcal{L}'\left(\varepsilon_{t+1}\right) y_{t+1-i-j} - b\mathcal{L}'\left(\varepsilon_{t+1}\right);$$

the result follows upon taking expectations.

Proof of Proposition 2.2

Begin by noting that $b_0 = 1$:

$$b_0 = \frac{\mathrm{E}\left[y_t \mathcal{L}'\left(\varepsilon_t\right)\right]}{\mathrm{E}\left[\varepsilon_t \mathcal{L}'\left(\varepsilon_t\right)\right]} = \frac{\mathrm{E}\left[\left(y_{t-1}\left(1\right) + \varepsilon_t\right) \mathcal{L}'\left(\varepsilon_t\right)\right]}{\mathrm{E}\left[\varepsilon_t \mathcal{L}'\left(\varepsilon_t\right)\right]}$$

where $y_{t-1}\left(1\right)$ is the optimal linear forecast of y_t given its infinite past and as such uncorrelated with $\mathcal{L}'\left(\varepsilon_t\right)$.

1. Write

$$\mathrm{E}\left[m_{t+h} \mathcal{L}'(\varepsilon_t)\right] = \mathrm{E}\left[\left(y_{t+h} - \sum_{j \geq 0} b_j \varepsilon_{t+h-j}\right) \mathcal{L}'(\varepsilon_t)\right],$$

and the result is immediate for $h < 0$ given Proposition 2.2 and Corollary 2.1. For $h \geq 0$, rewrite b_j as

$$b_j = \frac{\mathrm{E}\left[y_{t+j} \mathcal{L}'\left(\varepsilon_t\right)\right] - \sum_{k=0}^{j-1} b_k \mathrm{E}\left[\varepsilon_{t+j-k} \mathcal{L}'\left(\varepsilon_t\right)\right]}{\mathrm{E}\left[\varepsilon_t \mathcal{L}'\left(\varepsilon_t\right)\right]}$$

exploiting strict stationarity of y_t. Then,

$$\begin{aligned}
\mathrm{E}\left[m_{t+h}\mathcal{L}'(\varepsilon_t)\right] &= \mathrm{E}\left[y_{t+h}\mathcal{L}'(\varepsilon_t)\right] - \mathrm{E}\left[\mathcal{L}'(\varepsilon_t)\sum_{j\geq 0} b_j\varepsilon_{t+h-j}\right] = \\
&= b_0\,\mathrm{E}\left[\varepsilon_{t+h}\mathcal{L}'(\varepsilon_t)\right] + \ldots + b_h\,\mathrm{E}\left[\varepsilon_t\mathcal{L}'(\varepsilon_t)\right] - \mathrm{E}\left[\mathcal{L}'(\varepsilon_t)\sum_{j\geq 0} b_j\varepsilon_{t+h-j}\right] \\
&= -\,\mathrm{E}\left[\mathcal{L}'(\varepsilon_t)\sum_{j\geq h+1} b_j\varepsilon_{t+h-j}\right]
\end{aligned}$$

which is zero thanks to Corollary 2.1.

2. Since $m_t \in \mathcal{S}_t$ and m_t and $\mathcal{L}'(\varepsilon_t)$ are uncorrelated, $m_t \in \mathcal{S}_{t-1}$. Apply this inductively to conclude that $m_t \in \mathcal{S}_{-\infty}$.

3. Recall that $e_t = \varepsilon_t + \sum_{j\geq 1} b_j\varepsilon_{t-j}$. Since $\mathcal{L}'(\varepsilon_t)$ is orthogonal to $\mathcal{S}_{t-1}$ and $e_{t-j} \in \mathcal{S}_{t-1}$ $\forall j \geq 1$, ε_t must be the generalized innovation of e_t, while $\sum_{j\geq 1} b_j\varepsilon_{t-j}$ is its linear predictor, based on e_t's past. Hence, e_t is a regular process.

4. Since $m_t \in \mathcal{S}_{t-1}$, $\mathcal{S}_{t-1}$ contains all the information about m_t. Hence, the linear forecast of m_t given $\mathcal{S}_{t-1}$ can only be m_t itself. So m_t is predictable.

Proof of Proposition 2.3

In the OLS framework, where closed-form expressions for the estimators exist, the ℓ_2 vector norm (and the corresponding induced matrix norm) is the natural choice. For the general case of estimation under the relevant loss function, however, we prefer the use of the ℓ_1 norm (or city-block norm),

$$\|\boldsymbol{x}\|_1 = \sum_{j=1}^{m} |x_j| \quad \forall \boldsymbol{x} = (x_1, \ldots, x_m)' \in \mathbb{R}^m,$$

simplifying the arguments. Note that convergence in the ℓ_1-sense implies convergence in the ℓ_2-sense; the converse, however, does not always hold true.

Let with $\boldsymbol{a}^* = \left(a_1^*, a_2^*, \ldots, a_p^*, 0, \ldots\right)' \in \Theta$,

$$\begin{aligned}
\mathcal{Q}_T\left(\boldsymbol{a}^{*\prime}, b^*\right) &= \frac{1}{T}\sum_{t=h_T+1}^{T} \mathcal{L}\left(y_t - b^* - \sum_{j=1}^{h_T} a_j^* y_{t-j}\right) \\
&= \frac{1}{T}\sum_{t=h_T+1}^{T} \mathcal{L}\left(\varepsilon_t - (b^* - b) - \sum_{j=1}^{\infty}\left(a_j^* - a\right) y_{t-j}\right)
\end{aligned}$$

and assume that the result of the numerical optimization exists w.p. 1. Since, for any sample size, this procedure only delivers a vector of dimension h_T+1, while, in the limit, infinitely many elements are required, we set the 'missing estimates' equal to zero. Further, let $\boldsymbol{a} = (a_1, a_2, \ldots)' \in$

ℓ_1, the space of absolutely summable sequences. Since $\sum_{j>h_T}|a_j| \to 0$, convergence of $\hat{\boldsymbol{a}} = (\hat{a}_1, \ldots, \hat{a}_{h_T}, 0, \ldots)'$ to $\boldsymbol{a}$ in L_1 norm implies the convergence posited in Proposition 2.3.

We show in a first step that, for any $\boldsymbol{a}^*$,

$$\sup_{\Theta}\left|\mathcal{Q}_T\left(\boldsymbol{a}^{*\prime}, b^*\right) - \frac{1}{T}\sum_{t=1}^{T}\mathcal{L}\left(\varepsilon_t + u^*_{t-1}\right)\right| \xrightarrow{p} 0 \tag{2.15}$$

where $u^*_{t-1} = -\sum_{j\geq 1}\left(a^*_j - a_j\right)y_{t-j} - (b^* - b)$ (with $\boldsymbol{a}^* - \boldsymbol{a}$ obviously being absolutely summable). Using Lemma 2.1, one can in fact also conclude that u^*_{t-1} is uniformly L_p-bounded, stationary and ergodic. Note also that the L_p boundedness of u^*_{t-1} is uniform in $\boldsymbol{a}^*$ as well (and not only in t), since $\|\boldsymbol{a}^*\|_1 < M$. Let $R_{t,h_T} = \sum_{j=h_T+1}^{\infty} a^*_j y_{t-j}$ such that

$$\mathcal{Q}_T = \frac{1}{T}\sum_{t=h_T+1}^{T}\mathcal{L}\left(\varepsilon_t + u^*_{t-1} + R_{t,h_T}\right).$$

For $p = 1$, $\mathcal{L}$ is Lipschitz and we have immediately that

$$\left|\mathcal{Q}_T\left(\boldsymbol{a}^{*\prime}_{h_T}, b^*\right) - \frac{1}{T}\sum_{t=1+1}^{T}\mathcal{L}\left(\varepsilon_t + u^*_{t-1}\right)\right| \leq C\frac{1}{T}\sum_{t=h_T+1}^{T}|R_{t,h_T}| + C\frac{1}{T}\sum_{t=1}^{h_T}\mathcal{L}\left(\varepsilon_t + u^*_{t-1}\right).$$

Now, y_t is uniformly L_p-bounded, so

$$0 \leq \mathrm{E}\left[\frac{1}{T}\sum_{t=h_T+1}^{T}|R_{t,h_T}|\right] \leq \mathrm{E}\left(|y_t|\right)\sum_{j\geq h_T+1}\left|a^*_j\right| \to 0 \quad \text{for all } \boldsymbol{a}^*$$

and the first summand vanishes in L_1 norm at a uniform rate, since $\sum_{j\geq 1}\left|a^*_j\right| < C \ \forall \boldsymbol{a}^* \in \Theta$. For the second summand, we note that, for $p = 1$, $\mathcal{L}\left(\varepsilon_t + u_{t-1}\right) \leq C\left(|\varepsilon_t| + \left|u^*_{t-1}\right|\right)$ with ε_t and u_{t-1} uniformly L_1-bounded, and hence

$$0 \leq \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{h_T}\mathcal{L}\left(\varepsilon_t + u^*_{t-1}\right)\right] \leq \frac{Ch_T}{T}$$

for all $\boldsymbol{a}^*$ with $\|\boldsymbol{a}^*\|_1 < C$ as required for (2.15).

For $p > 1$, we exploit the power shape of the loss function. Concretely, use the mean value theorem to conclude that

$$\mathcal{L}\left(\varepsilon_t + u^*_{t-1} + R_{t,h_T}\right) = \mathcal{L}\left(\varepsilon_t + u^*_{t-1}\right) + \mathcal{L}'\left(\varepsilon_t + u^*_{t-1} + \upsilon_t\right)R_{t,h_T}$$

with $|\upsilon_t| \leq |R_{t,h_T}|$, where, thanks to Hölder's inequality,

$$\mathrm{E}\left[\left|\frac{1}{T}\sum\mathcal{L}'\left(\varepsilon_t + u^*_{t-1} + \upsilon_t\right)R_{t,h_T}\right|\right] \leq \frac{1}{T}\sum_{t=h_T+1}^{T}\sqrt[\frac{p}{p-1}]{\mathrm{E}\left[\left|\mathcal{L}'\left(\varepsilon_t + u^*_{t-1} + \upsilon_t\right)\right|^{\frac{p}{p-1}}\right]}\sqrt[p]{\mathrm{E}\left[|R_{t,h_T}|^p\right]}.$$

Note that, thanks to the Minkowski's inequality, $\sqrt[p]{\mathrm{E}\left[\left|R_{t,h_T}\right|^p\right]} \leq \sum_{j=h_T+1}^{\infty} \left|a_j^*\right| \sqrt[p]{\mathrm{E}\left[\left|y_t\right|^p\right]} \to 0$ like in the case $p = 1$, so we only need to show that $\mathrm{E}\left[\left|\mathcal{L}'\left(\varepsilon_t + u_{t-1}^* + \upsilon_t\right)\right|^{\frac{p}{p-1}}\right] < C$ for all t. To this end, note that $\mathcal{L}'$ has power tails with index $p-1$, so

$$\mathrm{E}\left[\left|\mathcal{L}'\left(\varepsilon_t + u_{t-1}^* + \upsilon_t\right)\right|^{\frac{p}{p-1}}\right] \leq C\,\mathrm{E}\left[\left|\varepsilon_t + u_{t-1}^* + \upsilon_t\right|^p\right].$$

Furthermore, Minkowski's inequality and the fact that $|\upsilon_t| \leq |R_{t,h_T}|$ imply

$$\begin{aligned}
\sqrt[p]{\mathrm{E}\left(\left|\varepsilon_t + u_{t-1}^* + \upsilon_t\right|^p\right)} &= \sqrt[p]{\mathrm{E}\left[\left|\varepsilon_t + u_{t-1}^* + \upsilon_t\right|^p\right]} \\
&\leq \sqrt[p]{\mathrm{E}\left[\left|\varepsilon_t\right|^p\right]} + \sqrt[p]{\mathrm{E}\left[\left|u_{t-1}^*\right|^p\right]} + \sqrt[p]{\mathrm{E}\left[\left|\upsilon_t\right|^p\right]} \\
&\leq \sqrt[p]{\mathrm{E}\left[\left|\varepsilon_t\right|^p\right]} + \sqrt[p]{\mathrm{E}\left[\left|u_{t-1}^*\right|^p\right]} + \sqrt[p]{\mathrm{E}\left[\left|R_{t,h_T}\right|^p\right]} < C
\end{aligned}$$

so indeed $\mathrm{E}\left[\left|\mathcal{L}'\left(\varepsilon_t + u_{t-1}^* + \upsilon_t\right)\right|^{\frac{p}{p-1}}\right] < C$ for all t as required. An equicontinuity argument similar to the one used for the case $p = 1$ leads to the desired result.

Furthermore, $\varepsilon_t + u_{t-1}^*$ is itself stationary and ergodic, and it follows with Lemma 2.1 that it also is L_p bounded, so $\mathrm{E}\left(\mathcal{L}\left(\varepsilon_t + u_{t-1}^*\right)\right) < \infty$. It then holds due to the ergodic theorem (see e.g. Davidson, 1994, Theorem 13.12) that

$$\frac{1}{T}\sum_{t=1}^{T} \mathcal{L}\left(\varepsilon_t + u_{t-1}^*\right) \xrightarrow{p} \mathrm{E}\left(\mathcal{L}\left(\varepsilon_t + u_{t-1}^*\right)\right)$$

as $T \to \infty$. We now establish uniformity of the above convergence. To accomplish this task, we show the sequence of target functions to be stochastically equicontinuous. See, among others, Andrews (1992) for a discussion on generic uniform convergence. We show the following condition to hold true

$$\sup_{\left\|\left(\boldsymbol{a}_1^{*\prime}, b_1^*\right)' - \left(\boldsymbol{a}_2^{*\prime}, b_2^*\right)'\right\|_1 < \delta_T} \left|\frac{1}{T}\sum_{t=1}^{T} \mathcal{L}\left(\varepsilon_t + u_{1,t-1}^*\right) - \frac{1}{T}\sum_{t=1}^{T} \mathcal{L}\left(\varepsilon_t + u_{2,t-1}^*\right)\right| \xrightarrow{p} 0 \tag{2.16}$$

for some deterministic sequence $\delta_T \to 0$ and $u_{i,t-1}^* = -\sum_{j\geq 1}\left(a_{i,j}^* - a_j\right) y_{t-j} - \left(b_i^* - b\right)$.

For $p = 1$, a Lipschitz argument immediately leads to the desired result, while for $p > 1$ we resort again to the mean value theorem to obtain for each t

$$\mathcal{L}\left(\varepsilon_t + u_{1,t-1}^*\right) - \mathcal{L}\left(\varepsilon_t + u_{2,t-1}^*\right) = \mathcal{L}'\left(\xi_t\right)\left(u_{1,t-1}^* - u_{2,t-1}^*\right)$$

for $\xi_t = w\left(\varepsilon_t + u_{1,t-1}^*\right) + (1-w)\left(\varepsilon_t + u_{2,t-1}^*\right)$ where $w \in [0,1]$. Therefore,

$$\left|\frac{1}{T}\sum_{t=1}^{T} \mathcal{L}\left(\varepsilon_t + u_{1,t-1}^*\right) - \frac{1}{T}\sum_{t=1}^{T} \mathcal{L}\left(\varepsilon_t + u_{2,t-1}^*\right)\right| \leq \frac{1}{T}\sum_{t=1}^{T}\left|\mathcal{L}'\left(\xi_t\right)\right|\left(\left|\sum_{j\geq 1}\left(a_{1,j}^* - a_{2,j}^*\right) y_{t-j}\right| + \left|b_1^* - b_2^*\right|\right).$$

With ξ_t easily shown to be uniformly L_p-bounded, and therefore $\mathcal{L}'\left(\xi_t\right)$ uniformly $L_{\frac{p}{p-1}}$-

bounded, Markov's inequality implies that

$$|b_1^* - b_2^*| \frac{1}{T} \sum_{t=1}^{T} |\mathcal{L}'(\xi_t)| \xrightarrow{p} 0;$$

at the same time, $\left(\sum_{j\geq 1} |a_{1,j}^* - a_{2,j}^*|\right)^{-1} \left|\sum_{j\geq 1} \left(a_{1,j}^* - a_{2,j}^*\right) y_{t-j}\right|$ is uniformly L_p-bounded, so Hölder's inequality implies that

$$\mathrm{E}\left[|\mathcal{L}'(\xi_t)| \left|\sum_{j\geq 1} \left(a_{1,j}^* - a_{2,j}^*\right) y_{t-j}\right|\right] \leq C \sum_{j\geq 1} \left|a_{1,j}^* - a_{2,j}^*\right| \leq C\delta_T \to 0$$

as required to establish (2.16).

Summing up, it holds that

$$\sup_{\left(\boldsymbol{a}_{h_T}^{*\prime}, b^*\right)\in\Theta} \left| \mathcal{Q}_T\left(\boldsymbol{a}_{h_T}^{*\prime}, b^*\right) - \mathrm{E}\left[\mathcal{L}\left(y_t - \sum_{j\geq 1} a_j^* y_{t-j} - b^*\right)\right]\right| \xrightarrow{p} 0 \qquad (2.17)$$

as $T, h_T \to \infty$. But the expectation is the target function in Proposition 2.1, which is uniquely minimized at parameters solving the f.o.c. from Proposition 2.1. Since the pseudo-true values a_j, b satisfy exactly these conditions according to Assumption 2.3, identification is provided for. Given uniform convergence, the result follows with Theorem 4.1.1 in Amemiya (1985) since Θ is compact and the target function is continuous.

Proof of Proposition 2.4

We drop b for simplicity, as it does not affect the derivations in an essential manner. Also, it is more convenient to now treat $\mathcal{Q}_T$ as having h_T arguments rather than $\boldsymbol{a} \in \ell_1$. For the case $p > 2$, the proof uses the usual argument of an elementwise first-order Taylor expansion of the gradient of the target function around $\boldsymbol{a}_{h_T}$. Let $s_j\left(\boldsymbol{a}_{h_T}^*\right)$, $j = 1, \ldots, h_T$, be the jth element of the gradient of $\mathcal{Q}_T$, $\boldsymbol{s} = \{s_j\}_{1\leq j\leq h_T}$. Then,

$$s_j\left(\hat{\boldsymbol{a}}_{h_T}\right) = s_j\left(\boldsymbol{a}_{h_T}\right) + \left.\frac{\partial s_j\left(\boldsymbol{a}_{h_T}^*\right)}{\partial \boldsymbol{a}_{h_T}^*}\right|_{\boldsymbol{\xi}_{j,h_T}}^{\prime} \left(\hat{\boldsymbol{a}}_{h_T} - \boldsymbol{a}_{h_T}\right),$$

where $\boldsymbol{\xi}_{j,h_T}$ is a convex combination of $\boldsymbol{a}_{h_T}$ and $\hat{\boldsymbol{a}}_{h_T}$. Since $\hat{\boldsymbol{a}}_{h_T}$ is the solution of (2.6), it holds that $\boldsymbol{s}\left(\hat{\boldsymbol{a}}_{h_T}\right) = \boldsymbol{0}$, and we obtain after pre-multiplication with the inverse of the Hessian that

$$\left\|\hat{\boldsymbol{a}}_{h_T} - \boldsymbol{a}_{h_T}\right\|_2 \leq \left\|\left(\Xi_{h_T}\right)^{-1}\right\|_2 \left\| \left.\frac{\partial \mathcal{Q}_T\left(\boldsymbol{a}_{h_T}^*\right)}{\partial \boldsymbol{a}_{h_T}^*}\right|_{\boldsymbol{a}_{h_T}} \right\|_2,$$

where $\|\cdot\|_2$ denotes the induced matrix norm and $\boldsymbol{\Xi}_{h_T}$ is the matrix having $\left.\frac{\partial s_j\left(\boldsymbol{a}^*_{h_T}\right)}{\partial \boldsymbol{a}^*_{h_T}}\right|'_{\boldsymbol{\xi}_{j,h_T}}$ as jth row. Now,

$$s_j\left(\boldsymbol{a}_{h_T}\right) = -\frac{1}{T}\sum_{t=h_T+1}^{T} y_{t-j}\mathcal{L}'\left(\varepsilon_t + \sum_{k>h_T} a_k y_{t-k}\right)$$

where

$$\mathcal{L}'\left(\varepsilon_t + \sum_{k>h_T} a_k y_{t-k}\right) = \mathcal{L}'\left(\varepsilon_t\right) + \mathcal{L}''\left(\psi_t\right)\sum_{k>h_T} a_k y_{t-k}$$

with ψ_t a convex combination of $\varepsilon_t + \sum_{k>h_T} a_k y_{t-k}$ and ε_t. The shape of the loss function, the s-summability of a_k and the uniform moment properties of ε_t and y_t imply that $\mathcal{L}''\left(\psi_t\right)$ is uniformly $L_{2p/(p-2)}$ bounded. We have therefore that

$$s_j = -\frac{1}{T}\sum_{t=h_T+1}^{T} y_{t-j}\mathcal{L}'\left(\varepsilon_t\right) - R_{j,T},$$

where

$$\sqrt{\mathrm{E}\left[\left|R_{j,T}\right|^2\right]} = \sqrt{\mathrm{E}\left[\left|\frac{1}{T}\sum_{t=h_T+1}^{T} y_{t-j}\mathcal{L}''\left(\psi_t\right)\sum_{k>h_T} a_k y_{t-k}\right|^2\right]} \leq \sum_{k>h_T}\left|a_k\right|\sqrt{\mathrm{E}\left[\left|y_{t-j}\mathcal{L}''\left(\psi_t\right)y_{t-k}\right|^2\right]}.$$

It is easily shown by using the Hölder's inequality that $\mathrm{E}\left[\left|y_{t-j}\mathcal{L}''\left(\psi_t\right)y_{t-k}\right|^2\right]$ is uniformly bounded, which implies

$$\sqrt{\mathrm{E}\left[\left|R_{j,T}\right|^2\right]} = o\left(h_T^{-s}\right)$$

uniformly in j. We also have uniformly in j that

$$\mathrm{Var}\left[\frac{1}{T}\sum_{t=h_T+1}^{T} y_{t-j}\mathcal{L}'\left(\varepsilon_t\right)\right] = O\left(\frac{1}{T}\right)$$

thanks to the md property of $\mathcal{L}'\left(\varepsilon_t\right)$ and the moment properties of $\mathcal{L}'\left(\varepsilon_t\right)$ and y_{t-j}, so, summing up,

$$\left\|\boldsymbol{s}\left(\boldsymbol{a}_{h_T}\right)\right\|_2 = O_p\left(\sqrt{h_T}\max\left\{h_T^{-s}, T^{-0.5}\right\}\right).$$

To discuss (the inverse of) $\boldsymbol{\Xi}_{h_T}$, note that its i,jth element is given by

$$\frac{1}{T}\sum_{t=h_T+1}^{T} y_{t-i}y_{t-j}\mathcal{L}''\left(\varepsilon_t + \sum_{k=1}^{h_T}\left(a_k - \xi_{k,j,h_T}\right)y_{t-k} + \sum_{k>h_T} a_k y_{t-k}\right)$$

where $\boldsymbol{\xi}_{j,h_T} = (\xi_{1,j,h_T}, \ldots, \xi_{h_T,j,h_T})'$. Let $\underline{L} = \min_{u\in\mathbb{R}} \mathcal{L}''(u)$ and write Ξ_{h_T} as the sum of two matrices,

$$\Xi_{h_T} = \mathbf{A} + \mathbf{B},$$

where

$$\mathbf{A} = \left\{ \underline{L} \frac{1}{T} \sum_{t=h_T+1}^{T} y_{t-i} y_{t-j} \right\}_{i,j}$$

and

$$\mathbf{B} = \left\{ \frac{1}{T} \sum_{t=h_T+1}^{T} y_{t-i} y_{t-j} \left(\mathcal{L}'' \left(\varepsilon_t + \sum_{k=1}^{h_T} (a_k - \xi_{k.j,h_T})\, y_{t-k} + \sum_{k>h_T} a_k y_{t-k} \right) - \underline{L} \right) \right\}_{i,j}.$$

Note that $\mathbf{A}$ is the scaled sample autocovariance matrix of y_t (with $\underline{L} > 0$), while $\mathbf{B}$ may be written as

$$\mathbf{B} = \mathbf{Y}'\mathbf{D}\mathbf{Y}$$

with $\mathbf{Y}$ stacking h_T lags of y_t, $t = h_T + 1, \ldots, T$ and

$$\mathbf{D} = \operatorname{diag} \left\{ \mathcal{L}'' \left(\varepsilon_t + \sum_{k=1}^{h_T} (a_k - \xi_{k.j,h_T})\, y_{t-k} + \sum_{k>h_T} a_k y_{t-k} \right) - \underline{L} \right\}$$

having nonnegative diagonal elements by construction. Therefore, both $\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite; for two positive semidefinite matrices $\mathbf{A}$ and $\mathbf{B}$ it can be shown that the smallest eigenvalue of the sum $\mathbf{A}+\mathbf{B}$ is not smaller than any of the eigenvalues of $\mathbf{A}$ or $\mathbf{B}$ (to see this, write min $eigenval(\mathbf{A}+\mathbf{B}) = \min_{\boldsymbol{x}} \frac{\boldsymbol{x}'(\mathbf{A}+\mathbf{B})\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} \geq \min_{\boldsymbol{x}} \left(\frac{\boldsymbol{x}'\mathbf{A}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} + \frac{\boldsymbol{x}'\mathbf{B}\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} \right)$ where the latter summands must both be nonnegative). Summing up, the smallest eigenvalue of Ξ_{h_T} is not smaller than the smallest eigenvalue of the autocovariance matrix of y_t, implying upon inversion that

$$\left\| (\Xi_{h_T})^{-1} \right\|_2 \leq \frac{1}{\underline{L}} \left\| \left(\left\{ \frac{1}{T} \sum y_{t-i} y_{t-j} \right\}_{i,j} \right)^{-1} \right\|_2 = O_p(1)$$

by Assumption 2.5. Summing up (again), we obtain

$$\| \hat{\boldsymbol{a}}_{h_T} - \boldsymbol{a}_{h_T} \|_2 = O_p \left(\max \left\{ h_T^{1/2-s}, \sqrt{\frac{h_T}{T^{1/2}}} \right\} \right).$$

The extension for $p = 2$, where for $\alpha \neq 0.5$ we assumed ε_t to have absolutely continuous conditional distribution, is tedious yet straightforward and we omit the details.

Additional figures

Figure 2.5: Relative forecast performance for strongly serially correlated ARMA-GARCH processes (Student $t(50)$ innovations, $p = 3$)

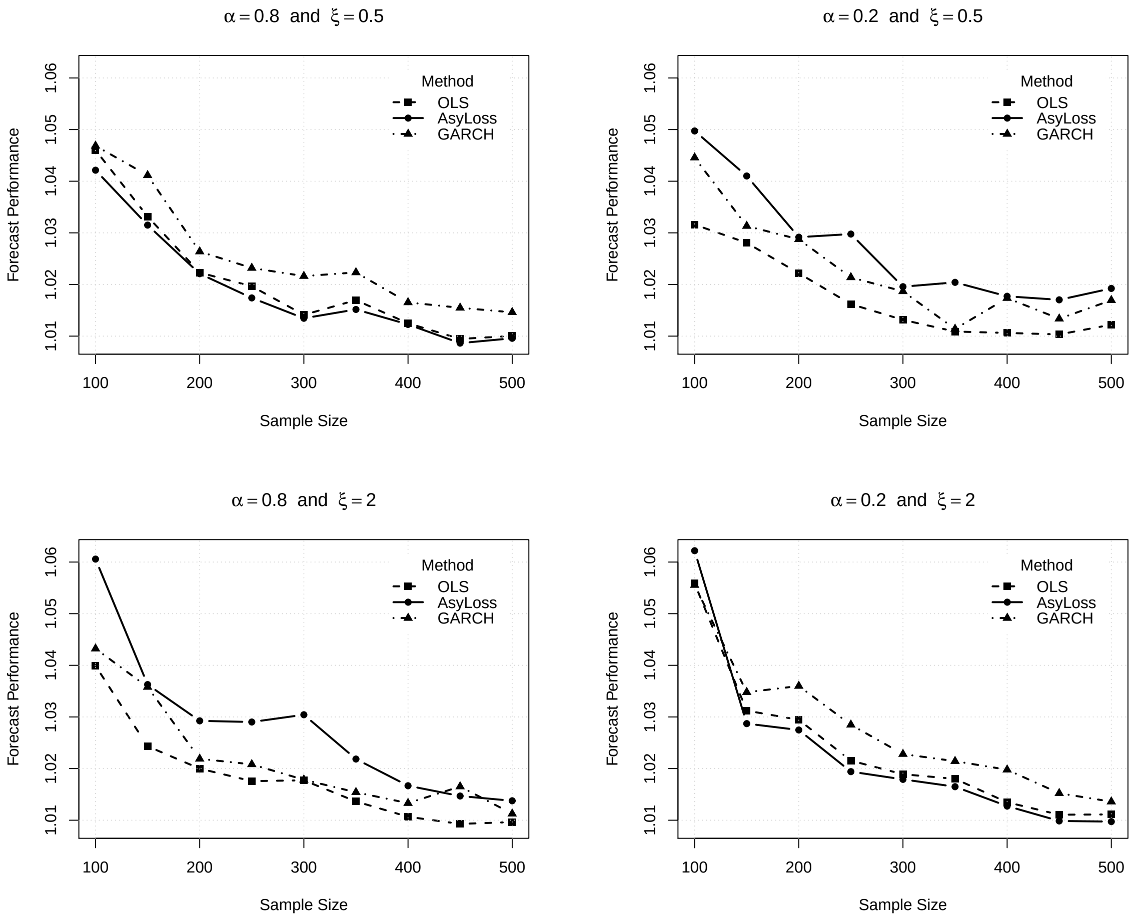


Figure 2.6: Relative forecast performance for weakly serially correlated ARMA-GARCH processes (Student $t(50)$ innovations, $p = 3$)

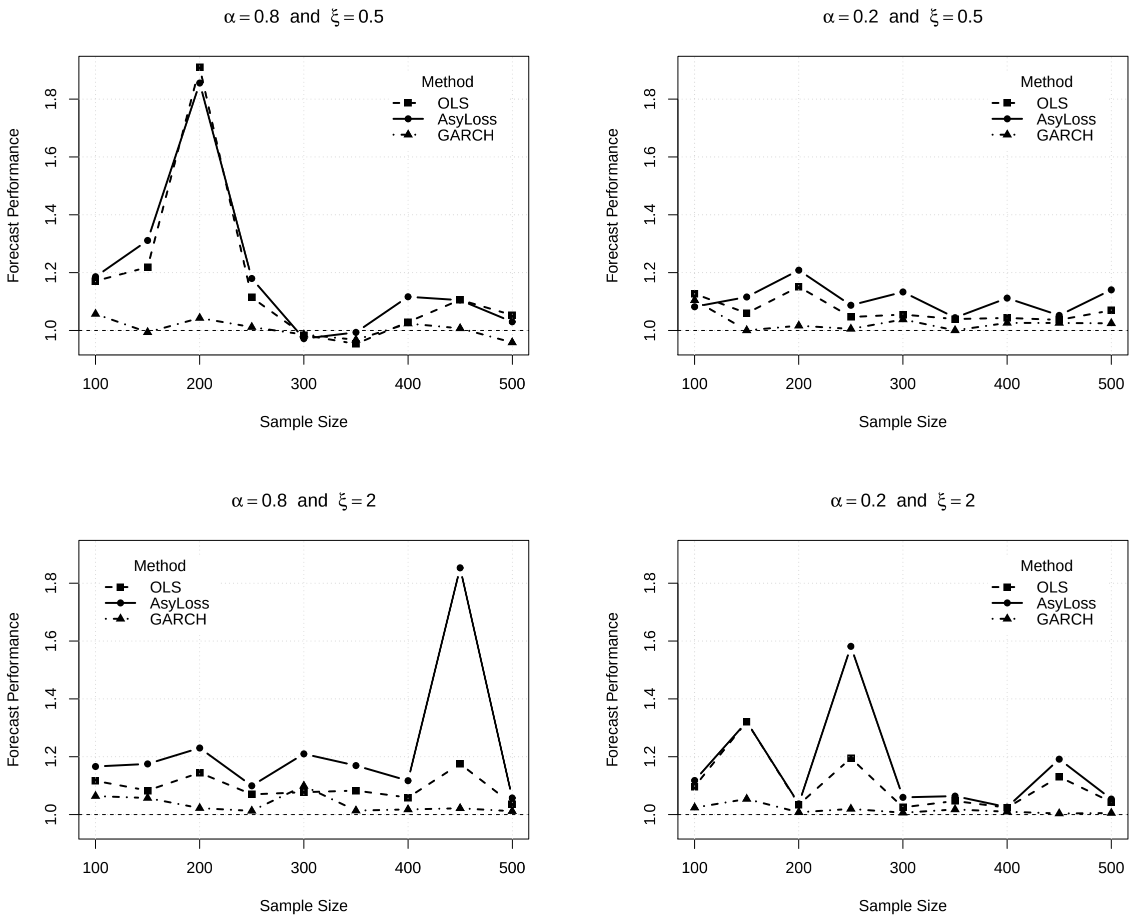


Figure 2.7: Relative forecast performance for strongly serially correlated ARMA-GARCH processes (Student $t(5)$ innovations, $p = 2$)

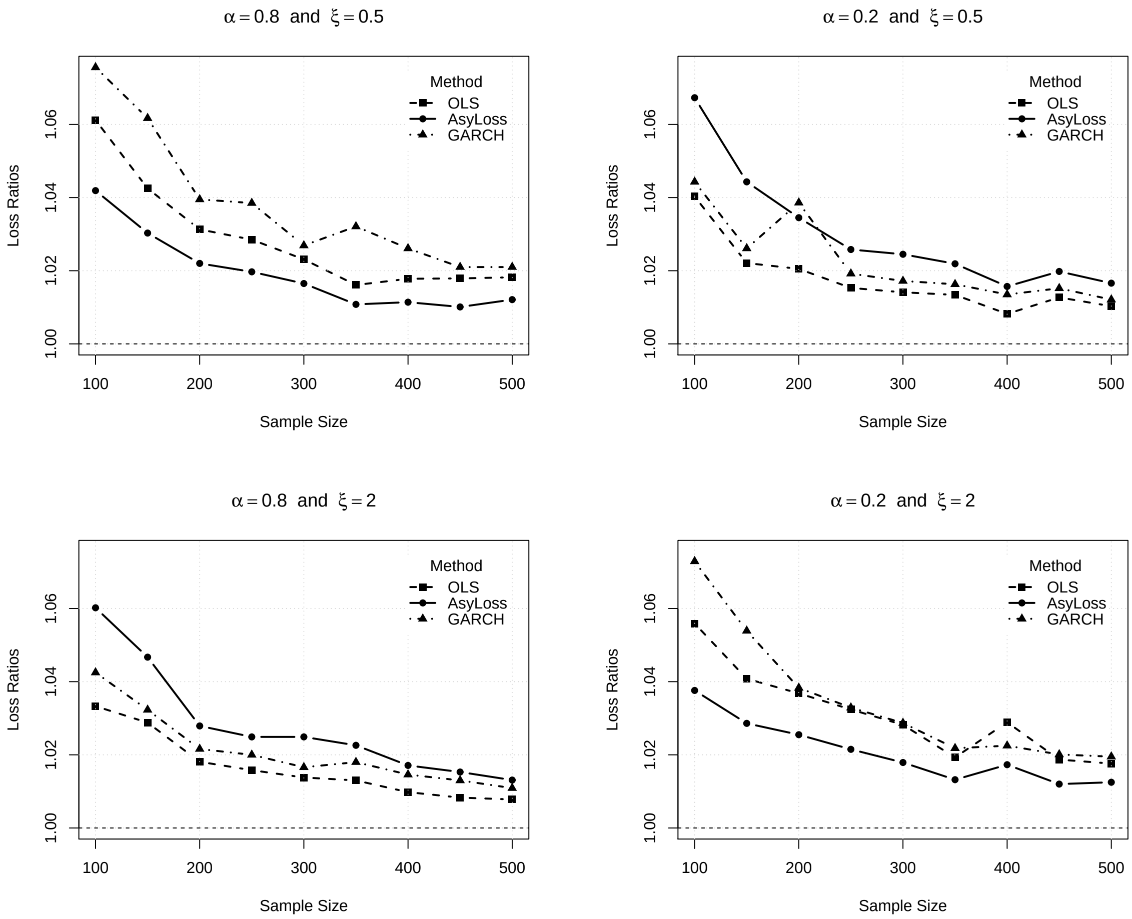
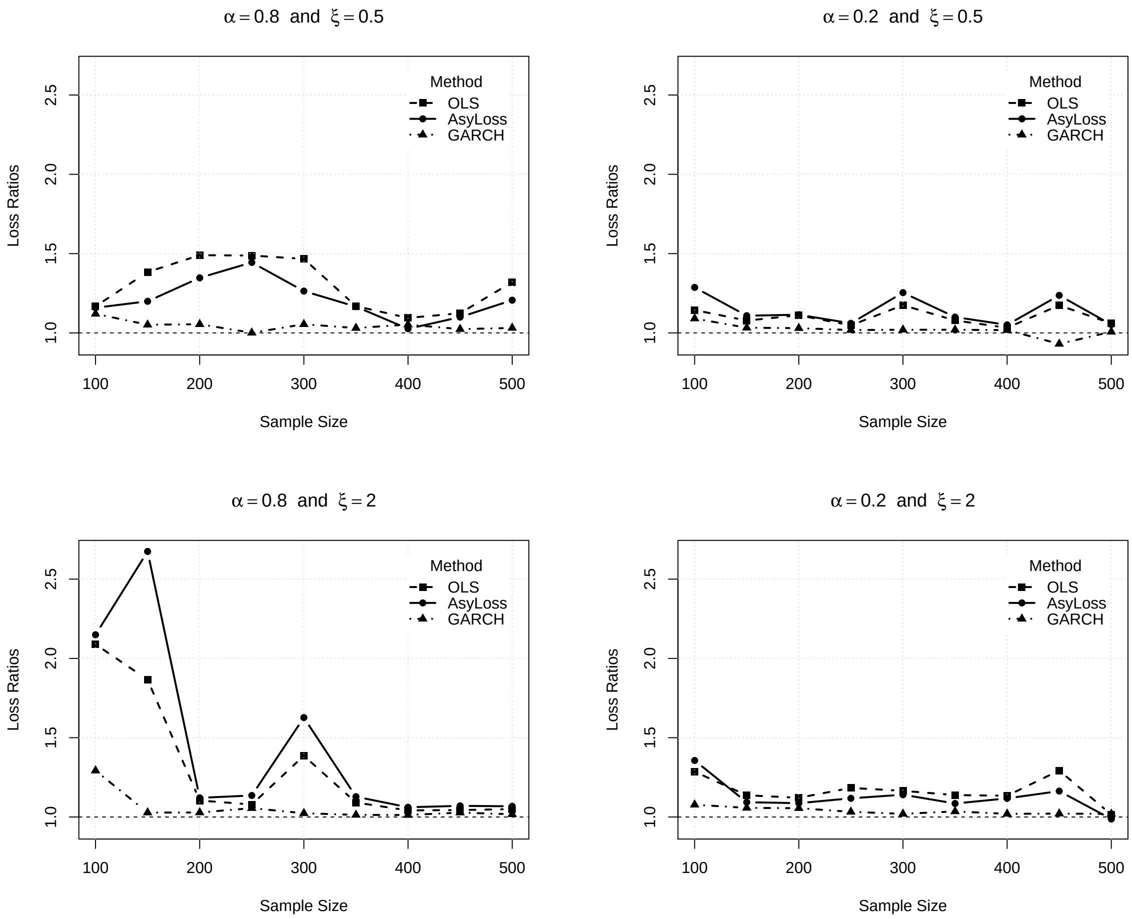


Figure 2.8: Relative forecast performance for weakly serially correlated ARMA-GARCH processes (Student $t(5)$ innovations, $p = 2$)



Chapter 3

Re-Evaluating the Prudence of Economic Forecasts in the EU: The role of instrument persistence

Coauthored by: Matei Demetrescu and Christoph Roling

3.1 Motivation

The evaluation of forecasts is one of the important feed-back loops in applied econometrics. It has e.g. become routine to test realized forecast error series for unbiasedness or lack of serial correlation. But most tests rely on the assumption that the relevant optimality criterion is based on the mean squared error [MSE]. Under other loss functions, and in particular asymmetric ones, it actually is quite rational to produce biased forecasts. Thus, it is of interest to learn the shape of the loss function underlying a given sequence of forecasts. Elliott et al. (2005) propose a class of loss functions indexed by two parameters: the asymmetry and the tail weight. GMM estimation of the asymmetry parameter is possible (for simplicity often assuming known tail weight), provided that variables are available, which would improve forecasts if the latter were not rational; such instrument variables are e.g. the original predictors, lagged forecast errors, or lagged target variables.

Building on this method, Christodoulakis and Mamatzakis (2008, 2009) find asymmetric preferences in series forecasts of EU institutions and countries. Clements et al. (2007) and Capistrán (2008) discuss the loss function of the Federal Reserve, while Pierdzioch et al. (2011) find evidence of asymmetry in the loss function of the Bank of Canada, and Tsuchiya (2016) provides evidence for Japan. Along the same lines, Elliott et al. (2008), Boero et al. (2008), Aretz et al. (2011), Clatworthy et al. (2012) or Fritsche et al. (2015) find asymmetric loss preferences of individual forecasters.

One limitation of the GMM method by Elliott et al. (2005) is that the instruments are assumed stationary. This may seem benign; but many typical instruments are, on the contrary, quite persistent. In their assessment of the EU Commission loss preferences, Christodoulakis and Mamatzakis (2009) use the lagged levels of inflation, unemployment, government balance, investment and current account as instruments, which are often regarded as having a stochastic trend; see Section 3.3 for concrete evidence on the persistence of these variables. We therefore re-examine the forecast preferences of the EU Commission in light of possibly highly persistent instruments.

To this end, we show in Section 3.2 that the limiting distribution of the GMM estimator is non-Gaussian in general. Moreover, the family of J-tests discussed by Elliott et al. (2005) loses the chi-square limiting distributions, and decisions based on χ^2 critical values lead to over-rejections of the respective null hypotheses. But we also find some cases where the limiting distribution of the t-statistic of the asymmetry parameter is standard normal in spite of instrument persistence, such that the usual critical values are still appropriate. This is the case under a homoskedasticity restriction, or when a persistent instrument is combined with an intercept only.

In Section 3.3 we extend the original data set to cover the full 1970–2016 period, and use our theoretical insights to re-evaluate the prudence of economic forecasts of the EU Commission in a robust fashion. Christodoulakis and Mamatzakis (2009) found that apparent irrational forecasting behavior of the EU commission can well be explained by asymmetries in the loss preferences. For the year-ahead forecasts, they find that forecasts tend to be optimistic, while the current-year forecasts are somewhat pessimistic (as if to counter-balance the year-ahead optimism). We confirm the conclusions of Christodoulakis and Mamatzakis for the original data. Considering the extended data set, we observe however an overall reduction of deviations from symmetry in the EU commission forecasts, as well as a somewhat reduced evidence of forecast irrationality. The final section concludes.

3.2 GMM inference under instrument persistence

3.2.1 Estimation of asymmetry

To keep the note self-contained, we briefly review the GMM estimation procedure of Elliott et al. (2005). The one-step ahead optimal predictor of a series y_t is given by

$$\hat{y}_t = \arg\min_{y^*} \mathrm{E}_{t-1} \left(\mathcal{L} \left(y_t - y^* \right) \right),$$

where $\mathcal{L}$ denotes the relevant loss function – which should be quasi-convex (see e.g. Granger, 1999) – and E_{t-1} denotes the expectation taken w.r.t. the conditional forecast density given the available information. The class of functions proposed by Elliott et al. (2005) is given by

$$\mathcal{L}(u) = (\alpha + (1 - 2\alpha) \mathbf{1}(u < 0)) |u|^p, \tag{3.1}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Let $u_t = y_t - \hat{y}_t$ denote the realized forecast error at time t. Should the forecast be indeed optimal, no information available at the time of the forecast can reduce forecasting risk, and the so-called generalized forecast error

$$\mathcal{L}'(u_t) = p\left(\alpha - \mathbf{1}(u_t < 0)\right)|u_t|^{p-1} =: \tilde{u}_t(\alpha)$$

is a martingale difference sequence, $\mathrm{E}\left(\tilde{u}_t | \tilde{u}_{t-1}, \ldots\right) = 0$ (Granger, 1999; Patton and Timmermann, 2007b). We sometimes write $\tilde{u}_t(\alpha) = \tilde{u}_t$ in the following, should dependence on α not be of essence.

Elliott et al. (2005) exploit this martingale difference property to estimate the asymmetry parameter α from a series of observed forecast errors u_t, $t = 1, \ldots, T$. Concretely, one employs for given p (typically chosen as $p = 1$ or $p = 2$) a set of D instrument variables gathered in the vector $\boldsymbol{v}_{t-1}$. The instruments may be, but are not restricted to, predictors from the original forecasting model; they cannot improve the forecasts when these are optimal, so $\boldsymbol{v}_{t-1}$ must belong to the information set at time $t-1$ and $\mathrm{E}\left(\tilde{u}_t | \tilde{u}_{t-1}, \ldots, \boldsymbol{v}_{t-1}, \ldots\right) = 0$ under rationality. This implies D moment restrictions,

$$\mathrm{E}\left(\boldsymbol{v}_{t-1}\tilde{u}_t(\alpha)\right) = p\,\mathrm{E}\left(\boldsymbol{v}_{t-1}\left(\alpha - \mathbf{1}(u_t < 0)\right)|u_t|^{p-1}\right) = 0,$$

leading for the loss functions from (3.1) to the GMM estimator

$$\hat{\alpha} = \frac{\hat{\boldsymbol{h}}'\,\hat{\mathbf{S}}^{-1}\left(\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\mathbf{1}(u_t < 0)|u_t|^{p-1}\right)}{\hat{\boldsymbol{h}}'\,\hat{\mathbf{S}}^{-1}\,\hat{\boldsymbol{h}}} = \alpha - \frac{1}{p}\frac{\hat{\boldsymbol{h}}'\,\hat{\mathbf{S}}^{-1}\left(\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\tilde{u}_t\right)}{\hat{\boldsymbol{h}}'\,\hat{\mathbf{S}}^{-1}\hat{\boldsymbol{h}}},$$

where

$$\hat{\boldsymbol{h}} = \frac{1}{T}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}|u_t|^{p-1} \quad \text{and} \quad \hat{\mathbf{S}} = \frac{1}{T}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\boldsymbol{v}_{t-1}'\left(\mathbf{1}(u_t < 0) - \hat{\alpha}\right)^2|u_t|^{2p-2}.$$

The matrix $\hat{\mathbf{S}}$ is nothing other than the scaled sample covariance matrix of the sample moment conditions $\boldsymbol{v}_{t-1}\tilde{u}_t$ (exploiting the zero expectation). Note that, for estimation, an iterative procedure is required since $\hat{\mathbf{S}}$ depends on α via $\tilde{u}_t = \tilde{u}_t(\alpha)$. Under the assumptions of Elliott et al. (2005), a limiting normal distribution holds for $\hat{\alpha}$ as $T \to \infty$,

$$\sqrt{T}\left(\hat{\alpha} - \alpha\right) \xrightarrow{d} \mathcal{N}\left(0, \operatorname{plim}\hat{V}\right),$$

where the standard error of $\hat{\alpha}$ is given by $\hat{V}^{1/2} = (\hat{\boldsymbol{h}}'\,\hat{\mathbf{S}}^{-1}\hat{\boldsymbol{h}})^{-1/2}$, and the plim exists and is positive. Confidence intervals with asymptotic coverage $1-\gamma$ are easily built as $\hat{\alpha} \pm z_{1-\gamma/2}T^{-1/2}\hat{V}^{1/2}$ with $z_{1-\gamma/2}$ the $1-\gamma/2$ quantile of the standard normal distribution. Moreover, J-statistics are available,

$$\mathcal{J}_{\cdot} = \frac{1}{p^2}\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}'\tilde{u}_t(\cdot)\right)\hat{\mathbf{S}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\tilde{u}_t(\cdot)\right),$$

for which χ^2 limiting distributions arise. Under stationarity assumptions, $\mathcal{J}_{\hat{\alpha}}$ (which may be interpreted as a rationality test) has a limiting null χ^2 distribution with $D-1$ degrees of freedom (so it requires the use of at least two instruments); if testing hypotheses of the type $\alpha = \alpha_0$, $\mathcal{J}_{\alpha_0}$

has a limiting null $\chi^2(D)$ distribution. Alternatively, null hypotheses on α can easily be checked using the t-ratio $\mathcal{T} = (\hat{\alpha} - \alpha_0)/\hat{V}^{1/2}$.[1] Note that, for one instrument only ($D = 1$), $\mathcal{J}_{\alpha_0}$ is the same as $\mathcal{T}^2$. Conversely, confidence intervals for α may be obtained by inverting the tests $\mathcal{J}_{\alpha_0}$.

The normal and χ^2 distributions hinge on whether a central limit theorem may be applied to the term $\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\tilde{u}_t$. The summands $\boldsymbol{v}_{t-1}\tilde{u}_t$ form by construction a martingale difference sequence, but a "classical" central limit theorem does not apply when elements of $\boldsymbol{v}_{t-1}$ are highly persistent (exhibit stochastic trends, e.g. random walks) and is thus nonstationary. We analyze in the following the limiting distribution of $\mathcal{J}_\alpha$ and $\mathcal{T}$ when some of the instruments are allowed to be strongly persistent in the sense that they possess a generic stochastic trend.

3.2.2 Assumptions and main result

To cover the case of both stationary and persistent instruments, partition

$$\boldsymbol{v}_{t-1} = \left(\boldsymbol{v}_{0,t-1}', 1, \boldsymbol{v}_{1,t-1}'\right)'$$

where the D_1 persistent instruments $\boldsymbol{v}_{1,t}$ satisfy the following technical requirement:

Assumption 3.1 *Let $\mathbf{N}_T$ be a diagonal matrix with $N_{T,ii} \to \infty$, and assume that there exists a continuous-time vector process $\boldsymbol{X}(s)$ such that the weak convergence $\mathbf{N}_T^{-1}\boldsymbol{v}_{1,[sT]} \Rightarrow \boldsymbol{X}(s)$ holds in a space of càdlàg functions endowed with a suitable norm.*

The assumption allows e.g. for near-integrated modelling of predictors (see e.g. Campbell and Yogo, 2006, and the references therein) but not exclusively. E.g. $\boldsymbol{X}(s)$ may be a (multivariate) Ornstein-Uhlenbeck [OU] process with non-zero starting value, corresponding in discrete time to a near-integrated process with initial condition drawn from the unconditional distribution (as employed by Müller and Elliott, 2003). Also, $\boldsymbol{X}(s)$ may be a fractional Brownian motion (see e.g. Maynard and Phillips, 2001, and the references therein). For the D_0 stationary instruments $\boldsymbol{v}_{0,t}$, we make

Assumption 3.2 *Let $\left(\tilde{u}_t, \boldsymbol{v}_{0,t-1}'\right)' \in \mathbb{R}^{D_0+1}$ be a zero-mean stationary, ergodic, uniformly $L_{4+\delta}$-bounded sequence for some $\delta > 0$, such that* $\mathrm{E}\left(\tilde{u}_t | \tilde{u}_{t-1}, \ldots, \boldsymbol{v}_{t-1}, \ldots\right) = 0$.

Theorem 27.14 in Davidson (1994) then implies

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{[sT]}\begin{pmatrix} \tilde{u}_t \\ \boldsymbol{v}_{0,t-1}\tilde{u}_t \end{pmatrix} \Rightarrow \begin{pmatrix} \tilde{W}(s) \\ \bar{\boldsymbol{W}}(s) \end{pmatrix},$$

with $\tilde{W}(s)$ and $\bar{\boldsymbol{W}}(s)$ Brownian motions. Take this to be joint with the weak convergence in Assumption 3.1, and let

$$\mathrm{Cov}\begin{pmatrix} \tilde{W}(1) \\ \bar{\boldsymbol{W}}(1) \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(\tilde{u}_t) & \mathrm{E}\left(\boldsymbol{v}_{0,t-1}'\tilde{u}_t^2\right) \\ \mathrm{E}\left(\boldsymbol{v}_{0,t-1}\tilde{u}_t^2\right) & \mathrm{E}\left(\boldsymbol{v}_{0,t-1}\boldsymbol{v}_{0,t-1}'\tilde{u}_t^2\right) \end{pmatrix} = \begin{pmatrix} \sigma_{\tilde{u}}^2 & \boldsymbol{\gamma}_0' \\ \boldsymbol{\gamma}_0 & \bar{\boldsymbol{\Omega}}_0 \end{pmatrix}$$

[1] When the purpose is testing rather than estimating, one may compute $\hat{\mathbf{S}}$ under the null – i.e. replace $\hat{\alpha}$ with α_0 – to reduce computational requirements.

for later reference; in case of conditional homoskedasticity of $\tilde{u}$, $\bar{\boldsymbol{\Omega}}_0 = \sigma^2_{\tilde{u}} \operatorname{Cov}(\boldsymbol{v}_{0,t-1})$ and $\boldsymbol{\gamma}_0 = \mathbf{0}$.

Having a constant in the vector of instruments is on the one hand common in practice; on the other hand, the constant stands in for $I(0)$ processes with nonzero mean since the purely stochastic component would be dominated in the limit; see the proof of Proposition 3.1. Hence requiring that $\boldsymbol{v}_{0,t-1}$ have zero mean does not imply any loss of generality, and one essentially has what one may call weakly persistent instruments.[2]

Proposition 3.1 *Under Assumptions 3.1 and 3.2, it holds as $T \to \infty$ that*

$$\sqrt{T}\left(\hat{\alpha} - \alpha\right) \Rightarrow \frac{1}{p} \frac{\boldsymbol{H}'\mathbf{S}^{-1}\boldsymbol{U}}{\boldsymbol{H}'\mathbf{S}^{-1}\boldsymbol{H}} \qquad \textit{and} \qquad \mathcal{T} \Rightarrow \frac{1}{p} \frac{\boldsymbol{H}'\mathbf{S}^{-1}\boldsymbol{U}}{\sqrt{\boldsymbol{H}'\mathbf{S}^{-1}\boldsymbol{H}}}$$

where

$$\boldsymbol{H} \equiv \begin{pmatrix} \operatorname{E}\left(\boldsymbol{v}_{0,t-1} \left|u_t\right|^{p-1}\right) \\ \operatorname{E}\left(\left|u_t\right|^{p-1}\right) \\ \operatorname{E}\left(\left|u_t\right|^{p-1}\right) \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s \end{pmatrix}, \quad \boldsymbol{U} \equiv \begin{pmatrix} \bar{\boldsymbol{W}}(1) \\ \tilde{W}(1) \\ \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}\tilde{W}(s) \end{pmatrix}$$

and

$$\mathbf{S} \equiv \frac{1}{p^2} \begin{pmatrix} \bar{\boldsymbol{\Omega}}_0 & \boldsymbol{\gamma}_0 & \boldsymbol{\gamma}_0 \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \boldsymbol{\gamma}_0' & \sigma^2_{\tilde{u}} & \sigma^2_{\tilde{u}} \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s\, \boldsymbol{\gamma}_0' & \sigma^2_{\tilde{u}} \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s & \sigma^2_{\tilde{u}} \int_0^1 \boldsymbol{X}(s) \boldsymbol{X}'(s)\,\mathrm{d}s \end{pmatrix}.$$

Proof: *See the Appendix.*

Proposition 3.2 *Under the assumptions of Proposition 3.1, it holds under the respective null hypotheses that, as $T \to \infty$,*

$$\mathcal{J}_{\alpha_0} \xrightarrow{d} \frac{1}{p^2} \boldsymbol{U}'\mathbf{S}^{-1}\boldsymbol{U} \quad \textit{and} \quad \mathcal{J}_{\hat{\alpha}} \xrightarrow{d} \frac{1}{p^2} \boldsymbol{U}' \left(\mathbf{S}^{-1} - \frac{1}{\boldsymbol{H}'\mathbf{S}^{-1}\boldsymbol{H}} \mathbf{S}^{-1}\boldsymbol{H}\boldsymbol{H}'\mathbf{S}^{-1}\right) \boldsymbol{U}.$$

Proof: *Analogous to the proof of Proposition 3.1 and omitted.*

3.2.3 Discussion

Above all, it should be emphasized that the estimator $\hat{\alpha}$ is still consistent; it is only the higher-order properties that are affected by instrument persistence. Yet, in spite of the nonstandard distribution of $\hat{\alpha}$, the GMM estimator is $\sqrt{T}$-consistent only, and not superconsistent as might have been expected given stochastically trending instruments.

Given the presence of the Itô-type integral in $\boldsymbol{U}$, Propositions 3.1 and 3.2 imply nonstandard distributions in general, for both $\hat{\alpha}$ and the J-statistics. Take as an extreme example the case

[2] This terminology may conflict with the persistence notion associated with long memory processes: stationary long memory is allowed for $\boldsymbol{v}_{0,t}$, uniform $L_{4+\delta}$-boundedness provided. We stick to it, though, since it complements strongly persistent processes which are not stationary or ergodic.

where there is exactly one persistent instrument such that, under the null $\alpha = \alpha_0$,

$$\mathcal{J}_{\alpha_0} \Rightarrow \frac{\left(\int_0^1 X(s)\,\mathrm{d}\tilde{W}(s)\right)^2}{\sigma_{\tilde{u}}^2 \int_0^1 X^2(s)\,\mathrm{d}s},$$

which is not χ^2 in general but has a structure akin to that of the Dickey-Fuller distribution.

The fact that the limiting distributions of the J- and $\mathcal{T}$-statistics change if some instruments are persistent has implications on the behavior of tests based on these statistics. Moreover, the distributions derived above depend on the properties of the process $\boldsymbol{X}(s)$, so it is quite difficult to provide practitioners with correct critical values, as these would depend on the particular data generating process [DGP]. In fact, it may well happen that the relevant characteristics of $\boldsymbol{X}(s)$ cannot even be consistently estimated. This is for instance the case with the mean reversion parameter of a near-integrated process (Phillips, 1987). (Note also that other types of persistence than near-integration generate similar behavior, and distinguishing among them – e.g. deciding between fractional integration and near integration – is difficult; see Müller and Watson, 2008.) In such situations, bootstrap schemes will fail too, as they cannot replicate the correct limiting distributions in general when the distributions depend on parameters that cannot be estimated consistently.

Yet standard limiting behavior may be recovered in particular cases. The first case is when the instruments $\boldsymbol{v}_{1,t}$ are "exogenous" in such a way that *mixed* Gaussianity of $\boldsymbol{U}$ is given:

Corollary 3.1 *Let* $\boldsymbol{X}$ *be independent of* $(\tilde{W}, \bar{\boldsymbol{W}}')'$. *Then,* $\mathcal{J}_{\hat{\alpha}} \stackrel{d}{\to} \chi^2_{D-1}$, $\mathcal{J}_{\alpha_0} \stackrel{d}{\to} \chi^2_D$, *and* $\mathcal{T} \stackrel{d}{\to} \mathcal{N}(0,1)$.

Proof: *Obvious and omitted.*

Moreover, if the forecast errors fulfil certain restrictions on the serial dependence of the conditional higher-order moments, the following corollary shows that normality of $\mathcal{T}$ is recovered even in cases where Corollary 3.1 does not apply.

Corollary 3.2 *Let* $\mathrm{E}(|u_t|^{p-1} | \tilde{u}_{t-1}, \ldots, \boldsymbol{v}_{t-1}, \ldots) = \mu_{p-1}$ *and* $\mathrm{E}\left(\tilde{u}_t^2 \middle| \tilde{u}_{t-1}, \ldots, \boldsymbol{v}_{t-1}, \ldots\right) = \sigma_{\tilde{u}}^2$ *be constant. Then,* $\mathcal{T} \stackrel{d}{\to} \mathcal{N}(0,1)$.

Proof: *See the Appendix.*

The corollary essentially requires constant conditional scale of u_t. Also, the presence of a constant instrument is paramount for the result; see the proof for details.

Finally, the constant alone also eliminates nonstandard distribution components from $\mathcal{T}$. The essential requirement is that *only* persistent instruments are employed in conjunction with a constant:

Corollary 3.3 *Let* $\boldsymbol{v}_{t-1} = \left(1, \boldsymbol{v}_{1,t-1}'\right)'$. *Then,* $\mathcal{T} \stackrel{d}{\to} \mathcal{N}(0,1)$.

Proof: *See Appendix.*

Corollaries 3.2 and 3.3 are particularly relevant for applied work under uncertain persistence of the instruments, since they allow for standard inference as long as the nonconstant instruments have the same kind of persistence, without having to specify whether the persistence is weak or strong. This holds irrespective of any conditional heteroskedasticity present in the data.

In what concerns the $\mathcal{J}$-statistics, the expressions in Proposition 3.2 do not appear to simplify under the conditions of either Corollary 3.2 or 3.3.

3.2.4 Gauging the behavior under persistence

We now highlight the extent of the departures from standard asymptotics of $\mathcal{J}$ and $\mathcal{T}$ by means of a Monte Carlo experiment. We combine the frameworks of Engle et al. (1987), Bollerslev (1990) and Gospodinov (2009) to obtain the following DGP:

$$\begin{bmatrix} s_t \\ f_t \end{bmatrix} = \begin{bmatrix} f_{t-1} + \delta\sqrt{h_{1,t}} \\ f_{t-1} \end{bmatrix} + \boldsymbol{\Omega}_t^{1/2}\boldsymbol{\epsilon}_t \tag{3.2}$$

for $t = 2, \ldots, T$ and $f_0 = 0$. We work with sample sizes $T \in \{100, 400, 1000\}$. Here, $\boldsymbol{\epsilon}_t = (\epsilon_{1,t}, \epsilon_{1,t})'$ is an i.i.d. Gaussian process with zero mean and identity covariance matrix. Furthermore,

$$\boldsymbol{\Omega}_t = \begin{bmatrix} \sqrt{h_{1,t}} & 0 \\ 0 & \sqrt{h_{2,t}} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sqrt{h_{1,t}} & 0 \\ 0 & \sqrt{h_{2,t}} \end{bmatrix} \tag{3.3}$$

where $h_{i,t} = \gamma_{0,i} + \gamma_{1,i}\epsilon_{i,t-1}^2$ for $i = 1, 2$. Consider now using f_{t-1} as a forecast for s_t.

(To fix ideas, s_t may denote the logarithm of the weekly spot rate of a given currency (relative to USD, say) and f_t may denote the corresponding forward rate formed at time $t-1$ for time t. Hence we adopt an ARCH in mean (ARCH-M) type specification for the series of excess returns $s_t - f_{t-1}$. Following Engle et al. (1987), the specification in (3.2) implies that the excess return is proportional to the conditional standard deviation of the spot rate.[3])

Now, the series $u_t = s_t - f_{t-1}$ is predictable in this framework due to the presence of the time-varying component $\delta\sqrt{h_{1,t}}$, so using f_{t-1} as a forecast for s_t is irrational under MSE loss. The f_{t-1} forecast may however be rational under a specific asymmetric loss function. By Theorem 1 in Patton and Timmermann (2007a), the optimal point forecast of s_t in this model is given under the loss function in (3.1) by $f_{t-1} + \delta\sqrt{h_{1,t}} + C\sqrt{h_{1,t}}$, where C is a constant that depends only on the distribution of the idiosyncratic error and the loss function $\mathcal{L}$. The optimal point forecast will thus be zero if $\delta = -C$. Under normality and given that $\delta = 0.5$, we can then select the asymmetry parameter α consistent with $u_t = s_{t-1} - f_t$ being unforecastable under an asymmetric power loss function. This value is $\alpha \approx 0.30854$ for $p = 1$, and $\alpha \approx 0.22066$ for $p = 2$.[4]

[3]See also Baillie and Bollerslev (2000) who show that in a consumption-based asset pricing model, the discrepancy between the (expected) spot and forward rate is a function of the conditional variance deviation of the spot rate.

[4]The constant C must satisfy $\mathrm{E}\left(\mathcal{L}'\left(\epsilon_{1,t} - C\right)\right) = 0$; see Granger (1999); Patton and Timmermann

The focus is on inference of the asymmetry parameter in this simplified framework using the test statistics $\mathcal{J}_{\hat{\alpha}}$, $\mathcal{J}_{\alpha_0}$ and $\mathcal{T}$, based on forecast errors $u_t = s_t - f_{t-1}$. To estimate the loss function parameter, we employ the GMM estimator described above and choose the instruments as

$$(v_{0,t-1}, 1, v_{1,t-1})' = (u_{t-1}, 1, f_{t-1})' \quad \text{for } t = 2, \ldots, T.$$

The instrument f_{t-1} is highly persistent[5] while the lagged forecast error u_{t-1} is stationary.

The ARCH parameters are given by $\gamma_{0,2} = \gamma_{0,1} = 0.01$, $\gamma_{1,2} = 0$, and $\delta = 0.5$. A nonzero correlation ρ translates in dependence between the innovations of the predictor and the forecast errors u_t, thus avoiding the conditions of Corollary 3.1, so we set the (constant) conditional correlation $\rho = 0.8$; Table 3.1 in the following section indicates that such correlations are not uncommon in real data. Furthermore, we consider the cases $\gamma_{1,1} = 0.95$ and $\gamma_{1,1} = 0$ separately. The case $\gamma_{1,1} = 0.95$ exhibits conditional heteroskedasticity, while $\gamma_{1,1} = 0$ does not, so we expect different behavior of $\mathcal{T}$ according to Corollary 3.2. The $\mathcal{J}$-statistics should behave differently under stationary and persistent instruments, irrespective of conditional heteroskedasticity.

All simulations have been performed in R, version 3.5.1, using RStudio (1.1.453) and the packages expm (0.999-2), ggplot2 (3.0.0) and readr (1.1.1). We draw 25,000 samples from the above DGP and estimate the loss function parameter $\hat{\alpha}$ and the variance parameter $\hat{V}$. We study three estimators: the first uses a constant and a lagged forward rate, the second resorts to a constant and a lagged forecast error, while the third estimator uses all three instruments.

We report here results for the class of asymmetric linear loss functions, i.e. $p = 1$.

The behavior of the $\mathcal{J}$-statistics is in line with the provided asymptotics. From Figure 3.1 we learn that the three-instrument case (with both the stationary and the persistent instrument) exhibits serious departures from the χ^2 limit derived by Elliott et al. (2005) under stationarity, leading to spurious rejections of the null. When leaving out the stationary instrument, the distribution is still distorted, even if less so than for the case with three instruments. The χ^2 distribution is only appropriate when no persistent instrument is used. Regarding the magnitude of the distortions, there is no obvious difference between the cases without, and with, conditional heteroskedasticity. Figure 3.2 plots the densities of the null distributions of the statistics $\mathcal{J}_{\alpha_0}$. Again, the use of persistent instruments shifts the distribution to the right leading to overrejections if using χ^2 critical values.[6]

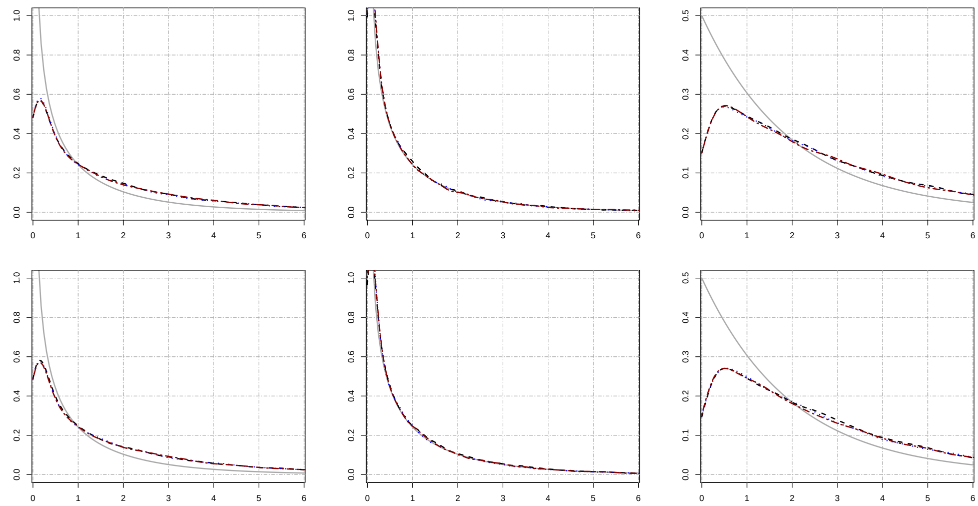
The results for asymmetric quadratic losses, $p = 2$, are virtually the same for the $\mathcal{J}$-statistics; for the precise results, see the Appendix. The only difference appears to be the somewhat slower convergence towards the respective limiting distribution; in any case, χ^2 critical values should

(2007a). From this condition and the normality assumption, we obtain $\Phi(C) = \alpha$ for $p = 1$ and $\Phi(C)\left(\mathrm{E}\left(\epsilon_{1,t}\middle|\epsilon_{1,t} < C\right)/C - 1\right) = \alpha/(1-2\alpha)$ with $\mathrm{E}\left(\epsilon_{1,t}\middle|\epsilon_{1,t} < C\right) = \frac{1}{\Phi(C)}\int_{-\infty}^{C} x\phi(x)\,\mathrm{d}x$ for $p = 2$, where ϕ and Φ are the std. normal pdf and cdf.

[5] This is in line with empirical evidence for the forward rate; see e.g. Liu and Maynard (2005) or Gospodinov (2009).

[6] Note that the plots seem to suggest a mismatch in the left tail between the simulated densities and the χ^2 distribution for the case of a stationary instrument as well, but this is an artifact of the kernel density estimator used for smoothing, which suffers from boundary bias at the origin.

Figure 3.1: Densities of $J_{\hat{\alpha}}$ under asymmetric linear loss and various instrument choices



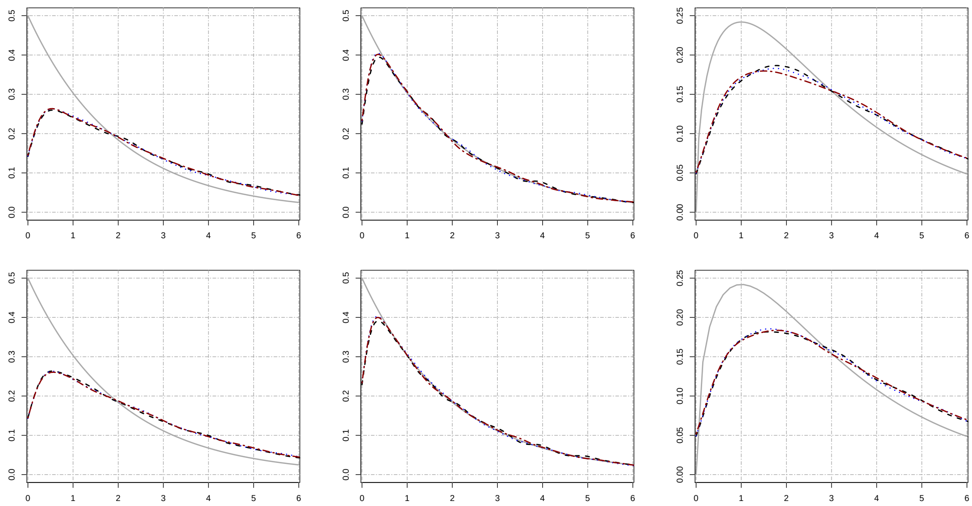
Notes: The plots show the density of χ^2 (solid grey) and kernel density estimates of the distribution of $J_{\hat{\alpha}}$ when $T = 100$ (dashed black), $T = 400$ (dotted blue) and $T = 1000$ (dot-dashed red). The instrument combinations are: a constant and a persistent instrument (left) and a constant a stationary instrument (middle), where the plotted χ^2 distribution has 1 d.o.f., as well as all three instruments (right), where the plotted χ^2 distribution has 2 d.o.f.. The DGP is given by (3.2) - (3.3) with $\delta = 0.5$, $\rho = 0.8$, $\gamma_{0,1} = \gamma_{0,2} = 0.01$, and $\gamma_{1,2} = 0$, with either conditional homoskedasticity of $\tilde{u}_t$ ($\gamma_{1,1} = 0$, top), or conditional heteroskedasticity ($\gamma_{1,1} = 0.95$, bottom).

not be used for testing with the $\mathcal{J}$-statistics under instrument persistence.

Figure 3.3 plots the finite-sample densities of $\mathcal{T}$, together with the standard normal benchmark. We observe for $p = 1$ a remarkable robustness of the $\mathcal{T}$-statistic to instrument persistence, even when this is not expected (one instrument of each kind, conditional heteroskedasticity). This is not the case for $p = 2$, however; see the Appendix. There, we note for the homoskedastic case ($\gamma_{1,1} = 0$) that convergence to the standard normal appears to occur in all cases (even if at a slightly lower pace for the case of three instruments). This is consistent with the statement of Corollary 3.2. At the same time, under conditional heteroskedasticity, the use of three instruments leads to larger departures from the standard normal, as expected from Proposition 3.1 in general. For the cases with two instruments, there is a small visible difference between the conditionally homoskedastic and conditionally heteroskedastic cases; at the same time, it makes no difference whether the non-constant instrument is stationary or not, as predicted by Corollary 3.3. Also, even in the asymptotically normal cases, the approximation of the finite-sample distribution by the standard normal limit is not ideal.[7] This effect is of similar magnitude as the persistence-induced distortions.

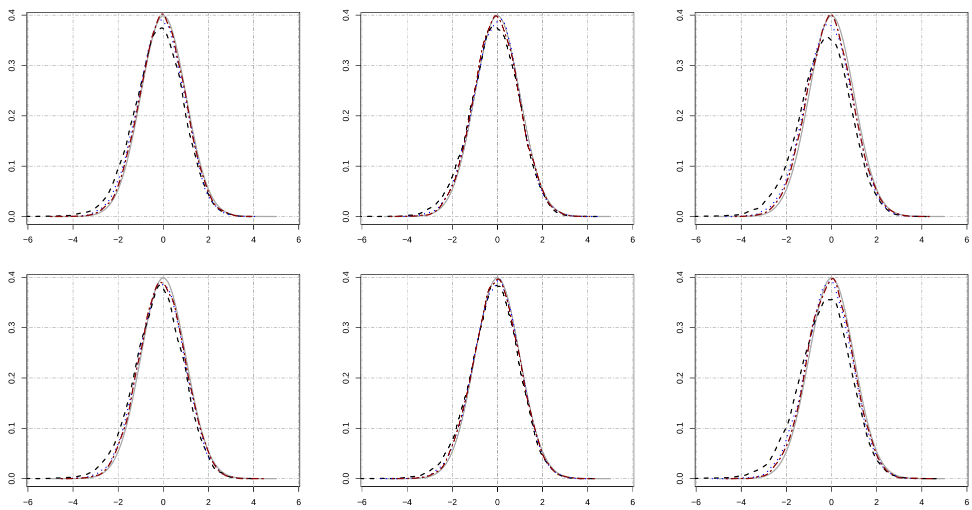
[7]The observed skewness is partly due to the iterative nature of $\hat{\alpha}$; when testing, using the null value α_0 reduces it.

Figure 3.2: Densities of J_{α_0} under asymmetric linear loss and various instrument choices



Notes: The plots show the density of χ^2 (solid grey) and kernel density estimates of the distribution of J_{α_0} when $T = 100$ (dashed black), $T = 400$ (dotted blue) and $T = 1000$ (dot-dashed red). The instrument combinations are: a constant and a persistent instrument (left) and a constant a stationary instrument (middle), where the plotted χ^2 distribution has 2 d.o.f., as well as all three instruments (right), where the plotted χ^2 distribution has 3 d.o.f.. Top: conditional homoskedasticity, bottom: conditional heteroskedasticity of $\tilde{u}_t$. See Fig. 3.1 for details.

Figure 3.3: Densities of $\mathcal{T}$ under asymmetric linear loss and various instrument choices



Notes: The plots show the density of the standard normal distribution (solid grey) and kernel density estimates of the distribution of $\mathcal{T}$ when $T = 100$ (dashed black), $T = 400$ (dotted blue) and $T = 1000$ (dot-dashed red). The used instrument combinations are: a constant and a persistent instrument (left) and a constant and a stationary instrument (middle), as well as a constant, a persistent and a stationary instrument (right). Top: conditional homoskedasticity, bottom: conditional heteroskedasticity of $\tilde{u}_t$. See Fig. 3.1 for details.

3.2.5 Recommendations for practitioners

To sum up, the $\mathcal{J}$-statistics are not reliable as soon as at least one instrument is persistent. Unless this case can be excluded, it is not recommendable to rely on $\mathcal{J}$-statistics for inference.

The situation is more finely nuanced for the $\mathcal{T}$-statistic. For the case $p = 2$, some care needs to be taken to ensure correctly-sized inference. First, practitioners could conduct a test of the null hypothesis of constant conditional moments of $\tilde{u}_t$ (e.g. in the spirit of Bierens, 1982, but a parametric test for no ARCH effects might also be used) to be able to justify χ^2 critical values via Corollary 3.2. Alternatively, the set of instruments may be split to separate weakly from strongly persistent instrument and run separate tests for the two sets of instruments. This would allow to exploit Corollary 3.3 without worrying about the conditional homoskedasticity requirement. For the case of linear asymmetric losses (i.e., $p = 1$), however, the $\mathcal{T}$-statistic appears to be quite robust to persistence, so the care taken for $p = 2$ is, by and large, unnecessary. In our replication of the analysis of the loss preferences of the EU Commission, we take the $\mathcal{T}$-ratio at face value for $p = 1$, but otherwise distinguish between persistent and stationary instruments to prevent spurious findings.

3.3 Re-assessing the EU Commission forecasts

This section reexamines the European Commission forecast data first analyzed by Christodoulakis and Mamatzakis (2009). They examine forecasts of five macroeconomic variables for 12 European Member States, namely inflation, unemployment, government balance, investment and current account, over the period 1970 – 2004. Each variable is predicted twice a year: in the spring for the current year, and in the autumn for the upcoming year. The number of observations for each country varies from 35 to 18 depending on the year of entering the EU (e.g. 1986 for Spain and Portugal). For this replication study, we augment the initial data set with more recent data, up to 2016,[8] which yields an increased number of observations between 31 and 48 data points. In order to encompass time evolution of asymmetries, we perform the evaluation for the original time span, for the full time span and for the last available 20 years only, i.e. 1997 – 2016.

Christodoulakis and Mamatzakis (2009) reported for the estimation under asymmetric linear loss ($p = 1$) with $D = 3$ instruments: a constant, a lagged realization and a lagged forecast error. The stationarity assumptions required by Elliott et al. (2005) may however be overly optimistic for the data at hand. Table 3.1 gives the sums of estimated coefficients of autoregressive [AR] processes fitted to the realized values of the variables to be forecast, as well as correlation between the AR residuals and the forecast errors (The model order p is selected individually using the Akaike information criterion). Quite often, the cumulated AR coefficients are seen to be close to, or even above, unity.[9] Moreover, the contemporaneous correlation between the forecast errors and the innovations to the instruments is strongly positive: for the year-ahead forecasts, the correlation ranges in a neighbourhood of 0.8, and is somewhat reduced, to about 0.5, for the

[8] The data were obtained from `http://ec.europa.eu/economy_finance/db_indicators/statistical_annex/index_en.htm`.

[9] For Spanish inflation we notice an extreme sum of about 3, which is explained by the specific pattern of this series, with high levels of inflation at the time Spain joined the European Community and a dominating downward trend in the years to follow. Also, investment series have a pronounced antipersistent behavior for some countries, e.g. for Belgium.

Table 3.1: Instrument persistence: Stylized facts, 1970 – 2016.

		Inflation		Unemployment		Gov. balance		Investment		Current account	
		Current	Ahead	Current	Ahead	Current	Ahead	Current	Ahead	Current	Ahead
Belgium	Σ	0.781	0.793	0.778	0.747	0.284	0.795	-1.437	-0.833	0.564	0.621
	r	0.292	0.771	0.336	0.183	0.573	0.782	0.237	0.561	0.585	0.364
Denmark	Σ	1.038	1.017	0.680	0.555	0.654	0.665	-1.026	-0.605	1.065	1.182
	r	0.339	0.596	0.368	0.407	0.545	0.501	0.558	0.683	0.563	0.537
Germany	Σ	0.811	0.986	0.888	0.921	0.775	0.673	-0.092	-0.120	0.901	0.924
	r	0.434	0.880	0.380	0.203	0.365	0.682	0.589	0.919	0.570	0.801
Greece	Σ	1.037	1.009	0.761	0.926	0.848	0.621	-0.323	0.435	0.774	0.381
	r	0.358	0.437	0.212	0.538	0.423	0.597	0.350	0.519	0.596	0.547
Spain	Σ	2.968	3.411	0.835	0.778	–	0.736	0.361	-4.075	0.790	0.677
	r	0.331	0.039	0.087	0.555	0.688	0.294	-0.130	0.147	0.309	0.668
France	Σ	0.989	0.948	0.910	0.915	0.240	0.724	-0.416	0.043	0.686	0.824
	r	0.339	0.796	0.592	0.461	0.700	0.847	0.466	0.741	0.720	0.792
Ireland	Σ	1.044	0.964	0.893	0.909	0.793	0.729	-1.283	0.465	1.000	0.725
	r	0.471	0.807	0.533	0.559	0.486	0.804	0.575	0.623	0.403	0.590
Italy	Σ	0.979	0.974	0.835	0.882	0.780	0.779	–	–	-0.061	0.440
	r	0.332	0.662	0.534	0.505	-0.015	0.015	0.600	0.851	0.575	0.450
Luxembourg	Σ	0.754	0.890	0.987	0.950	0.666	0.652	–	–	0.783	0.842
	r	0.307	0.597	0.728	0.711	0.402	0.709	0.802	0.934	0.193	0.142
Netherlands	Σ	0.969	1.068	0.574	0.659	0.075	0.242	-1.195	–	0.979	0.847
	r	0.392	0.474	0.417	0.266	0.346	0.656	0.585	0.760	0.554	0.787
Portugal	Σ	0.976	1.206	0.735	0.686	0.578	0.100	-0.691	-0.999	-0.063	0.703
	r	0.367	0.335	0.151	0.335	0.072	0.594	0.119	0.150	0.415	0.380
UK	Σ	0.975	0.942	0.792	0.831	0.720	0.729	–	0.301	0.805	0.701
	r	0.701	0.406	0.353	0.496	0.605	0.563	0.676	0.926	0.080	0.619
EU	Σ	1.008	1.005	0.911	0.894	0.726	0.610	-1.101	-0.043	0.578	0.539
	r	0.577	0.808	0.464	0.397	0.715	0.863	0.287	0.874	0.511	0.662

Notes: Σ denotes the sums of OLS AR(p) coefficient estimates; r denotes the contemporaneous correlation between the AR(p) residuals and the forecast errors. The order p was selected via AIC. Missing values indicate $p = 0$.

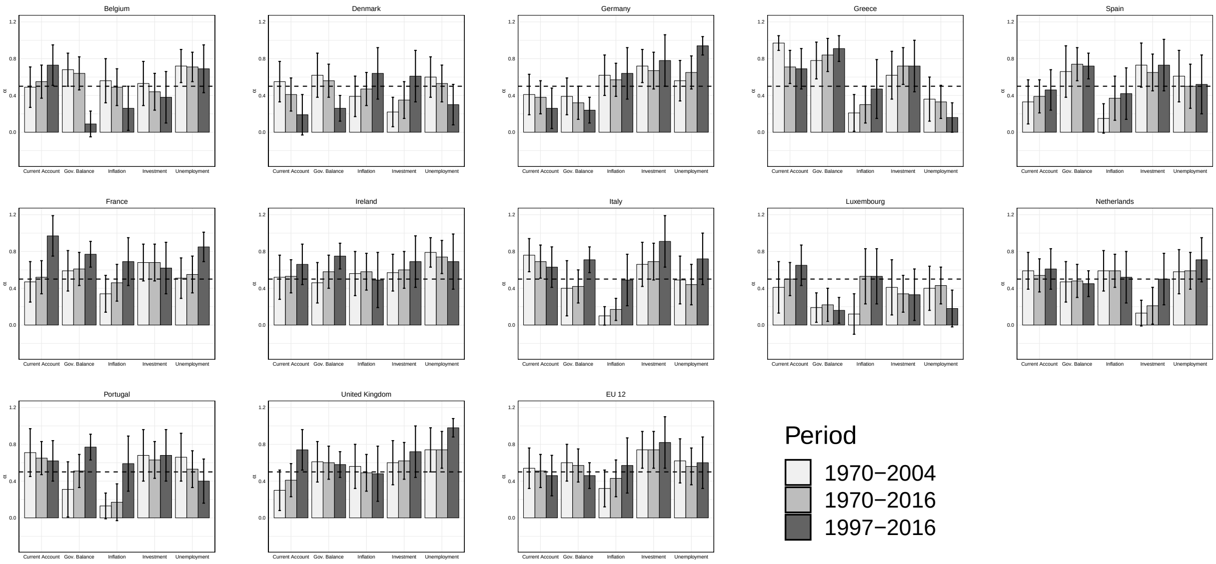
current-year forecasts.

We first replicate the study of Christodoulakis and Mamatzakis (2009). We use the dataset and the Matlab estimation routine of Christodoulakis and Mamatzakis.[10] The results of the replication are given in Tables 3.2 – 3.6 of the Appendix (where we also report values for the $\mathcal{T}$-statistic testing the null $\alpha = 0.5$ as it is robust). Our computations lead to somewhat different figures. In particular, the three $\mathcal{J}_{\alpha_0}$ statistics are different by about 5 to 10% for most variables and both the current-year and the year-ahead forecasts. For government balance, some of the point estimates are also different, although the differences do not change the overall picture. As a robustness check, we conducted this replication in R, obtaining the same numbers as in Tables 3.2 – 3.6. The differences are likely due to different Matlab versions. We used Matlab 6.1, as later versions did not run the codes without modifications, but it is not clear which version was used in the original study. It should be noted at this point, that certain cases display convergence problems when $D = 3$. In particular, estimation for inflation, unemployment and government balance sometimes yields values of $\alpha \notin (0, 1)$ or fails to converge all together due to singularity of $\hat{\mathbf{S}}$.

We now focus on the time evolution of the estimates and report outcomes for the three different periods we consider. We use several sets of instruments: additionally to the original set of $D = 3$

[10]The data from the original study, as well as the codes, were kindly provided by the authors on `http://qed.econ.queensu.ca/jae/2009-v24.4/christodoulakis-mamatzakis/`.

Figure 3.4: Asymmetry estimates for the EU–12 Member States over different time spans, linear asymmetric loss and $D = 3$ instruments (current year forecasts)



Notes: The figure shows α estimates with error bars (± 2 standard errors). The estimation was conducted using three instruments: a constant, a lagged forecast error and a lagged realization; $p = 1$.

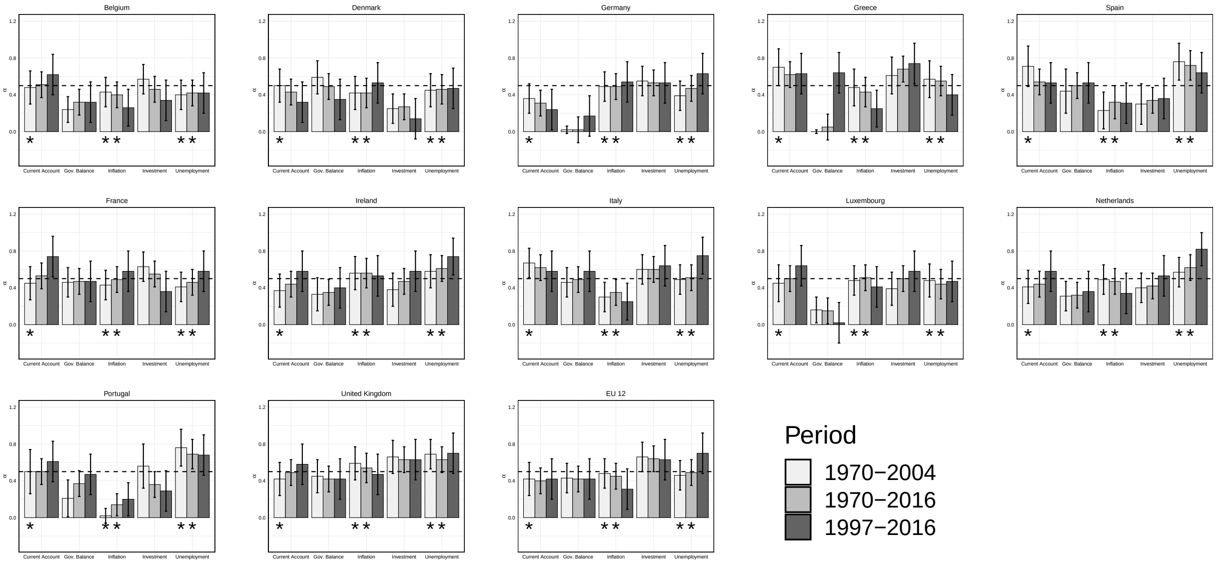
instruments, we resort to two sets of $D = 2$ instruments: constant and lagged realization, as well as constant and lagged forecast error. A nice side effect of using $D = 2$ instruments is that singularity issues seem to be alleviated; the main reason to conduct the estimation with $D = 2$ is that we may argue a) that the estimates resorting to a constant and the lagged realization (which tends to be persistent) lead to a robust $\mathcal{T}$-statistic, while b) the estimates building on a constant and the lagged forecast error likely fulfill stationarity assumptions and therefore deliver interpretable results for the $\mathcal{J}$-statistics. Also, to ensure numerical convergence of the estimator, we use a different starting value for iterative calculation of the matrix $\hat{\mathbf{S}}$: Christodoulakis and Mamatzakis (2009), following Elliott et al. (2005), start their iterations by choosing the identity matrix; we calculate the starting value by plugging in $\alpha = 0.5$, which improves the convergence behavior of the estimation procedure.

Figures 3.4 – 3.5 give point estimates $\hat{\alpha}$ together with confidence intervals based on standard normality of the $\mathcal{T}$-ratio for the current year forecasts.

The results for the $\mathcal{J}$-statistic $\mathcal{J}_{\hat{\alpha}}$ testing for rationality under possible loss asymmetry are only given in Figure 3.5, which builds on stationary instruments and thus allow for χ^2 inference following Elliott et al. (2005). Nicely confirming Corollary 3.3, the results for the case $D = 2$ with the nonstationary instrument only are almost identical to those for $D = 2$ using the stationary instrument, so we only report them in the Appendix.

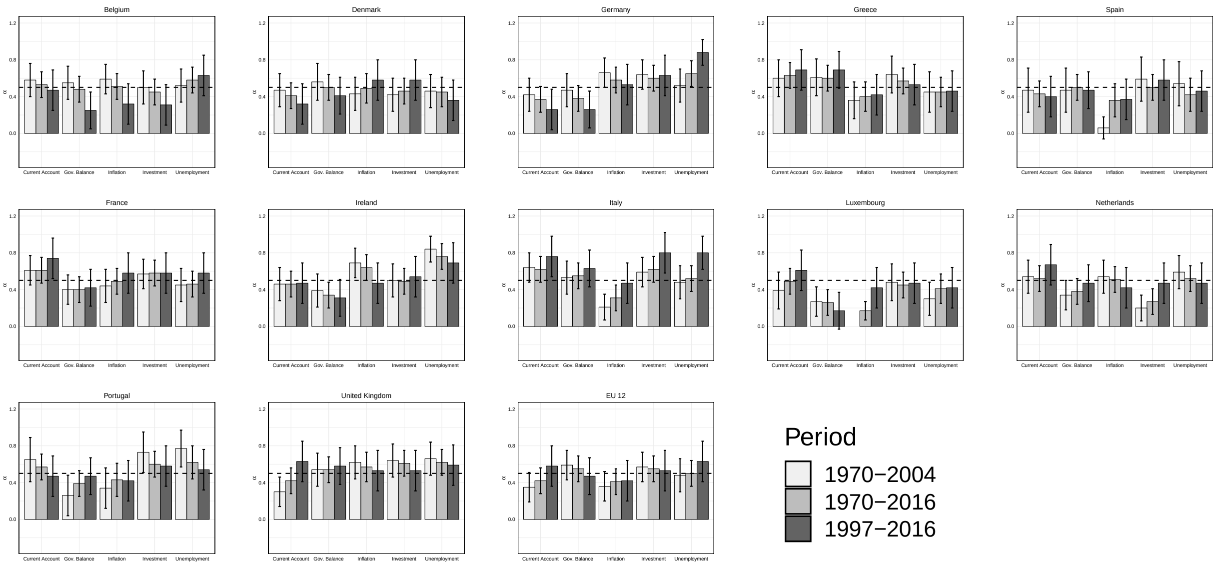
Although the results are mostly similar for the three sets, we do notice that there are some distinctions between Figure 3.4 and Figures 3.5 and 3.11. In particular for the current account, the estimator with $D = 3$ delivers estimates that may be seen as a bit too extreme. Given that robustness is not given for $D = 3$, one should prefer the latter. The analogous findings for the year ahead forecasts are given in Figures 3.6 – 3.7.

Figure 3.5: Asymmetry estimates for the EU–12 Member States over different time spans, linear asymmetric loss and $D = 2$ instruments (current year forecasts)



Notes: The figure shows α estimates with error bars (± 2 standard errors). The estimation was conducted using two instruments: a constant and a lagged forecast error; $p = 1$. Asterisks represent significance at 5% level.

Figure 3.6: Asymmetry estimates for the EU–12 Member States over different time spans, linear asymmetric loss and $D = 3$ instruments (year ahead forecasts)



Notes: The figure shows α estimates with error bars (± 2 standard errors). The estimation was conducted using three instruments: a constant, a lagged forecast error and a lagged realization; $p = 1$. Missing bars indicate failed convergence of the estimation algorithm.

To interpret the results, we follow Christodoulakis and Mamatzakis (2009). They observe optimistic tendencies in year-ahead forecasts of the EU Commission and more prudent ones for the current year forecasts. This is reflected in the significant differences from 0.5 of the estimated α. We resort to $\mathcal{T}$-statistics rather than $\mathcal{J}$-statistics to assess significance, as the $\mathcal{J}$-statistics are not robust to persistence. Generally, a value above 0.5 represents a higher preference of the forecaster for over-prediction, since negative forecast errors are deemed more costly. Whether it can be interpreted as a prudent attitude, depends on the variable in question. For instance, variables such as inflation and unemployment a pessimistic forecaster would have a tendency to over-forecast, expecting a worse outcome than the one actually occurring. The same holds for government balance that yields negative values in case of a deficit. Here, underestimating the deficit is more costly in case of a prudent strategy. As for investment and current account, the picture is reversed. Optimism corresponds to higher values of α, i.e. over-prediction. Rationality is discussed only in the case where the set of instruments contains a constant and a lagged forecast error (see Figures 3.5 and 3.7); in other cases we cannot rely on the $\mathcal{J}_{\hat{\alpha}}$ statistic to distinguish between asymmetry and irrationality.

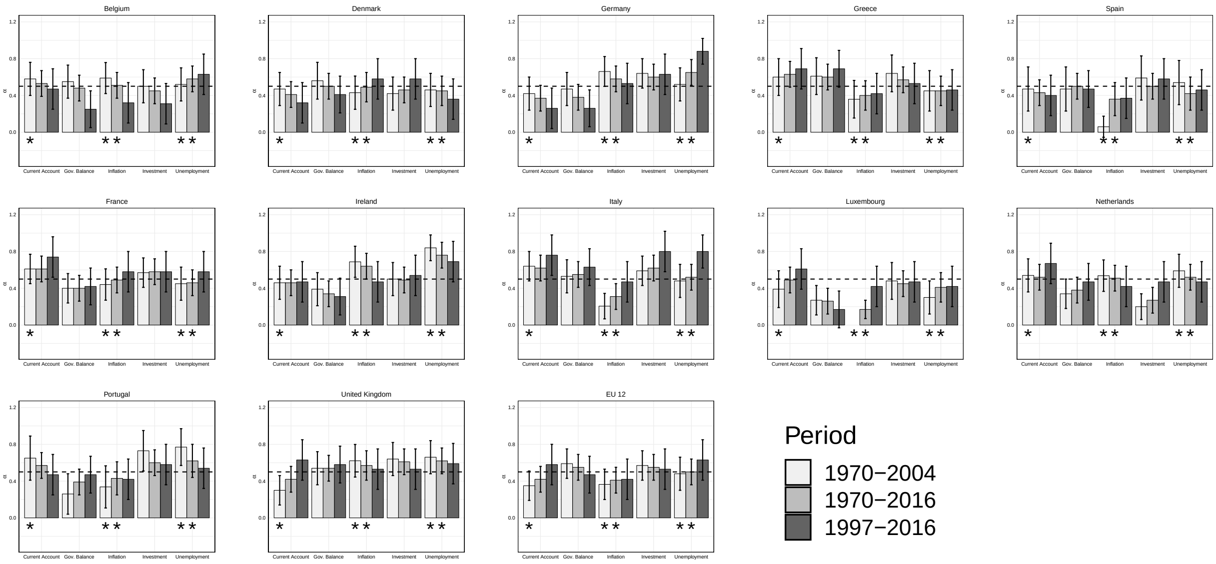
Forecasts appear to have different properties in time. For both year-ahead and current-year inflation forecasts the hypothesis of rationality can be rejected in the first two samples (original and augmented). We observe no significant deviations from symmetry for year-ahead forecasts in the last twenty years, though. For the current year we detect optimistic preferences for Belgium, Greece, Italy and Portugal. Similarly, the hypothesis of rationality can be rejected for current-year as well as year-ahead unemployment forecasts in the first two samples. As for the last sample, we only find evidence of prudence for Ireland, Italy and Netherlands in the current year, and Germany and Italy in the year ahead (Figures 3.5 and 3.7).

For government balance we observe no deviations from rationality. For the current year there appears to be an optimistic tendency for Germany and Luxembourg. Forecasts for Belgium, Greece, Netherlands and Portugal are significantly optimistic in the beginning with a tendency towards symmetry in the last twenty years. As for the year-ahead forecasts, preferences for Luxembourg remain overall optimistic, but for Belgium, Ireland and Germany they shift to optimism in the last sample. For Portugal, however, there is a shift is from optimism to symmetry over time.

Investment was highlighted by Christodoulakis and Mamatzakis (2009) as being forecast overly optimistic. Here we don't see any deviations from rationality, but results for some countries exhibit strong asymmetric preferences. In particular for the current-year forecasts, cases of Greece, UK, Portugal and the EU average show optimistic tendencies, while Netherlands and Denmark deliver evidence for a more prudent forecast strategy. The year-ahead results reveal optimistic preferences for Italy and Portugal, and pessimistic ones for the Netherlands.

For the current account we see rejection of rationality in the original sample for both forecast horizons. Otherwise the evidence from the augmented sample and the last twenty years supports rational and symmetric preferences of the Commission with few minor exceptions. The notable ones are France and Italy for which the forecasts are more optimistic, and Germany, with a tendency towards prudence.

Figure 3.7: Asymmetry estimates for the EU–12 Member States over different time spans, linear asymmetric loss and $D = 2$ instruments (year ahead forecasts)



Notes: The figure shows α estimates with error bars (± 2 standard errors). The estimation was conducted using two instruments: a constant and a lagged forecast error; $p = 1$. Asterisks represent significance at 5% level. Missing bars indicate failed convergence of the estimation algorithm.

Finally, the results for $p = 2$ are qualitatively very similar (see Appendix, Figures 3.13 – 3.14) and we do not discuss them in detail. (We only provide results for $D = 2$ with stationary instruments and therefore robust $\mathcal{J}$-statistics to save space.)

3.4 Summary

The note replicated and extended the study of Christodoulakis and Mamatzakis (2009) on the loss preferences of EU commission forecasts for several member states and key economic indicators.

To this end, we investigate the distribution of the estimator of the asymmetry parameter in the GMM framework of Elliott et al. (2005). We focus on the empirically relevant case when some of the instruments are persistent. In theoretical derivations and Monte Carlo experiments, we found that $\mathcal{J}$-statistics tend to overreject their respective null hypotheses whenever at least one instrument is persistent. For the $\mathcal{T}$-statistic, we provide theoretical arguments that robustness to instrument persistence is given in several practically relevant cases; moreover, Monte Carlo evidence for the case of the asymmetric linear loss suggests that the $\mathcal{T}$-statistic is not really affected by persistence when $p = 1$, even in theoretically not so clear cut situations. All in all, we recommend the use of the $\mathcal{T}$-statistic for inference and for all types of instruments. The use of $\mathcal{J}$-statistics for inference is however only recommended after persistent instruments have been eliminated from the set of instruments.

We then find that the original conclusions of Christodoulakis and Mamatzakis (2009) are largely confirmed for the extended data. However, the departures from symmetry appear to have somewhat decreased compared to the original (shorter) data set, and rationality is rejected less often, such that EU commission forecasts could be seen as increasedly reliable.

Appendix

Proofs

Before proving the main result, we state and prove a useful lemma.

Lemma 3.1 *For any strictly stationary ergodic process z_t, uniformly $L_{1+\delta}$-bounded for some $\delta > 0$ and any and v_t strongly persistent in the sense of Assumption 3.1, it holds that*

$$\frac{1}{Tn_T}\sum_{t=2}^{T} v_{t-1}z_t \xrightarrow{d} \mathrm{E}\left(z_t\right)\int_0^1 X\left(s\right)\mathrm{d}s.$$

Note that the lemma, applied elementwise, implies under the assumptions of Proposition 3.1 that

1. $\frac{1}{T}\mathbf{N}_T^{-1}\sum_{t=2}^{T}\boldsymbol{v}_{1,t-1}\left|u_t\right|^{p-1} \xrightarrow{d} \mathrm{E}\left(\left|u_t\right|^{p-1}\right)\int_0^1 \boldsymbol{X}\left(s\right)\mathrm{d}s$
2. $\frac{1}{T}\mathbf{N}_T^{-1}\sum \boldsymbol{v}_{0,t-1}\boldsymbol{v}_{1,t-1}'\tilde{u}_t^2 \xrightarrow{p} \gamma_0\int_0^1 \boldsymbol{X}'\left(s\right)\mathrm{d}s$
3. $\frac{1}{T}\mathbf{N}_T^{-1}\sum_{t=2}^{T}\boldsymbol{v}_{1,t-1}\tilde{u}_t^2 \xrightarrow{d} \sigma_{\tilde{u}}^2\int_0^1 \boldsymbol{X}\left(s\right)\mathrm{d}s$
4. $\frac{1}{T}\mathbf{N}_T^{-1}\left(\sum_{t=2}^{T}\boldsymbol{v}_{1,t-1}\boldsymbol{v}_{1,t-1}'\tilde{u}_t^2\right)N_T^{-1} \xrightarrow{d} \sigma_{\tilde{u}}^2\int_0^1 \boldsymbol{X}\left(s\right)\boldsymbol{X}'\left(s\right)\mathrm{d}s.$

Proof of Lemma 3.1

Write

$$\frac{1}{Tn_T}\sum_{t=1}^{T-1} v_t z_t = \frac{1}{Tn_T}\sum_{t=1}^{T-1} v_t\left(z_t - \mathrm{E}\left(z_t\right)\right) + \mathrm{E}\left(z_t\right)\frac{1}{Tn_T}\sum_{t=1}^{T-1} v_t.$$

Should the first term vanish as $T\to\infty$, the desired result follows directly from Assumption **??** with the continuous mapping theorem. Let then $\tilde{z}_t = z_t - \mathrm{E}\left(z_t\right)$ and note that, since z_t is ergodic, $\mathrm{E}\left(\tilde{z}_t|\tilde{z}_{t-m},\tilde{z}_{t-m-1},\ldots\right)\xrightarrow{p} 0$ as $m\to\infty$. Furthermore, since $\tilde{z}_t$ is uniformly $L_{1+\delta}$-bounded for some $\delta > 0$, it is uniformly integrable and thus $\mathrm{E}\left(\left|\mathrm{E}\left(\tilde{z}_t|\tilde{z}_{t-m},\tilde{z}_{t-m-1},\ldots\right)\right|\right)\to 0$. Then, Theorem 3.3 of Hansen (1992) applies, such that, as required, $\left|\frac{1}{Tn_T}\sum_{t=1}^{T-1} v_t\tilde{z}_t\right| \leq \sup_{s\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{[sT]}\frac{v_t}{n_T}\tilde{z}_t\right|\xrightarrow{p} 0.$

Proof of Proposition 3.1

It holds that $\sqrt{T}\left(\hat{\alpha}-\alpha\right) = \frac{1}{p}\frac{\hat{\boldsymbol{h}}'\hat{\mathbf{S}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\tilde{u}_t\right)}{\hat{\boldsymbol{h}}'\hat{\mathbf{S}}^{-1}\hat{\boldsymbol{h}}}$ and $t_{\hat{\alpha}} = \frac{1}{p}\frac{\hat{\boldsymbol{h}}'\hat{\mathbf{S}}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\tilde{u}_t\right)}{\sqrt{\hat{\boldsymbol{h}}'\hat{\mathbf{S}}^{-1}\hat{\boldsymbol{h}}}}$. We have, regularity conditions assumed, that

$$\mathbf{D}_T^{-1}\hat{\boldsymbol{h}} \Rightarrow \left(\begin{array}{c}\mathrm{E}\left(\boldsymbol{v}_{0,t-1}\left|u_t\right|^{p-1}\right)\\ \mathrm{E}\left(\left|u_t\right|^{p-1}\right)\\ \mathrm{E}\left(\left|u_t\right|^{p-1}\right)\int_0^1 \boldsymbol{X}\left(s\right)\mathrm{d}s\end{array}\right) := \boldsymbol{H}\quad\text{with}\quad \mathbf{D}_T = \left(\begin{array}{ccc}\mathbf{I} & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & \mathbf{N}_T\end{array}\right).$$

Since $\hat{\mathbf{S}} = \frac{1}{p^2}\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\boldsymbol{v}_{t-1}'\tilde{u}_t^2$, we also have that

$$\mathbf{D}_T^{-1}\hat{\mathbf{S}}\mathbf{D}_T^{-1} \Rightarrow \frac{1}{p^2}\begin{pmatrix} \bar{\boldsymbol{\Omega}}_0 & \boldsymbol{\gamma}_0 & \boldsymbol{\gamma}_0 \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \boldsymbol{\gamma}_0' & \sigma_{\tilde{u}}^2 & \sigma_{\tilde{u}}^2 \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \left(\int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s\right)\boldsymbol{\gamma}_0' & \sigma_{\tilde{u}}^2 \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s & \sigma_{\tilde{u}}^2 \int_0^1 \boldsymbol{X}(s)\boldsymbol{X}'(s)\,\mathrm{d}s \end{pmatrix}.$$

Finally,

$$\mathbf{D}_T^{-1}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\boldsymbol{v}_{t-1}\tilde{u}_t \Rightarrow \begin{pmatrix} \bar{\boldsymbol{W}}(1) \\ \tilde{W}(1) \\ \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}\tilde{W}(s) \end{pmatrix},$$

see Hansen (1992, Theorem 2.1), and the result follows with the continuous mapping theorem.

Proof of Corollary 3.2

From the assumptions of the corollary it follows that $\mathrm{E}\left(\boldsymbol{v}_{0,t-1}\,|u_t|^{p-1}\right) = \mathbf{0}$ just like $\boldsymbol{\gamma}_0$ and, with $\mathrm{E}\left(\tilde{u}_t^2\right) = \sigma_{\tilde{u}}^2$, $\bar{\boldsymbol{\Omega}}_0 = \sigma_{\tilde{u}}^2\,\mathrm{E}\left(\boldsymbol{v}_{0,t-1}\boldsymbol{v}_{0,t-1}'\right)$ one obtains

$$\begin{aligned} \boldsymbol{H}\mathbf{S}^{-1} &\equiv \mathrm{E}\left(|u_t|^{p-1}\right)\left(\mathbf{0}', 1, \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s\right) p^2 \begin{pmatrix} \frac{1}{\sigma_{\tilde{u}}^2}\mathrm{E}\left(\boldsymbol{v}_{0,t-1}\boldsymbol{v}_{0,t-1}'\right)^{-1} & \mathbf{0}' \\ \mathbf{0} & Q^{-1} \end{pmatrix} \\ &\equiv p^2\,\mathrm{E}\left(|u_t|^{p-1}\right)\left(\mathbf{0}', \left(1, \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s\right)\mathbf{Q}^{-1}\right) \end{aligned}$$

with $\mathbf{Q} = \sigma_{\tilde{u}}^2\begin{pmatrix} 1 & \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s \\ \int_0^1 \boldsymbol{X}(s)\,\mathrm{d}s & \int_0^1 \boldsymbol{X}(s)\boldsymbol{X}'(s)\,\mathrm{d}s \end{pmatrix}$. Now, $\left(1, \int_0^1 \boldsymbol{X}'(s)\,\mathrm{d}s\right)'$ is the first column of $\sigma_{\tilde{u}}^{-2}\mathbf{Q}$ so its transpose, postmultiplied with the inverse of $\mathbf{Q}$, gives $\sigma_{\tilde{u}}^{-2}\left(1, \mathbf{0}'\right)'$ where there are exactly as many zeros as elements of $\boldsymbol{v}_{1,t}$. Hence

$$\boldsymbol{H}'\mathbf{S}^{-1}\boldsymbol{U} = \frac{\mathrm{E}\left(|u_t|^{p-1}\right)}{p^2\sigma_{\tilde{u}}}\tilde{W}(1).$$

The same reasoning indicates that

$$\boldsymbol{H}'\mathbf{S}^{-1}\boldsymbol{H} = p^2\frac{\left(\mathrm{E}\left(|u_t|^{p-1}\right)\right)^2}{\sigma_{\tilde{u}}^2}$$

such that $\mathcal{T} \Rightarrow \mathcal{N}(0,1)$ whenever $\tilde{u}_t$ has constant conditional scale in the sense that the conditional expectation of the relevant powers of $|\tilde{u}_t|$ are constant.

To understand why the result hinges on the presence of the constant instrument, consider the simple bivariate case with one weakly and one strongly persistent instrument; let also the weakly persistent instrument have non-zero mean. Then

$$\begin{aligned} \boldsymbol{H}'\mathbf{S}^{-1} \equiv & \left(\mathrm{E}\left(v_{0,t-1} \left| u_t \right|^{p-1}\right); \quad \mathrm{E}\left(\left| u_t \right|^{p-1}\right) \int_0^1 X(s)\,\mathrm{d}s \right) \\ & \times p^2 \begin{pmatrix} \bar{\omega}_0 & \gamma_0 \int_0^1 X(s)\,\mathrm{d}s \\ \gamma_0 \int_0^1 X(s)\,\mathrm{d}s & \sigma_{\tilde{u}}^2 \int_0^1 X^2(s)\,\mathrm{d}s \end{pmatrix}^{-1} \end{aligned}$$

while

$$\boldsymbol{U} \equiv \begin{pmatrix} \int_0^1 \mathrm{d}\bar{W}(s) \\ \int_0^1 X(s)\,\mathrm{d}\tilde{W}(s) \end{pmatrix}.$$

The coefficient of $\int_0^1 X(s)\,\mathrm{d}\tilde{W}(s)$ in $\boldsymbol{H}'\mathbf{S}^{-1}\boldsymbol{U}$ should be zero for normality to be recovered in general. Some algebra indicates this to be the case when

$$-\mathrm{E}\left(v_{0,t-1} \left| u_t \right|^{p-1}\right) \gamma_0 \int_0^1 X(s)\,\mathrm{d}s + \mathrm{E}\left(\left| u_t \right|^{p-1}\right) \int_0^1 X(s)\,\mathrm{d}s\,\bar{\omega}_0 = 0,$$

or

$$\frac{\mathrm{E}\left(v_{0,t-1} \left|u_t\right|^{p-1}\right)}{\mathrm{E}\left(\left|u_t\right|^{p-1}\right)} = \frac{\mathrm{E}\left(v_{0,t-1}^2 \tilde{u}_t^2\right)}{\mathrm{E}\left(v_{0,t-1} \tilde{u}_t^2\right)}.$$

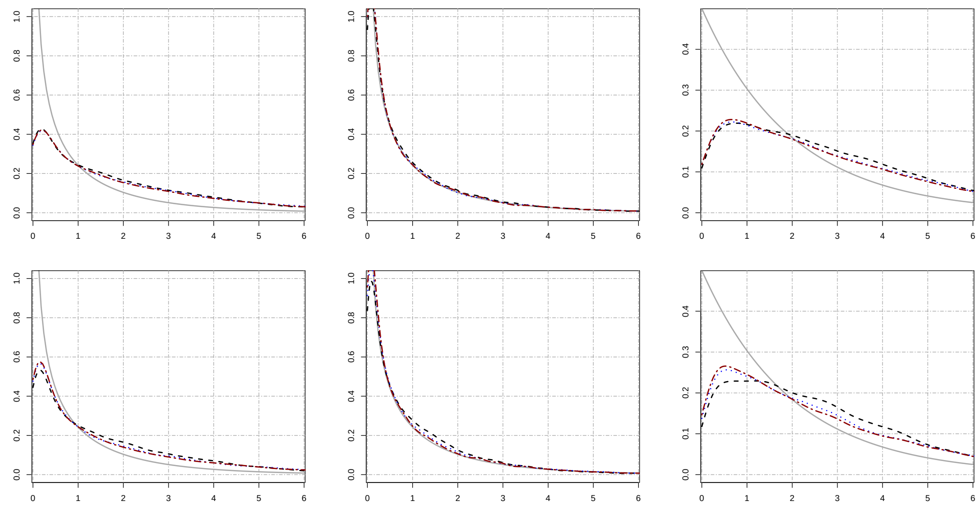
For constant conditional scale this reduces to $\mathrm{E}\left(v_{0,t-1}^2\right) = \mathrm{E}\left(v_{0,t-1}\right)^2$ and the weakly persistent instrument is constant w.p. 1.

Proof of Corollary 3.3

The result follows by noting that, without stationary instruments $\boldsymbol{v}_{0,t-1}$, $\boldsymbol{H}$ is proportional to the first row of $\mathbf{S}$, such that $\boldsymbol{H}'\mathbf{S}^{-1}$ is proportional to the first row of the identity matrix, which then cancels out all nonstandard terms in $\mathcal{T}$ and the result follows.

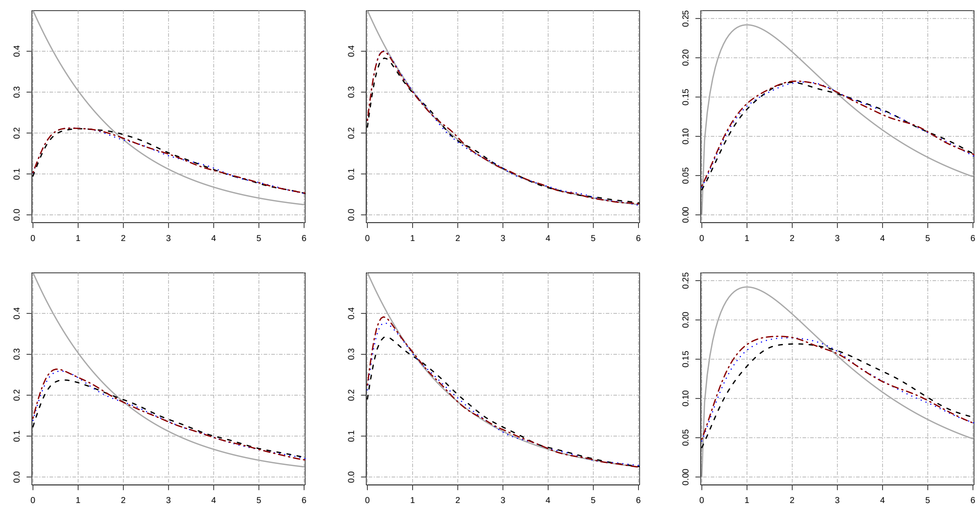
Additional simulations: quadratic asymmetric loss

Figure 3.8: Densities of $J_{\hat{\alpha}}$ under asymmetric quadratic loss and various instrument choices



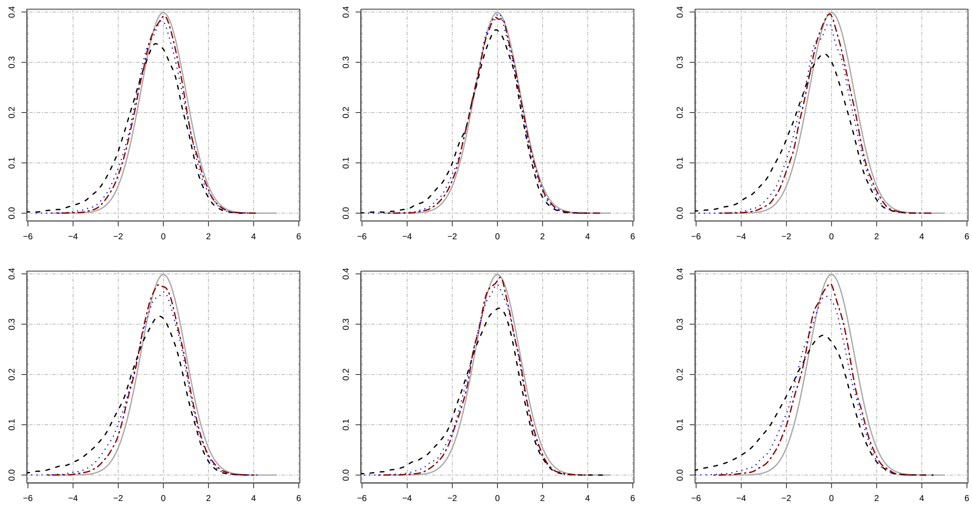
Notes: The figures show the density of the χ^2 (solid grey) and the kernel density estimates of the distribution of $J_{\hat{\alpha}}$ when $T = 100$ (dashed black), $T = 400$ (dotted blue) and $T = 1000$ (dot-dashed red). The used instrument combinations are: a constant and a persistent instrument (left) and a constant a stationary instrument (middle), where the plotted χ^2 distribution has 1 d.o.f., as well as all three instruments (right), where the plotted χ^2 distribution has 2 d.o.f.. The underlying DGP is given by (3.2) - (3.3) with $\delta = 0.5$, $\rho = 0.8$, $\gamma_{0,1} = \gamma_{0,2} = 0.01$, and $\gamma_{1,2} = 0$, exhibiting conditional homoskedasticity ($\gamma_{1,1} = 0$, top) or conditional heteroskedasticity ($\gamma_{1,1} = 0.95$, bottom).

Figure 3.9: Densities of J_{α_0} under asymmetric quadratic loss and various instrument sets



Notes: The figures show the density of the χ^2 (solid grey) and the kernel density estimates of the distribution of J_{α_0} when $T = 100$ (dashed black), $T = 400$ (dotted blue) and $T = 1000$ (dot-dashed red). The used instrument combinations are: a constant and a persistent instrument (left) and a constant a stationary instrument (middle), where the plotted χ^2 distribution has 2 d.o.f., as well as all three instruments (right), where the plotted χ^2 distribution has 3 d.o.f.. See Figure 3.8 for the DGP.

Figure 3.10: Densities of $\mathcal{T}$ under asymmetric quadratic loss and various instrument choices



Notes: The figures show the density of the standard normal distribution (solid grey) and the kernel density estimates of the distribution of $\mathcal{T}$ when $T = 100$ (dashed black), $T = 400$ (dotted blue) and $T = 1000$ (dot-dashed red). The used instrument combinations are: a constant and a persistent instrument (left) and a constant and a stationary instrument (middle), as well as a constant, a persistent and a stationary instrument (right). See Figure 3.8 for the DGP.

Replication study: EU Commission forecast data 1970 – 2004

Tables 3.2 – 3.6 below replicate Tables I – V from Christodoulakis and Mamatzakis (2009) using the data and Matlab codes provided by the authors. In addition, we included a column with the $\mathcal{T}$ statistic for the null $\alpha = 0.5$ in each table.

Table 3.2: Inflation under asymmetric linear loss function

	Current year							Year ahead						
	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$
Bel.	0.42	0.08	−0.92	1.17	6.02	1.82	12.31	0.59	0.08	1.11	6.32	14.46	8.06	7.81
Den.	0.42	0.09	−0.93	0.33	5.10	1.10	10.63	0.41	0.09	−0.96	3.40	6.88	3.36	10.87
Ger.	0.49	0.08	−0.17	0.25	7.97	0.27	9.60	0.68	0.08	2.19	5.61	15.41	8.36	6.25
Gr.	0.47	0.10	−0.30	3.59	5.97	3.68	6.32	0.36	0.10	−1.42	0.65	2.48	2.36	10.01
Sp.	0.23	0.10	−2.74	3.48	3.36	5.23	8.57	0.03	0.04	−12.08	3.75	6.89	7.34	11.00
Fr.	0.43	0.08	−0.89	0.64	6.15	1.32	11.65	0.42	0.08	−0.93	4.26	8.10	4.31	12.12
Ire.	0.56	0.09	0.70	3.41	10.66	3.98	6.78	0.69	0.08	2.24	4.43	15.09	8.74	4.77
Ital.	0.27	0.08	−3.05	3.32	3.54	8.80	17.11	0.20	0.07	−4.27	0.06	0.07	11.79	23.24
Lux.	0.48	0.08	−0.25	5.62	8.92	5.70	11.24	−	∞	−	−	23.19	30.30	32.45
Neth.	0.48	0.08	−0.21	3.49	8.66	3.52	10.31	0.54	0.09	0.45	4.00	10.90	4.00	8.85
Port.	0.02	0.03	−14.53	2.75	6.48	8.34	11.80	0.33	0.11	−1.52	1.26	2.63	2.12	7.97
UK	0.59	0.09	1.01	1.48	11.12	2.02	5.11	0.62	0.09	1.39	2.79	12.53	4.46	5.01
EU	0.48	0.08	−0.21	3.53	8.62	3.63	10.51	0.36	0.08	−1.67	6.08	7.73	6.56	13.70

Notes: Estimates are based on three instruments ($D = 3$): a constant, a lagged forecast error and a lagged realization. $J_{\hat{\alpha}} \sim \chi^2(2)$ and $J_{\alpha|H_0} \sim \chi^2(3)$, see Christodoulakis and Mamatzakis (2009), p. 589.

Table 3.3: Unemployment under asymmetric linear loss function

	Current year							Year ahead						
	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$
Bel.	0.19	0.07	−4.59	11.79	11.55	12.7	13.94	0.56	0.09	0.67	12.18	11.57	12.19	10.53
Den.	0.44	0.09	−0.67	2.9	6.85	2.91	9.56	0.45	0.09	−0.58	5.49	8.03	5.36	9.63
Ger.	0.31	0.08	−2.47	8.42	8.6	7.69	13.79	0.52	0.09	0.22	3.16	9.41	3.05	8.23
Gr.	0.58	0.1	0.74	1.61	7.52	2.16	4.31	0.45	0.11	−0.44	0.27	4.04	0.43	6.79
Sp.	0.78	0.1	2.93	1.96	9.51	5.58	1.85	0.54	0.12	0.33	2.17	5.21	2.26	3.59
Fr.	0.38	0.08	−1.43	4.4	6.21	5.24	12.91	0.45	0.09	−0.59	1.82	7.08	1.98	10.04
Ire.	0.6	0.09	1.14	3.02	11.39	3.55	5.74	0.89	0.06	6.64	7.22	15.31	8.63	7.77
Ital.	0.49	0.08	−0.17	0.16	8.39	0.18	9.14	0.48	0.09	−0.21	2.83	8.37	2.75	9.03
Lux.	0.48	0.09	−0.19	0.1	6.88	0.15	7.71	0.27	0.08	−2.68	5.15	4.75	10.07	14.19
Neth.	0.58	0.08	1.01	2.66	12.35	3.17	6.37	0.6	0.09	1.14	3.65	11.93	4.09	6.23
Port.	0.85	0.08	4.18	3.3	9.28	4.81	3.23	0.89	0.08	5.09	5.28	7.03	4.42	5.3
UK	0.69	0.08	2.36	1.39	15.08	4.72	2.51	0.66	0.09	1.84	5.58	12.05	6.76	5.84
EU	0.43	0.08	−0.83	6.71	9.12	6.33	11.12	0.47	0.09	−0.33	7.87	10.14	7.47	9.73

Notes: Estimates are based on three instruments ($D = 3$): a constant, a lagged forecast error and a lagged realization. $J_{\hat{\alpha}} \sim \chi^2(2)$ and $J_{\alpha|H_0} \sim \chi^2(3)$, see Christodoulakis and Mamatzakis (2009), p. 590.

Table 3.4: Government balance under asymmetric linear loss function

	Current year							Year ahead						
	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$
Bel.	0.22	0.07	−3.83	6.59	7.08	5.4	15.48	0.55	0.09	0.56	1.14	10.19	1.49	6.96
Den.	0.63	0.09	1.41	3.96	9.95	5.32	4.92	0.60	0.09	1.05	5.92	10.76	7.64	6.16
Ger.	0.02	0.02	−22.39	5.61	19.53	22.3	25.52	0.46	0.09	−0.47	4.5	8.23	4.49	10.69
Gr.	0	0.01	−64.6	4.98	12.34	11.02	12.94	0.65	0.1	1.52	4.53	9.24	6.25	4.57
Sp.	0.44	0.12	−0.48	0.17	3.55	0.37	5.26	0.47	0.12	−0.25	0.37	3.78	0.41	4.43
Fr.	0.45	0.08	−0.55	1.15	7.1	1.4	11.1	0.39	0.08	−1.3	3.17	5.87	3.91	12.75
Ire.	0.25	0.08	−3.11	4.88	4.76	7.39	14.3	0.39	0.09	−1.27	1.69	4.43	2.87	11.7
Ital.	0.45	0.08	−0.63	3.27	7.59	3.33	11.18	0.53	0.09	0.37	1.32	10.22	1.44	7.34
Lux.	0.12	0.06	−6.46	1.89	2.62	13.09	20.88	0.24	0.08	−3.23	1.86	2.03	6.7	15.88
Neth.	0.3	0.08	−2.59	1.27	2.4	5.76	17.5	0.33	0.08	−2.14	2.46	3.83	5.23	15.29
Port.	0.19	0.09	−3.29	0.85	0.82	5.5	10.71	0.25	0.11	−2.37	3.49	3.3	3.78	7.23
UK	0.44	0.09	−0.62	1.99	6.5	2.3	9.67	0.55	0.09	0.51	4.22	9.59	4.28	7.14
EU	0.42	0.08	−0.94	1.64	5.99	2.37	12.31	0.59	0.08	1.11	0.95	12.01	1.95	5.37

Notes: Estimates are based on three instruments ($D = 3$): a constant, a lagged forecast error and a lagged realization. $J_{\hat{\alpha}} \sim \chi^2(2)$ and $J_{\alpha|H_0} \sim \chi^2(3)$, see Christodoulakis and Mamatzakis (2009), p. 591.

Table 3.5: Investment under asymmetric linear loss function

	Current year							Year ahead						
	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$
Bel.	0.58	0.08	0.9	0.84	12.26	1.49	5.94	0.5	0.09	0	1.65	8.43	1.61	9.2
Den.	0.24	0.08	−3.38	2.98	3.4	6.14	16.33	0.42	0.09	−0.84	1.79	5.69	2.06	10.28
Ger.	0.56	0.08	0.72	5.03	11.57	5.09	8.58	0.64	0.08	1.74	3	13.99	3.92	5.46
Gr.	0.62	0.1	1.21	1.25	8.69	2.22	3.56	0.64	0.1	1.42	0.62	9.38	2.02	2.33
Sp.	0.14	0.08	−4.38	6.22	5.95	4.15	6.73	0.62	0.12	1.02	2.24	6.28	3.03	2.77
Fr.	0.64	0.08	1.78	1.85	14.44	4.04	4.6	0.58	0.08	0.91	4.02	11.14	4.07	7.52
Ire.	0.34	0.08	−1.93	4.83	5.5	6.38	12.58	0.5	0.09	0	2.59	7.55	2.51	8.22
Ital.	0.6	0.08	1.24	0.44	13.53	1.74	4.75	0.59	0.08	1.08	0.51	11.97	1.64	5.24
Lux.	0.39	0.09	−1.18	0.24	3.46	1.46	10.96	0.48	0.1	−0.22	1.65	6.63	2.31	8.49
Neth.	0.39	0.08	−1.28	0.95	4.78	2.4	13.54	0.09	0.05	−8.33	4.8	5.96	13.09	21.91
Port.	0.56	0.12	0.5	0.38	5.25	0.5	3.57	0.76	0.1	2.46	1.67	8.27	4.23	3.26
UK	0.66	0.09	1.88	1.44	13.79	3.58	3.42	0.64	0.09	1.66	1.2	12.18	3.16	3.47
EU	0.66	0.08	2.06	0.81	16.13	3.85	3.18	0.57	0.08	0.83	2.78	10.94	3.12	6.92

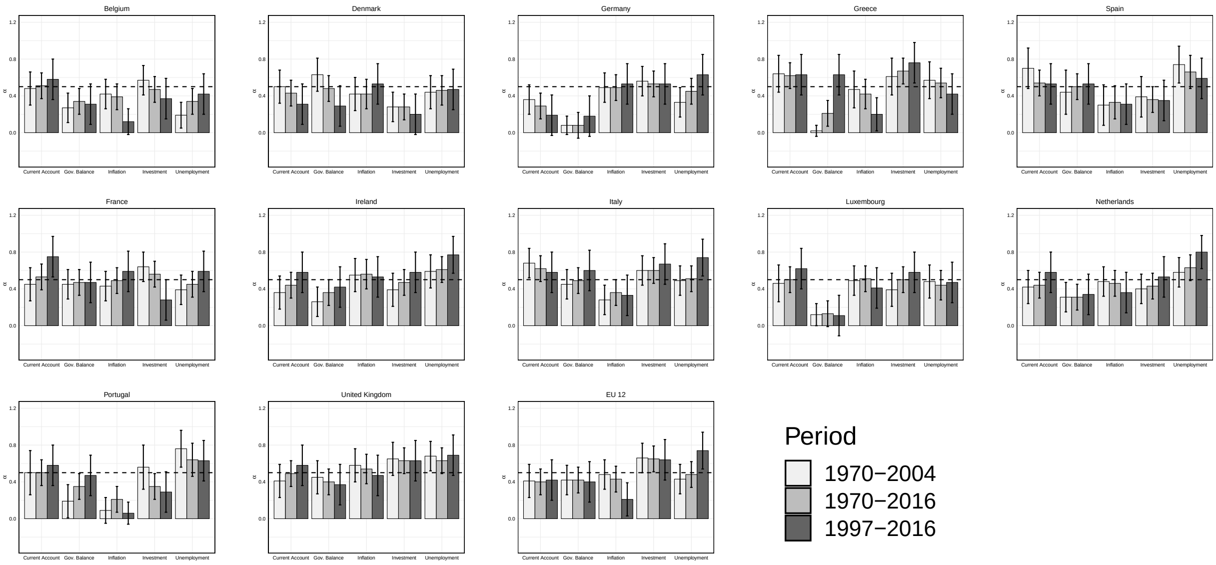
Notes: Estimates are based on three instruments ($D = 3$): a constant, a lagged forecast error and a lagged realization. $J_{\hat{\alpha}} \sim \chi^2(2)$ and $J_{\alpha|H_0} \sim \chi^2(3)$, see Christodoulakis and Mamatzakis (2009), p. 592.

Table 3.6: Current account under asymmetric linear loss function

	Current year							Year ahead						
	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$	$\hat{\alpha}$	SE	$\mathcal{T}$	$J_{\hat{\alpha}}$	$J_{\alpha=0.2}$	$J_{\alpha=0.5}$	$J_{\alpha=0.8}$
Bel.	0.48	0.09	−0.25	4.93	10.34	5.41	9.69	0.52	0.09	0.29	6.44	9.89	6.39	9.1
Den.	0.5	0.09	0	0.76	8.28	0.77	7.82	0.43	0.09	−0.8	1.24	5.46	1.54	10.24
Ger.	0.36	0.08	−1.64	0.11	3.2	2.41	14.26	0.41	0.09	−1.04	2.46	5.97	3.01	11.88
Gr.	0.73	0.1	2.39	4.38	9.47	4.95	4.35	0.61	0.1	1.03	1.64	8.56	2.97	6.25
Sp.	0.92	0.06	6.75	5.46	7.95	4.85	5.83	0.41	0.12	−0.72	5.6	5.97	8.37	8.36
Fr.	0.45	0.09	−0.55	0.85	6.87	1.06	9.93	0.64	0.08	1.7	4.19	13.16	4.66	6.15
Ire.	0.34	0.08	−1.91	4.7	5.77	5.7	12.22	0.46	0.09	−0.4	1.35	6.8	1.4	8.59
Ital.	0.69	0.08	2.36	2.03	15.1	5.37	3.3	0.67	0.08	2.06	3.19	14.35	4.54	4.75
Lux.	0.45	0.1	−0.53	3.42	10.26	3.78	7.96	0.43	0.1	−0.64	4.32	5.97	4.39	7.85
Neth.	0.4	0.09	−1.14	3.63	6.82	5.14	12.29	0.55	0.09	0.57	11.41	11.99	11.65	10.26
Port.	0.5	0.12	0	3.79	4.96	3.75	4.81	0.7	0.11	1.79	2.23	7.38	3.9	2.22
UK	0.41	0.09	−1.03	1.72	5.44	2.13	11.18	0.25	0.08	−3.18	1.06	1.45	6.7	16.83
EU	0.41	0.09	−1.02	2.25	5.89	2.77	11.8	0.2	0.07	−4.37	10.72	10.68	15.5	14.84

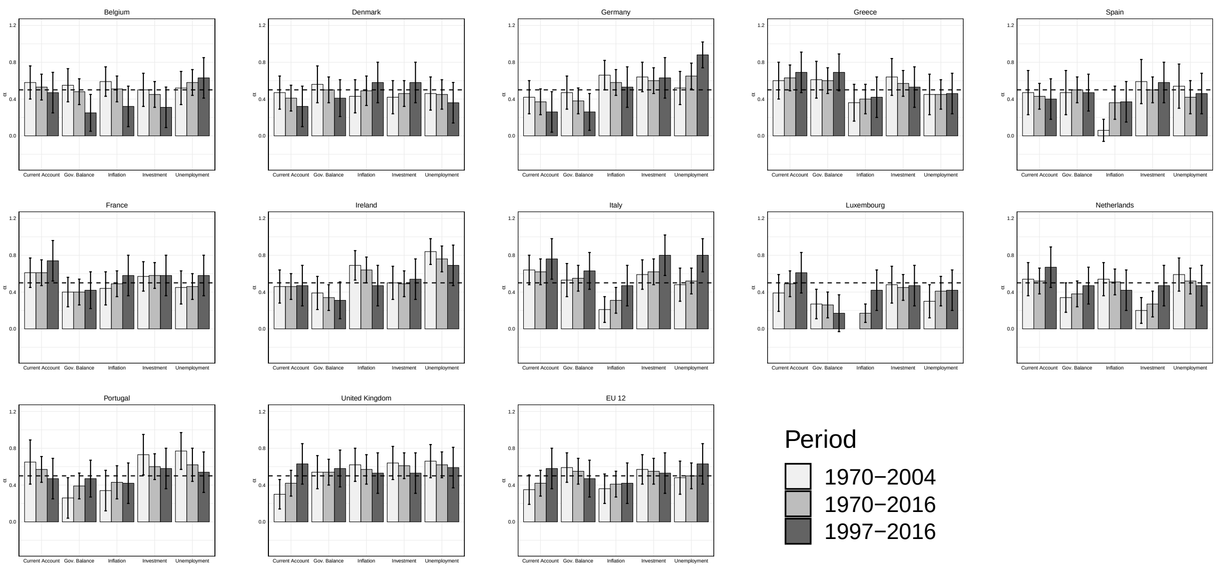
Notes: Estimates are based on three instruments ($D = 3$): a constant, a lagged forecast error and a lagged realization. $J_{\hat{\alpha}} \sim \chi^2(2)$ and $J_{\alpha|H_0} \sim \chi^2(3)$, see Christodoulakis and Mamatzakis (2009), p. 592.

Figure 3.11: Asymmetry estimates for the EU–12 Member States over different time spans, linear asymmetric loss and $D = 2$ instruments (current year forecasts)



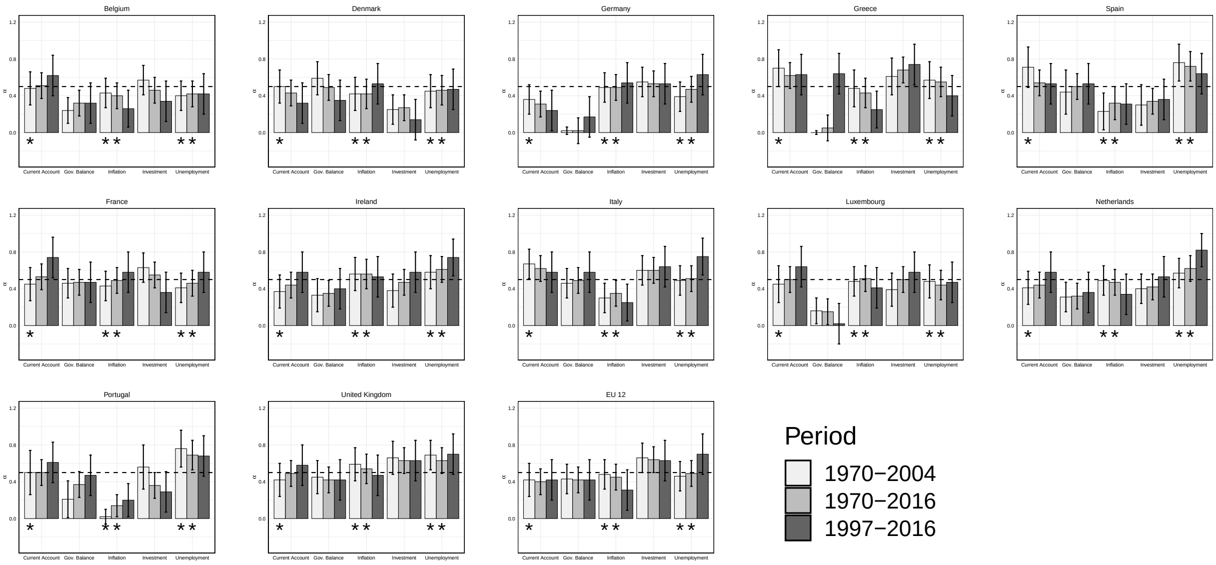
Notes: The figure shows α estimates with error bars (± 2 standard errors). The estimation was conducted using two instruments: a constant and a lagged realization; $p = 1$.

Figure 3.12: Asymmetry estimates for the EU–12 Member States over different time spans, linear asymmetric loss and $D = 2$ instruments (year ahead forecasts)



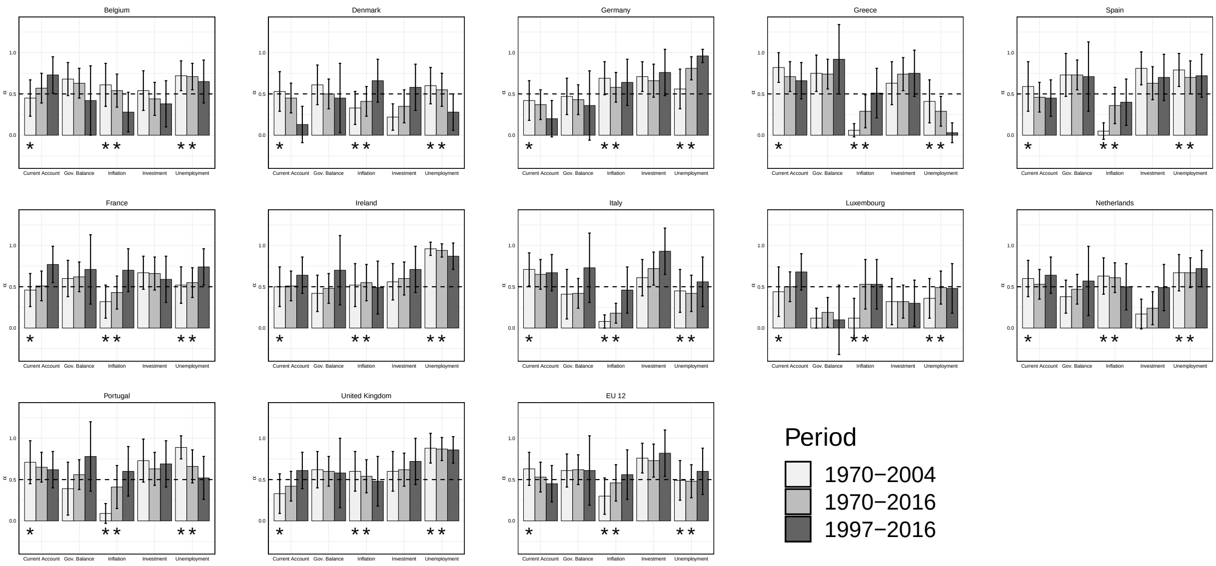
Notes: The figure shows α estimates with error bars (± 2 standard errors). The estimation was conducted using two instruments: a constant and a lagged realization; $p = 1$. Missing bars indicate failed convergence of the estimation algorithm.

Figure 3.13: Asymmetry estimates for the EU–12 Member States over different time spans, asymmetric quadratic loss and $D = 2$ instruments (current year forecasts)



Notes: The figure shows α estimates with error bars (± 2 standard errors). The estimation was conducted using two instruments: a constant and a lagged forecast error; $p = 2$. Asterisks represent significance at 5% level.

Figure 3.14: Asymmetry estimates for the EU–12 Member States over different time spans, asymmetric quadratic loss and $D = 2$ instruments (year ahead forecasts)



Notes: The figure shows α estimates with error bars (± 2 standard errors). The estimation was conducted using two instruments: a constant and a lagged forecast error; $p = 2$. Asterisks represent significance at 5% level.

Chapter 4

Asymmetric-Loss-Based Evaluation of Daily Value-at-Risk Models

4.1 Motivation

Foreseeing and quantifying market risk has always been important, even more so in light of the recent global financial instability and its aftermath. Value-at-risk (VaR) is defined as a potential loss of a portfolio that can occur with a fixed probability on a given day in the future. It has received a lot of attention in the literature due to its interpretability and convenient mathematical properties. VaR is widely used by risk managers, financial and non-financial institutions, and is suggested by The Basel Committee on Banking Supervision as a guideline for satisfying market risk capital requirements. It is conventionally defined as the negative of the $100\%\tau$ quantile of the conditional distribution of a daily log return r_t,

$$\mathrm{VaR}_t^\tau = -\sup\left(r|P(r_t \leq r) \leq \tau\right). \tag{4.1}$$

The difficulty of accurately modelling value-at-risk lies in its latent nature. The precision of a VaR forecast cannot be established by observing its realized values; assessing the quality of a forecast is demanding, but there is, however, a variety of methods to choose from. Nieto and Ruiz (2016) provide an extensive summary of existing modelling methods and their comparison techniques. The authors distinguish between one and two-step procedures: the former yields a direct forecast of the τ quantile, while the latter estimates conditional mean and variance, first, and combines it with a distributional assumption to obtain the forecast. The most popular examples of the one-step approach are historical simulation (see e.g. Dowd, 2007), extreme value theory (see e.g. Embrechts et al., 2013) and conditional autoregressive value-at-risk (Engle and Manganelli, 2004). The two-step approach relies heavily on a conditional volatility forecast. However, predicting volatility and value-at-risk are two completely different objectives, meaning, a sound volatility forecast does not automatically lead to a good VaR forecast (e.g. Bams et al., 2017).

There has been a growing interest in employing information beyond the return series in the VaR forecasting literature. Giot and Laurent (2004) include realized volatility into the information set used for capturing value-at-risk. Later, in their review, Kuester et al. (2006) mentioned the importance of realized volatility models for VaR prediction. Since then multiple applications of realized measures in VaR modelling have been considered in the literature (see e.g. Louzis et al., 2014; Haugom et al., 2016; Wong et al., 2016). In particular, Žikeš and Baruník (2014) utilize heterogeneous autoregression (HAR) model of Corsi (2009) in a quantile autoregression framework. The authors argue that realized measures, as well as certain exogenous regressors (e.g. option implied volatility), possess substantial predictive power when it comes to forecasting value-at-risk. In addition, they augment the well-known CAViaR model of Engle and Manganelli (2004) with realized and implied volatility and call it realized CAViaR. This idea of extending the CAViaR model has been realized in a few other studies. Rubia and Sanchis-Marco (2013) examine the influence of information beyond historical returns on the forecasts of VaR. They adapt CAViaR methodology of Engle and Manganelli (2004) and enrich their models with trading activity and market liquidity variables. They find that including data related to trading volume significantly enhances the out-of-sample performance of VaR models. Furthermore, Jeon and Taylor (2013) use implied volatility indices as external regressors in order to include market expectations. They find that combining time series information delivered by CAViaR models with the information on implied volatility yields a superior value-at-risk forecast. Bams et al. (2017) also examine the role of implied volatility in predicting value-at-risk, but do not detect any significant improvement.

The main goal of this paper is to provide a flexible, computationally effective and data-driven model for daily VaR forecasting. Here, the focus is mainly on 1% and 5% quantiles of the conditional return distribution. For that purpose, a linear quantile autoregression with various realized measures is estimated. In that way, a direct quantile forecast can be obtained by relying solely on the quantile loss function. The main contributions of this paper are the following. First, this combination of weakly exogenous predictors, such as daily trading volume, day of the week and implied volatility, has not yet been employed in predictive models. Second, as opposed to the majority of VaR studies, a larger number of return series, of both stock indices and single stocks, is used. Third, the forecasting performance of suggested models is compared to popular benchmarks, such as GARCH, apARCH, CAViaR and GAS models, using three types of criteria. The latter enables the models to be scored and shows the suggested quantile autoregressions to deliver the best performance overall.

The paper is structured as follows. The model setup and properties are presented in Section 4.2. Then, the model validation methods are reviewed in Section 4.3. Section 4.4

is devoted to empirical analysis and its results. Finally, Section 4.5 concludes.

4.2 Model setup

This section introduces the theoretical framework and model setup. The focus lies on the so-called one-step approaches that are based on the quantile autoregression by Koenker and Xiao (2006). The idea is to hereby avoid any dependence of the forecast procedure on the optimality of the volatility forecast and only rely on a loss function appropriate for value-at-risk. The quantile regression framework allows for regressors that are based on historical returns, as well as on information regarding market characteristics.

4.2.1 Heterogeneous quantile autoregression

Assume conditional τ-quantile of the future returns distribution to be a linear function of variables based on past quadratic variation and external predictors,

$$Q_\tau(r_{t+1}|\mathcal{F}_t) \quad = \quad \beta_0(\tau) + \boldsymbol{\beta}_v(\tau)'\boldsymbol{v_t} + \boldsymbol{\beta}_z(\tau)'\boldsymbol{z_t}, \quad \tau \in (0,1), \tag{4.2}$$

where $\boldsymbol{v_t}$ collects various realized measures, $\boldsymbol{z_t}$ is a vector of external regressors and $\boldsymbol{\beta}_{\cdot}(\tau)$ are the autoregression coefficients to be estimated. The estimation is performed using quantile regression methodology of Koenker and Bassett (1978). The parameter estimator is given by

$$\widehat{\boldsymbol{\beta}}(\tau) \quad = \quad \text{argmin}_{\beta \in R^d} \Big\{ \sum_{t=1}^{T} \rho_\tau(r_{t+1} - \beta_0(\tau) - \boldsymbol{\beta}_v(\tau)'\boldsymbol{v_t} - \boldsymbol{\beta}_z(\tau)'\boldsymbol{z_t}) \Big\},$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}(u < 0))$ is the quantile loss function with $\mathbf{1}(\cdot)$ being a common indicator function.[1]

Given the relation between the conditional quantile of a daily return and its conditional standard deviation (assuming a zero mean),

$$Q_\tau(r_{t+1}|\mathcal{F}_t) = \sigma_{t+1} F_\varepsilon^{-1}(\tau),$$

where σ_{t+1} is the day ahead volatility and F_ε^{-1} is the inverse cdf of the innovations, it is convenient to use the heterogeneous autoregressive (HAR) model of Corsi (2009) for the linear dependence of the value-at-risk on the past volatility. The HAR model captures the persistent nature of volatility in a simple manner, which is easily adapted to a quantile regression (see e.g Haugom et al. (2016); Žikeš and Baruník (2014)). There are various

[1] This optimization problem doesn't have a closed-form solution and requires a linear programming algorithm, which is provided by Koenker (2012).

proxies for daily volatility that have been discussed in the VaR literature: from simple squared returns to realized measures computed with high frequency intraday data. In this paper the following realized measures are used: realized variance (Andersen and Bollerslev, 1998), bipower variation (Barndorff-Nielsen and Shephard, 2004, 2006) and median realized variance (Andersen et al., 2012).

4.2.2 Realized measures

Realized measures are computed from high frequency intraday returns. They aim to consistently estimate and predict quadratic variation that consists of two main components: (continuous) integrated variance and (discrete) jump variation, $RV_t = RC_t + RJ_t$. Lately, it has become common practice to estimate these components separately. It has been well documented that these two sources of variation affect volatility quite differently (e.g. Corsi, Pirino, and Reno, 2010; Andersen, Bollerslev, and Diebold, 2007; Giot and Laurent, 2007; Busch, Christensen, and Nielsen, 2011).

The realized measures chosen for this paper are asymptotically equivalent in the absence of jumps. Notably, realized variance is not jump-robust, while median realized variance and bipower variation are. All three measures can be used to construct a simple HAR-like model for value-at-risk:

$$\mathrm{VaR}_{t+1}^{\tau} = \beta_{\tau,0} + \beta_{\tau}^{(d)} RM_t^{(d)} + \beta_{\tau}^{(w)} RM_t^{(w)} + \beta_{\tau}^{(m)} RM_t^{(m)} + \boldsymbol{\beta}_z(\tau)' \boldsymbol{z_t} \tag{4.3}$$

where $RM_t^{(\cdot)}$ represent a daily, weekly and monthly realized measure component according to the HAR-RV model of Corsi (2009). The components are defined as $RM_t^{(d)} = RM_t$, $RM_t^{(w)} = (RM_t + \ldots + RM_{t-4})/5$ and $RM_t^{(m)} = (RM_t + \ldots + RM_{t-21})/22$ [2].

These realized measures and many more are freely available to researchers at the Oxford-Man Institute's realized library (Heber et al., 2009).[3] The library contains volatility estimators for a variety of stock indices. However, if one wishes to analyze single stocks, as is done here in Section 4.4, high frequency intraday data required for computation of the realized measures might be hard to come by. To circumvent this issue the measures of a particular stock index can be used in Equation (4.3) as proxies for market volatility. This has shown to be an appropriate action (see Section 4.4.2 for results).

It is important to note at this point that estimation of the original HAR model is sometimes performed in logs. However, motivated by the linear relationship between quantiles and volatility, the quantile regression is estimated here using levels of the realized

[2] The model can be constructed using $RM_t^{(\cdot)}$ or $\sqrt{RM_t^{(\cdot)}}$. Due to linear relationship between quantiles and standard deviations, I decided to specify the VaR model in terms of square roots (see Corsi et al. (2012))

[3] The mathematical definitions of the realized measures are included in the Appendix.

measures, but estimation with their logs is performed as a robustness check.

4.2.3 Additional predictors

In addition to HAR components, the role of several observables in VaR forecasting, such as implied volatility, downward semivariance, trading volume and day of the week, is examined. Giot and Laurent (2007) documented high explanatory power of implied volatility when forecasting realized volatility. Bams et al. (2017), Žikeš and Baruník (2014) and Jeon and Taylor (2013) used implied volatility in building VaR forecasts. The evidence, however, is mixed. Most stock market indices have corresponding implied volatility indices (e.g. VDAX for DAX, V2TX for EUROSTOXX 50, VFTSE for FTSE 100, and VXD for Dow Jones). Implied volatility reflects rational expectations on future volatility of the market. The indices are usually annualized, so they are divided $\sqrt{252}$ to obtain daily implied volatility as in Jeon and Taylor (2013) (also see Busch et al., 2011).

Similarly to the CAViaR framework and various other volatility models (see e.g. Engle and Manganelli, 2004; Martens et al., 2009; Wong et al., 2016), some functions of a lagged return are included. It is a standard practice to include the absolute value of the lagged return, but sometimes the effect of a negative return is different from the effect of a positive one of the same magnitude. This asymmetry can be captured by including a sign effect variable. At last, an interaction effect of the sign and the magnitude is included.

This asymmetry may, however, be captured in a different way. There has been some evidence in favour of realized downward semivariance (Barndorff-Nielsen et al., 2008). It is defined as a sum of squared negative intraday returns,

$$RS_t^- = \sum_{j=1}^{M-1} (r_{t,j})^2 \mathbf{1}(r_{t,j} < 0), \quad t = 1, \ldots, T,$$

and is said to be more informative than its positive counterpart, realized upside semivariance (Žikeš and Baruník, 2014; Patton and Sheppard, 2015). This measure reflects an asymmetric effect of past negative returns in a more sophisticated manner than the sign-magnitude interaction. It is possible that including the downward semivariance will substantially dampen the effect of the sign variable. Nevertheless, if this realized measure is unavailable, the asymmetry has to be accounted for.

As mentioned above, Rubia and Sanchis-Marco (2013) examine roles of market activity and liquidity variables in VaR forecasting. They use CAViaR setting, extend it, and find robust evidence for including volume-related variables into the relevant information set.

Finally, there is a distinct U–shape of the average volatility through the week with its lowest point on Wednesdays (see Andersen and Bollerslev, 1998; Martens et al., 2009). Also, volatility appears to be higher on days when macroeconomic news announcements are made, which mostly happens on Fridays (see Andersen et al., 2003, 2007; Martens et al., 2009). For these reasons, variables for Wednesday and Friday effects are included in the quantile autoregression.

4.2.4 Model specifications

To sum up, a full set of additional predictors includes daily trading volume, implied volatility, dummy variables for Wednesday and Friday, realized downward semivariance, absolute value of a lagged return, its sign and an interaction of the latter two. The full QREG-RM model is then given by:

$$\begin{aligned} Q_\tau(r_{t+1}|\mathcal{F}_t) &= \beta_{\tau,0} + \beta_\tau^{(d)} RM_t^{(d)} + \beta_\tau^{(w)} RM_t^{(w)} + \beta_\tau^{(m)} RM_t^{(m)} \\ &+ \beta_{\tau,1} RS_t^- + \beta_{\tau,2} IV_t + \beta_{\tau,3}|r_t| + \beta_{\tau,4}\mathbb{I}(r_t < 0) \\ &+ \beta_{\tau,5}|r_t|\mathbb{I}(r_t < 0) + \beta_{\tau,6} Vol_t + \beta_{\tau,7} Wed + \beta_{\tau,8} Fr. \end{aligned} \tag{4.4}$$

Depending on the availability of information and different realized measures I distinguish between fourteen specifications of the quantile autoregression. The first six models, QREG-RM, represent quantile HAR regressions with a full set of additional predictors. Here the HAR components are composed using realized volatility, bipower variation and median realized volatility, first in levels, then in logs. Secondly, in order to check the effect of exogenous predictors, the HARQ (see Haugom et al., 2016) models are examined, also with levels and logs. Finally, if high frequency data is unavailable no realized measures can be computed, additional specifications, QREG-IV1 and QREG-IV2, are introduced where the load is carried by the lagged return, external predictors and log implied volatility. The goal here is to compare the explanatory power of realized measures, exogenous predictors and a combination of the two. Table 4.1 summarizes all specifications of interest.

In the following section, the model evaluation criteria and backtesting procedures used in this analysis are discussed. These procedures help assess the performance of the competing models from both a statistical and a regulatory point of view.

Table 4.1: Model specifications

Model →	**QREG**								**HARQ**					
Regressor ↓	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
RV^d	*			log					*			log		
RV^w	*			log					*			log		
RV^m	*			log					*			log		
BV^d		*			log					*			log	
BV^w		*			log					*			log	
BV^m		*			log					*			log	
$MedRV^d$			*			log					*			log
$MedRV^w$			*			log					*			log
$MedRV^m$			*			log					*			log
RS^-	*	*	*	log	log	log								
IV	*	*	*	log	log	log	log	log						
$\lvert r_t\rvert$	*	*	*	*	*	*		*						
$\mathbf{1}(r_t < 0)$	*	*	*	*	*	*	*	*						
$\lvert r_t\rvert\mathbf{1}(r_t < 0)$	*	*	*	*	*	*		*						
r_t^2							*							
Volume$\times 10^{-8}$	*	*	*	*	*	*	*	*						
Wednesday	*	*	*	*	*	*	*	*						
Friday	*	*	*	*	*	*	*	*						

4.3 Comparing model performances

Since the variety of models to choose from is large, one needs a comprehensive strategy and criteria to select the most suitable ones for a given purpose. This purpose can be rooted in achieving statistical accuracy, e.g. minimizing a loss function, or meeting regulatory requirements, e.g. good backtesting results. It has been mentioned that standard statistical procedures for VaR backtesting often fail to discriminate between different forecasting strategies, especially when performed on data from calm periods (Danıelsson, 2002; Laurent, 2017). In most of the value-at-risk literature, a comparison of models is performed on a few index and stock time series with only two to three tests to evaluate the results. Arguably, this is not sufficient to find a robust model and make practical suggestions. For that reason the evaluation is conducted in three stages: model confidence set procedure, lowest average loss, and five backtesting procedures. The models are then ranked by their performance in all three stages.

4.3.1 Statistical criteria

First, the model confidence set (MCS) procedure of Hansen et al. (2011) is applied to a representative cross section of value-at-risk. It has been recently used for comparing VaR models by Bernardi and Catania (2016)[4]. The MCS approach consists of a sequence of tests that allows constructing a so-called superior set of models (SSM). The null hypothesis of this test sequence is the equal predictive ability (EPA) of models in question. The principle is similar to the test of Diebold and Mariano (1995), although it can be used to compare more than two models at once. The MCS procedure is very flexible in terms of selection criteria. A test statistic is constructed from a series of losses produced by

[4] The authors provide an R package for the MCS procedure (see Bernardi and Catania, 2018)

each model and a preselected loss function. In that way models of different classes can be simultaneously compared to one another. For the purpose of VaR forecasting, it is best to consider the asymmetric loss function of González-Rivera et al. (2004), defined as

$$\mathcal{L}(r_t, \widehat{\mathrm{VaR}}_t^{\tau}) = (\tau - d_t^{\tau})(r_t - \widehat{\mathrm{VaR}}_t^{\tau}), \quad t = T+1, \ldots, T+H, \tag{4.5}$$

where $\widehat{\mathrm{VaR}}_t^{\tau}$ is the predicted value-at-risk, $d_t^{\tau} = \mathbf{1}(r_t < \widehat{\mathrm{VaR}}_t^{\tau})$ is a violation variable. Note, that the subsample $[T+1, T+H]$ corresponds to the validation sample of length H with T being the number of observations used to estimate the models. The asymmetric loss function of González-Rivera et al. (2004) is an appropriate choice for assessing VaR models, since it penalizes downward deviations from τ-level quantile more heavily.

As for any test, a significance level α should be fixed for the MCS procedure. It follows, then, the final SSM can contain models inferior to others. In the best case scenario, the SSM contains only one model, which, however, rarely happens. Otherwise, all remaining models posses equal predictive accuracy, and the selection process requires more criteria to discriminate further[5]. Conveniently, the MCS procedure also ranks the models in the SSM by the minimum average loss according to (4.5). While there can only be one model with the minimum loss in a bunch, this is a much more helpful step in decision making.

4.3.2 Regulatory criteria

Since value-at-risk is an unobservable quantity, predicted values cannot be compared to its realizations. The Basel accords suggest backtesting the series of predicted values and require it to be performed on at least 250 one-step-ahead VaR forecasts. Many well-known tests are based on a binary hit variable, $I_t(\tau) = \mathbb{I}(r_t < -\mathrm{VaR}_t^{\tau})$, e.g. the unconditional coverage test (UC) of Kupiec (1995) and the conditional coverage test (CC) of Christoffersen (1998). It is, however, well documented, that these tests suffer from a substantial lack of power in finite samples (see e.g. Gaglianone et al., 2011; Nieto and Ruiz, 2016). The dynamic quantile test (DQ) of Engle and Manganelli (2004) links the series of violations to the lagged hits, and it is said to be the best procedure for the 1% VaR (Berkowitz et al., 2011). All of the above mentioned backtests have been thoroughly discussed in the literature, so I refrain from further details on this account.

Another interesting and comprehensive approach was suggested by Dumitrescu et al. (2012) that has not yet received much attention. They argue that linear regression, as in DQ, is not a suitable tool for a binary dependent variable such as VaR violation. The authors propose several specifications of a dynamic binary test. In this setup the

[5]For further details see Hansen et al. (2011)

conditional probability of a violation is assumed to be dependent on some index π_t:

$$P(I_t(\tau)|\mathcal{F}_{t-1}) = \mathrm{E}[I_t(\tau)|\mathcal{F}_{t-1}] = F(\pi_t), \tag{4.6}$$
$$\pi_t = c + \sum_{j=1}^{q_1} \rho_j \pi_{t-j} + \sum_{j=1}^{q_2} \delta_j I_{t-j}(\tau) + \sum_{j=1}^{q_3} \phi_j l(y_{t-j}, \phi) + \sum_{j=1}^{q_4} \gamma_j l(y_{t-j}, \phi) I_{t-j}, \tag{4.7}$$

where $F(\cdot)$ denotes a cdf, $l(\cdot)$ a function of lagged observables, and y_t is a vector of explanatory variables, e.g. lagged returns (Dumitrescu et al., 2012). In their paper, the authors propose seven versions of the dynamic binary model, from which I choose two, namely the least and the most restrictive ones:

$$DB_1 : \pi_t = c + \rho_1 \pi_{t-1}, \tag{4.8}$$
$$DB_7 : \pi_t = c + \rho_1 \pi_{t-1} + \delta_1 I_{t-1}(\tau) + \phi_1 VaR_{t-1} + \gamma_1 Var_{t-1} I_{t-1}. \tag{4.9}$$

The first specification is a simple AR(1) representation and the last one reflects an asymmetric effect lagged VaR values can have on the index π_t. In terms of estimation, constrained maximum likelihood can be applied to obtain estimates for $\theta = (\rho', \gamma', \phi', \delta')$. Then one can test the assumption of conditional temporal independence of violations with the null hypothesis

$$H_0 : \rho = 0,\ \delta = 0,\ \phi = 0,\ \gamma = 0 \text{ and } c = F^{-1}(\tau).$$

Under the null hypothesis the percentage of violations is equal to τ on average, and the violations are uninformative. If the null is rejected, the model delivering the corresponding VaR forecast can be improved, since the hit sequence contains omitted information[6].

In the backtesting stage of model validation five backtests are used: unconditional coverage test, conditional coverage test, dynamic quantile test and two specifications of dynamic binary test mentioned above. A model is considered admissible if none of the five null hypotheses is rejected at 10% significance level. The results of my extensive empirical analysis of model performance are presented in the following section.

4.4 Empirical analysis

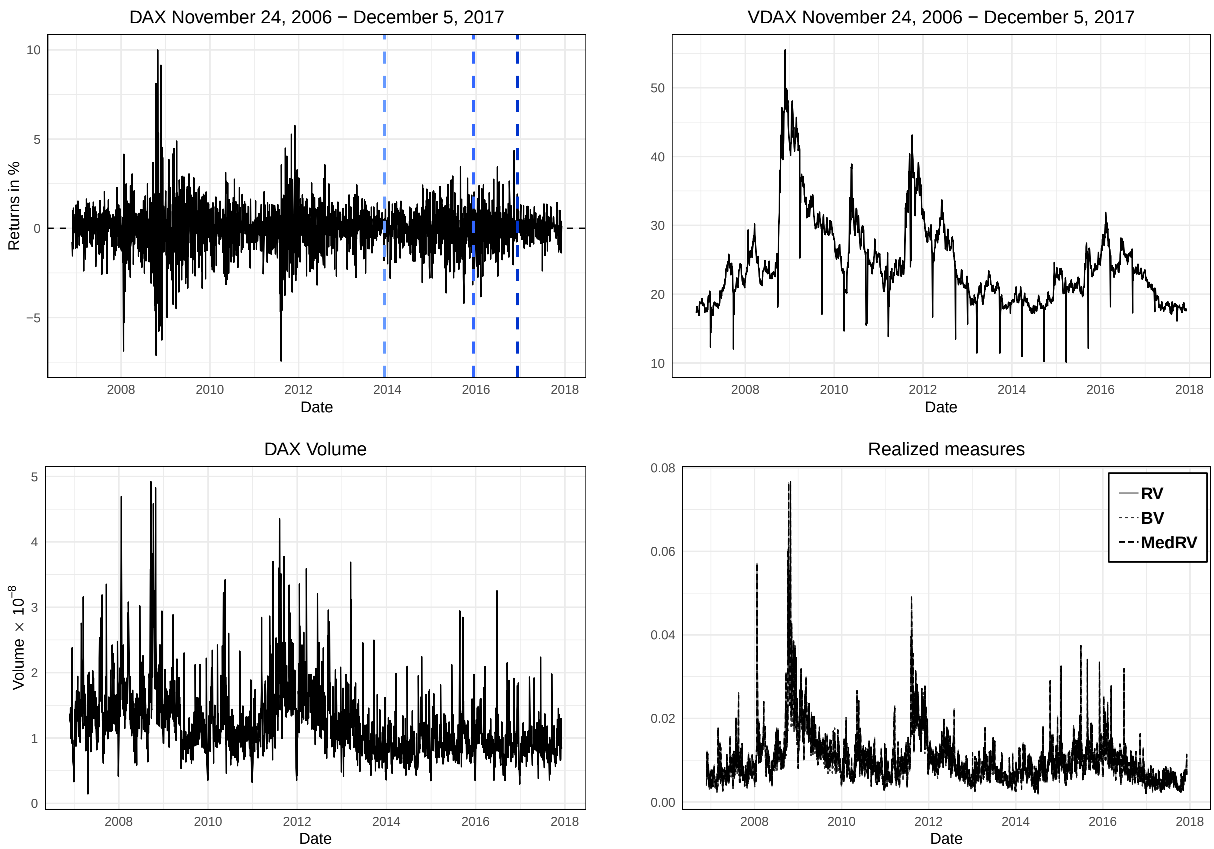
In order to encompass all the information obtained from the model confidence set and backtesting procedures a rather simple additive scoring system is used. Starting with the MCS, each model receives a score of one each time it is included in the superior set of models. An additional score of one is awarded to the model with the lowest average loss.

[6] The Matlab code for estimation and testing is available on http://www.runmycode.org/companion/view/35

At last, if a model passes all of the five backtests, its score is, again, increased by one. This scoring is conducted for a variety of cases depending on quantile levels and validation sample sizes. Additionally, there are two major datasets: an index dataset containing four international stock indices, and a stock dataset with sixteen different stocks[7]. In most of the VaR forecasting studies, the model comparison is performed on a small number of return series. A substantially larger data pool should help make a more meaningful statement.

4.4.1 Data

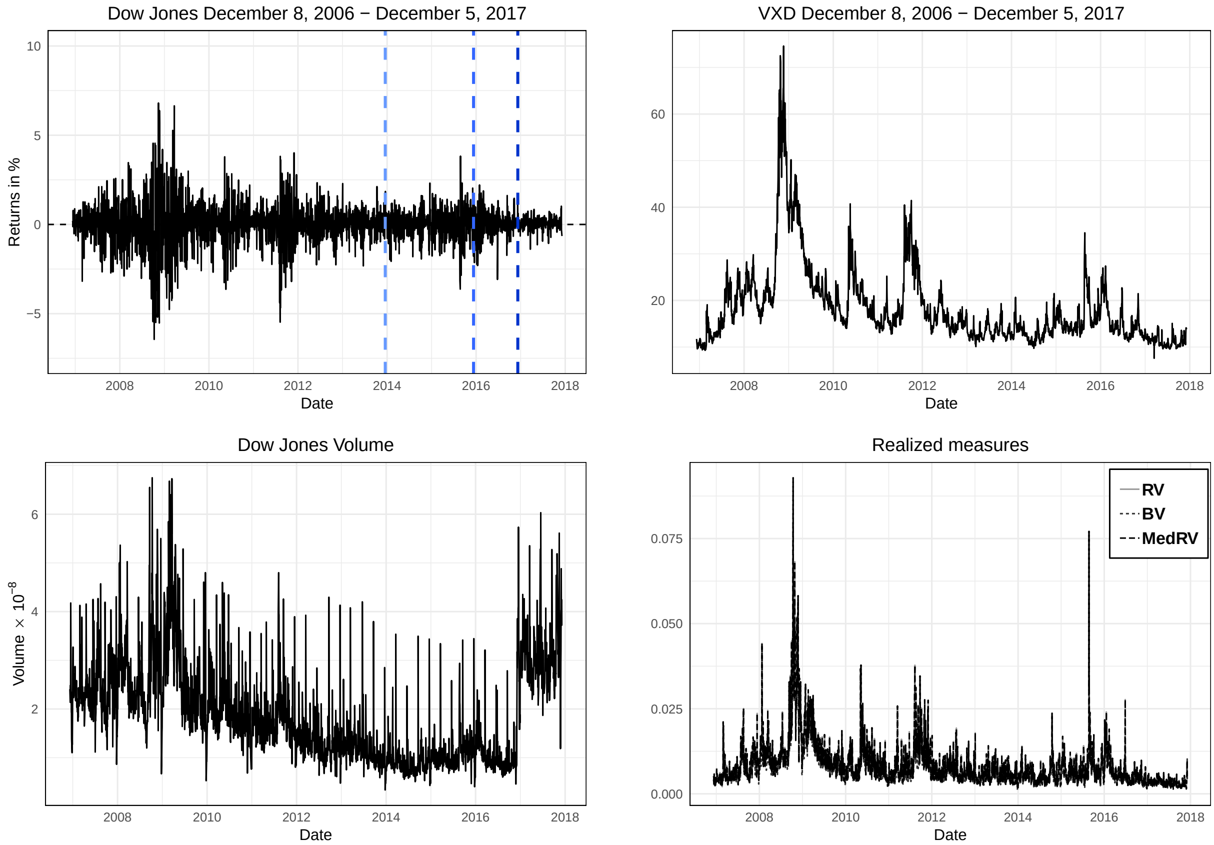
The index dataset contains four stock indices over a time period between 2006 and 2017: DAX 30, Dow Jones, FTSE 100 and EUROSTOXX 50. The set contains between 2760 and 2717 observations. The stocks dataset consists of sixteen different stock returns from several countries and industries. The descriptive statistics for both datasets are presented in Tables 4.2 and 4.3.



Note: The dashed blue lines in the upper left plot represent subsamples reserved for backtesting, $H \in \{250, 500, 1000\}$.

Figure 4.1: DAX 30, daily returns and related series

[7]Stock data, option implied volatility and daily trading volume are obtained from *Thomson Reuters Datastream*



Note: The dashed blue lines in the upper left plot represent subsamples reserved for backtesting, $H \in \{250, 500, 1000\}$.

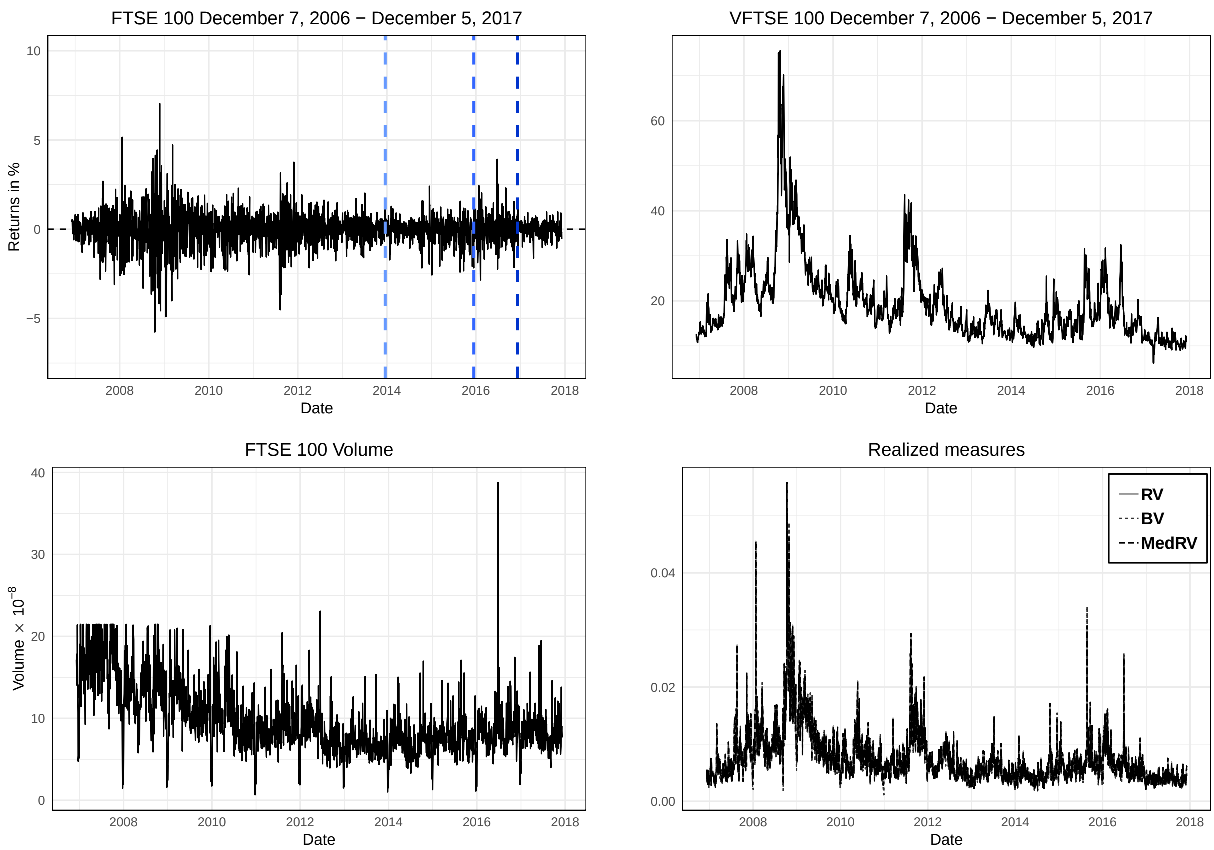
Figure 4.2: Dow Jones, daily returns and related series

Table 4.2: Descriptive statistics of the index data (in percentage points)

	AR(1) coef.	(Std. err.)	$\bar{\hat{\epsilon}}$	$\hat{\sigma}^2$	$\hat{\xi}$	$\hat{\kappa}$	p^{LB}	1%	5%	N obs.
DAX	0.027	(0.035)	0	1.465	0.092	10.215	0.001	−3.469	−1.939	2759
Dow Jones	−0.026	(0.033)	0	1.275	−0.105	14.020	0.000	−3.489	−1.797	2760
FTSE 100	0.012	(0.028)	0	0.777	−0.077	8.823	0.000	−2.504	−1.483	2760
EUROSTOXX 50	−0.001	(0.023)	0	1.508	−0.377	10.257	0.023	−3.583	−1.885	2717

Note: AR(1) coefficient is estimated by OLS, HAC robust standard errors in parentheses; $\bar{\hat{\epsilon}}$ is the sample mean of OLS residuals, $\hat{\sigma}^2$ the sample variance, $\hat{\xi}$ the sample skewness, $\hat{\kappa}$ the sample kurtosis. p^{LB} are the p-values of the Ljung-Box test for 20 lags. 1% and 5% represent empirical quantiles.

Figures 4.1–4.4 show series from the index dataset. Daily returns are depicted in the upper left panel with blue lines marking beginnings of different validation periods reserved for backtesting, $H \in \{250, 500, 1000\}$. The upper right plots show corresponding implied volatility indices. The lower left and right graphs are daily trading volume and realized measures, respectively. All return series were filtered using an AR(1) model, which is a standard practice (see e.g. Kuester et al., 2006 and references therein). The corresponding demeaned returns exhibit high kurtosis and mostly negative skewness.

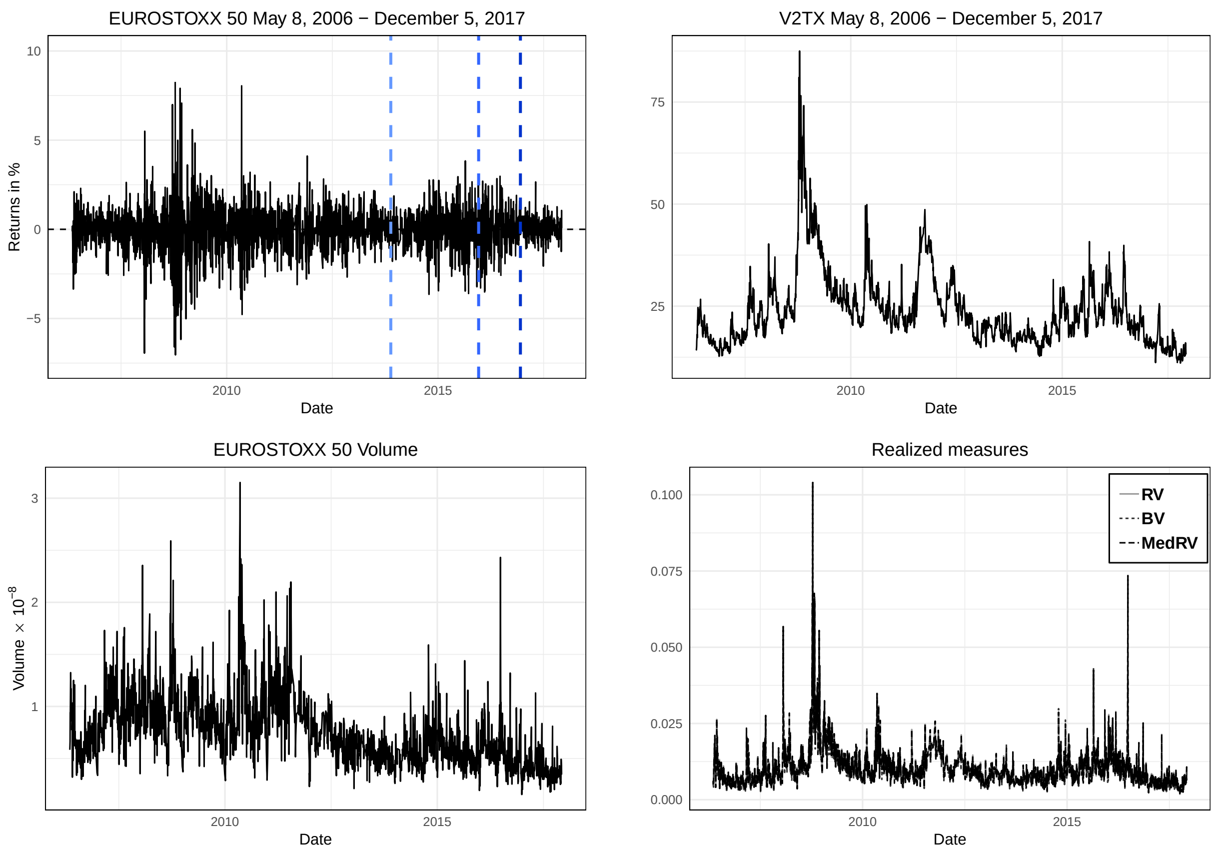


Note: The dashed blue lines in the upper left plot represent subsamples reserved for backtesting, $H \in \{250, 500, 1000\}$.

Figure 4.3: FTSE 100, daily returns and related series

The realized measures discussed in Section 4.2.2 are taken from Oxford-Man Institute's realized library (Heber et al., 2009). The library contains several realized measures computed for different stock indices. In general, it is not easy to obtain high frequency intraday return data for single stocks. Hence, the single stocks I picked are included in the calculation of the four indices from the index set. The purpose of this is to use the HAR components of the indices as market volatility proxies in the single stock regressions.

As it is obvious from Figures 4.1–4.4, trading volume and realized measures series are not always well-behaved, and it is possible that, besides additional information, they will introduce undesired noise into VaR forecasts. It might be a good idea to filter out some of that noise before using these variables in quantile autoregression, but this is beyond the scope of this paper.



Note: The dashed blue lines in the upper left plot represent subsamples reserved for backtesting, $H \in \{250, 500, 1000\}$.

Figure 4.4: EUROSTOXX 50, daily returns and related series

Table 4.3: Descriptive statistics of the stocks data (in percentage points)

Index	Company	Tick	Industry	AR(1)	(Std. err.)	$\hat{\sigma}^2$	$\hat{\xi}$	$\hat{\kappa}$	p^{LB}	1%	5%
DAX	Bayer	BAYN	Pharma	-0.060	(0.025)	3.43	0.648	17.22	0.026	-4.98	-2.78
	BMW	BMW	Manufac.	0.041	(0.025)	4.36	0.082	7.35	0.003	-5.49	-3.27
	Deutsche Telekom	DTE	Commun.	-0.023	(0.023)	2.49	0.173	11.12	0.000	-4.04	-2.36
	SAP	SAP	Software	0.033	(0.035)	2.28	-0.591	16.12	0.004	-4.25	-2.25
DJI	Apple	AAPL	Info tech.	-0.001	(0.022)	4.10	-0.429	10.22	0.002	-5.91	-3.03
	IBM	IBM	Info tech.	-0.011	(0.027)	1.92	-0.133	8.97	0.021	-4.43	-2.11
	Microsoft	MSFT	Info tech.	-0.058	(0.029)	2.95	0.165	12.77	0.000	-4.76	-2.53
	Pfizer	PFE	Pharma	-0.058	(0.029)	1.91	-0.040	9.30	0.000	-3.75	-2.13
	Exxon Mobil	XOM	Oil & gas	-0.135	(0.035)	2.31	-0.273	17.24	0.000	-4.34	-2.24
FTSE 100	BP	BP	Oil & gas	-0.029	(0.025)	3.69	-0.504	14.31	0.000	-5.45	-2.89
	Halma	HLMA	Electronics	-0.061	(0.023)	2.54	0.101	6.11	0.004	-4.11	-2.52
	Marks & Spencer	MKS	Retail	0.026	(0.025)	3.85	-1.473	25.45	0.008	-4.79	-2.86
	Unilever	ULVR	Pers. goods	-0.040	(0.027)	1.92	0.295	9.49	0.282	-3.64	-2.01
	Vodafone	VOD	Commun.	-0.065	(0.028)	3.11	-0.390	10.73	0.000	-5.26	-2.60
EURO STOXX 50	Airbus	AIR	Aerospace	-0.010	(0.026)	5.10	-1.047	17.77	0.001	-5.82	-3.59
	Nokia	NOKIA	Technology	-0.001	(0.021)	6.70	-0.265	16.22	0.042	-7.65	-3.63
	Sanofi	SAN	Pharma	-0.061	(0.029)	2.41	-0.256	9.50	0.040	-4.27	-2.44
	TOTAL S.A.	FP	Oil & gas	-0.055	(0.025)	2.70	0.178	9.96	0.000	-4.44	-2.48

Note: The number of observations are 2686 for DAX stocks, 2759 for DJI stocks, 2699 for FTSE 100 stocks, and 2680 for EUROSTOXX 50 stocks. AR(1) coefficient is estimated by OLS, HAC robust standard errors in parentheses; sample mean of OLS residuals $\bar{\hat{\epsilon}}$ is approx. zero for all assets, $\hat{\sigma}^2$ is the sample variance, $\hat{\xi}$ is the sample skewness, $\hat{\kappa}$ is the sample kurtosis. p^{LB} are the p-values of the Ljung-Box test for 20 lags. 1% and 5% represent empirical quantiles.

4.4.2 Results

In order to perform a more accurate model comparison, value-at-risk is forecast by means of few other well-established models. First, three specifications of CAViaR of Engle and Manganelli (2004) are used, namely the symmetric absolute value (SAV), asymmetric slope (ASlope) and integrated GARCH (iGARCH). This class of models represents a one-step semi-parametric approach where the conditional quantile is also modelled directly. Second, GARCH of Bollerslev (1986) and apARCH of Ding et al. (1993) with skew-normal and skew-t distributed innovations are considered. Here, the quantile forecast relies on a conditional volatility forecast and a distributional assumption. Both of these model classes have been analyzed in the literature in detail and shown quite successful with the task (Nieto and Ruiz, 2016). Finally, there is a relatively new class of parametric models, generalized autoregressive score models (GAS) of Creal et al. (2013). GAS models are rather competitive when it comes to VaR forecasts (Ardia et al., 2019; Bernardi and Catania, 2016). Here, three specifications of the GAS are employed: normal, Student t and skew-t.

The results of model performances for the index data are summarized in Table 4.4 and the detailed results on backtesting and MCS procedure can be found in Tables 4.20–4.43 of the Appendix. The estimation results of the quantile autoregressions are presented in Tables 4.12–4.19 (see the Appendix). For the stocks data the results are presented in Tables 4.5 – 4.10 for different quantiles and validation sample sizes, and the overall model scores for the stock data are presented in Table 4.11. In each case, the inclusion of a model in the SSM is represented by a light-blue cell, which assigns a score of one to this model.[8] A dark-blue cell indicates a model with the lowest average loss. Number one in a light-blue cell means that the model passed all validation tests; number one in a dark-blue cell corresponds to the model with the lowest loss, but not all passed backtesting procedures. Finally, number two in a dark-blue cell indicates a model with both lowest loss and backtesting success.

First, I will discuss the results for the index data (Table 4.4). It is evident that the MCS procedure keeps almost all of the QREG-RM models in the superior set in every case. For DAX, QREG models with log bipower variation and median realized variance, both in logs and levels, appear to be the better choices when the sample is small. For $H = 1000$, however, skew-normal apARCH is a more suitable alternative.

For the Dow Jones the MCS procedure appears to be rather discriminative, but the QREG models in levels keep their seat at the table. Here, bipower variation, realized volatility and median realized variance contain the highest predictive power.

For FTSE 100, QREG models with implied volatility perform rather well for the 5% quantile, but so do the QREG-RM models. The latter tend to do better for the 1% quantile

[8] Here I refrain from adding a number one to a coloured cell for visual purposes

in logs. The competitors deliver inconsistent results, and no pattern is recognizable.

Finally, for EUROSTOXX 50, QREG models only manage to remain in the SSM but fail to score otherwise. The best alternatives here are HARQ-BV, apARCH, GARCH and GAS with Student t innovations.

Looking at overall scores, QREG models with median realized variance, both in logs and levels, and bipower variation are the best choices for this dataset, especially for the shorter validation periods.

As for single stocks dataset, the results are rather diverse for different periods and quantiles. In terms of MCS, QREG model consistently remain in the SSM with only a few exceptions. Score-wise, quantile regressions perform better for the 5% quantile and smaller sample sizes. For the 1% VaR, apARCH and GAS models appear to be more successful. CAViaR models become competitive only for $H = 1000$ and perform poorly otherwise. The same holds for GARCH models. The worst performing model overall is GAS with Student t innovations.

Looking at the overall scores in Table 4.11, it is evident that extending a models information set beyond past returns can be beneficial. Still, the HARQ model is rather successful on its own but holds the same score as asymmetric slope CAViaR. There is a definite trade-off between the amount of predictive information and the complexity of the model. Involving weakly exogenous regressors and applying a flexible semiparametric model has shown robustness so far. However, there is a definite need for further analysis, and clearer patterns have to be revealed.

4.5 Summary

In this paper, I compare a class of predictive quantile autoregressions with exogenous explanatory variables to a number of return-based models for VaR forecasting. I focus on modelling 1% and 5% conditional quantiles of a large set of stock and index returns. I propose a simple ranking system of forecasting performances based on the model confidence set procedure, lowest average loss criterion and a battery of backtests. The models with the best overall performance are heterogeneous quantile autoregression with bipower variation and median realized measures and additional exogenous regressors, such as implied volatility, downward semivariance, daily trading volume, day-of-the-week effect and various transformations of lagged returns. The estimation process is performed by means of quantile regression and is shown to be fast and quite simple. These models work better in smaller samples and are competitive against apARCH, GAS and CAViaR in larger ones. The analysis is performed on recent stock market data with a calm validation period, which normally complicates discrimination among different forecasting methods. While the role of additional information, i.e. market characteristics, liquidity variables, cannot be discounted, there is definitely a need for further research on relationships between value-at-risk and possible predictive regressors.

Table 4.4: Model selection and backtesting results for stock indices.

Model	DAX						Dow Jones						FTSE 100						EUROSTOXX 50						Total
	0.01			0.05			0.01			0.05			0.01			0.05			0.01			0.05			
	250	500	1000	250	500	1000	250	500	1000	250	500	1000	250	500	1000	250	500	1000	250	500	1000	250	500	1000	Score
QREG-RV	1					1			1		1		1		1		1	1	1						33
QREG-BV	1					1	2	1	2	1						2	1	1	1						**36**
QREG-MedRV	1			1	2	1			1			1	1	1	1		1	1	1						**37**
QREG-logRV	1	1											2	1		1	1	1							31
QREG-logBV	2	1		1									1			1	1	1							30
QREG-logMedRV	1	2			1								1	2	2	1	1	2							**36**
HARQ-RV			1			1	1	1							1			1		1	1				25
HARQ-BV	1		1			1		1					1		1			1	2	2	1	1	1		34
HARQ-MedRV			1	1		1		1							1	1		1	1	1					27
HARQ-logRV			1										1			1	1	1							25
HARQ-logBV			1										1				1	1			1				25
HARQ-logMedRV					1								1					1							22
QREG-IV																	2	1							24
QREG-logIV																1	1	1							23
CAViaR-SAV						1		1							1					1	1			1	14
CAViaR-ASlope								1						1	1				1	1	2				19
CAViaR-iGARCH						1		1					1	1	1					1	1			1	17
GARCH-SN	1	1	1			1	1	1						1					1	1	1			1	28
GARCH-ST			1		1	1		1						1	1					1	1		1	1	24
apARCH-SN	1	1	2			2	1	1	1				1						1	1	1			1	32
apARCH-ST					1	1	1	1					1	1	1		1	1		1	1			2	30
GAS-N		1	1			1	1							1						1	1			1	23
GAS-t	1	1	1	1	1									1		1			1	1	1		1	1	31
GAS-ST																				1					7

Note: The table summarizes the results for model confidence set procedure and backtesting for the stock index data. The forecasting of value-at-risk was performed for three rolling-window lengths, $\{250, 500, 1000\}$, and two quantile levels, 1% and 5%. Light blue cells mark the models included in the SSM and dark blue cells mark the models with the lowest average loss. A value of one represents a model passing all the backtests. A score of one is always added for a model with a lowest average loss.

Table 4.5: Model selection and backtesting results for stock returns, $\tau = 0.01$ and $H = 250$.

Model	DAX				Dow Jones					FTSE 100					EUROSTOXX 50				Score
	BMW	DTE	SAP	BAYN	AAPL	IBM	MSFT	PFE	XOM	BP	HLMA	MKS	ULVR	VOD	AIR	NOKIA	SAN	FP	
QREG-RV	1	1	1	2	1	1		1		1	1	1	2						25
QREG-BV	1	1	1	1	1	1	1			1	1	2	1		1	1			28
QREG-MedRV	2	1	1	1	1	1	1	1		1	1	1	1		2				**29**
QREG-logRV	1					1		1	1		1					1			23
QREG-logBV	1				1	1		1	1		1					1	1	1	27
QREG-logMedRV					1	1		1	1		1			1		1	1	1	27
HARQ-RV		1	1	1	1		2			1	1	1					1		21
HARQ-BV		1	1	1	1		1				1	1				1	1		19
HARQ-MedRV		1	2	1	1		1			1	1	1							18
HARQ-logRV		1			1						1	1		1					21
HARQ-logBV	1					1					2	1							21
HARQ-logMedRV					1	1			1		1	1							19
QREG-IV				1	1	1		1	2					1	1				22
QREG-logIV				1	1	2		1	1					1	1	1	2		26
CAViaR-SAV	1				1		1				1	1	1			1			17
CAViaR-ASlope	1				1		1			1	1	1	1			1			19
CAViaR-iGARCH							1				1	1	1			1			12
GARCH-SN	1	1	1		1	1	1	1		1	1	1							26
GARCH-ST	1	1	1	1	1		1			1	1	1	1						21
apARCH-SN	1	2	1		1	1	1	1		1	1	1	1			1	1		**30**
apARCH-ST	1	1			1	1	1			2	1	1				2			22
GAS-N	1				1						1	1				1			15
GAS-t	1	1	1		1		1		1	1	1	1				1		2	**30**
GAS-ST	1			1	1	1	1					1	1			1			17

Note: The table summarizes the results for model confidence set procedure and backtesting for the stocks data. The forecasting of value-at-risk was performed for a rolling-window length of 250 and a quantile level of 1%. Light blue cells mark the models included in the SSM and dark blue cells mark the models with the lowest average loss. A value of one represents a model passing all the backtests. A score of one is always added for a model with a lowest average loss.

Table 4.6: Model selection and backtesting results for stock returns, $\tau = 0.05$ and $H = 250$.

Model	DAX				Dow Jones					FTSE 100					EUROSTOXX 50				Score
	BMW	DTE	SAP	BAYN	AAPL	IBM	MSFT	PFE	XOM	BP	HLMA	MKS	ULVR	VOD	AIR	NOKIA	SAN	FP	
QREG-RV		1	1	1	1	1	1	1	1		1	1	1	1		1	1		**32**
QREG-BV	1	1	1	1	1	1	1		1		1	1		1	2	1	1		**33**
QREG-MedRV		1	1	1		1	1		1		1	1		1		1	1		28
QREG-logRV				2		1			2		1			1					25
QREG-logBV				1					1		1			1				1	23
QREG-logMedRV				1		1			1		1		1	1		1			25
HARQ-RV	1		1	1		1	2	2	1	1	1						1		28
HARQ-BV	1		1	1			1	1	1	1	1	2					1		**29**
HARQ-MedRV	1		1	1					1	2	1	1					1	1	27
HARQ-logRV			1							1	2	1		1		1	1		26
HARQ-logBV										1	1	1				2		1	24
HARQ-logMedRV											1	1				1			20
QREG-IV					1	2			1		1		1	1					22
QREG-logIV					1	1			1		1		1	1					21
CAViaR-SAV				1				1	1	1	1			1		1			25
CAViaR-ASlope			1				1		1	1	1		1			1	1		25
CAViaR-iGARCH									1				1			1			15
GARCH-SN				1					1										9
GARCH-ST		1					1		1										18
apARCH-SN									1	1	1		1						13
apARCH-ST						1	1		1	1	1		1						20
GAS-N									1										9
GAS-t			1		1		1		1	1						1		1	23
GAS-ST					1		1		1				1						13

Note: The table summarizes the results for model confidence set procedure and backtesting for the stocks data. The forecasting of value-at-risk was performed for a rolling-window length of 250 and a quantile level of 5%. Light blue cells mark the models included in the SSM and dark blue cells mark the models with the lowest average loss. A value of one represents a model passing all the backtests. A score of one is always added for a model with a lowest average loss.

Table 4.7: Model selection and backtesting results for stock returns, $\tau = 0.01$ and $H = 500$.

	DAX				Dow Jones					FTSE 100				EUROSTOXX 50				Score
	BMW	DTE	SAP	BAYN	AAPL	IBM	MSFT	PFE	XOM	BP	HLMA	MKS	ULVR	VOD	AIR	NOKIA	SAN	FP	
QREG-RV	1		1		1	1		1	1					1		1	1		27
QREG-BV	1		1		1	2	1		1					1	1	1	2	1	31
QREG-MedRV	1		1		1	1	1	1	1				1	1		1	1		29
QREG-logRV	1				1	1								1		1	1		23
QREG-logBV	1				1	1								1	1		1	1	24
QREG-logMedRV	1				1	1								1	1	1		1	24
HARQ-RV	1	1	1		1	1	1	2	1	1				1		1	1		**30**
HARQ-BV	1	1	1		1	1	2		2	1				1		1	1		**30**
HARQ-MedRV	1	1	2		1	1	1		1	2				1		1	1	1	**32**
HARQ-logRV	1	2	1		1									1		1	1		25
HARQ-logBV	2	1			2	1								1		1	1		27
HARQ-logMedRV	1	1			1	1	1			1						1	1		26
QREG-IV	1	1			1	1								1		1	1		24
QREG-logIV	1	1			1	1										2	1		24
CAViaR-SAV	1	1	1		1		1		1	1			1		2	1		1	28
CAViaR-ASlope	1	1	1		1		1		1	1	1		1			1		1	26
CAViaR-iGARCH		1				1	1			1	1				1	1		1	23
GARCH-SN	1	1	1		1	1	1		1	1						1	1	1	29
GARCH-ST	1	1	1		1		1		1	1			1			1	1	1	27
apARCH-SN	1	1	1		1	1	1		1		1				1	1	1	1	**30**
apARCH-ST	1	1	1		1	1	1		1		1					1	1		28
GAS-N	1	1	1	1	1	1	1		1	1	1	1		1	1	1		1	**33**
GAS-t		1	1				1		1	1	1			1		1		1	27
GAS-ST	1	1	1		1	1	1		1				1			1		1	26

Note: The table summarizes the results for model confidence set procedure and backtesting for the stocks data. The forecasting of value-at-risk was performed for a rolling-window length of 500 and a quantile level of 1%. Light blue cells mark the models included in the SSM and dark blue cells mark the models with the lowest average loss. A value of one represents a model passing all the backtests. A score of one is always added for a model with a lowest average loss.

Table 4.8: Model selection and backtesting results for stock returns, $\tau = 0.05$ and $H = 500$.

Model	DAX				Dow Jones					FTSE 100					EUROSTOXX 50				Score
	BMW	DTE	SAP	BAYN	AAPL	IBM	MSFT	PFE	XOM	BP	HLMA	MKS	ULVR	VOD	AIR	NOKIA	SAN	FP	
QREG-rv	1	1	1	2		1			1		1	1		1	1	1	1		**31**
QREG-bv	1	1	1	1	2	1	1		1		2	1			2	1	1		**34**
QREG-mrv	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	2	**35**
QREG-lrv		1				1								1			1	1	23
QREG-lbv	1	1				1											1	1	22
QREG-lmrv		1				1								1		1		1	23
HARQ-RV	1	1	1	1		1	1		1	1	1	1							26
HARQ-BV	1	1	1	1			2	2	1	1	1	1					1		**31**
HARQ-MedRV	1	1	1	1		1	1	1	1	1		1						1	29
HARQ-logRV		1	1	1						1	1	1						1	25
HARQ-logBV	1	1	1	1						1	1	1				2		1	27
HARQ-logMedRV		1	2							1	1	1						1	25
QREG-IV				1		2		1	1		1								24
QREG-logIV				1		1			2		1			1					24
CAViaR-SAV	1	1		1		1	1		1	1	1	1			1	1			28
CAViaR-ASlope		1	1		1	1	1			1	1	2			1	1			29
CAViaR-iGARCH	1		1		1				1	1	1	1				1			24
GARCH-SN	1			1	1				1	1	1								18
GARCH-ST	1	1	1	1	1		1		1	2	1				1				27
apARCH-SN									1	1	1								17
apARCH-ST		1	1		1		1	1	1	1	1				1				25
GAS-N																			10
GAS-t	1		1		1		1	1	1	1									22
GAS-ST					1		1												10

Note: The table summarizes the results for model confidence set procedure and backtesting for the stocks data. The forecasting of value-at-risk was performed for a rolling-window length of 500 and a quantile level of 5%. Light blue cells mark the models included in the SSM and dark blue cells mark the models with the lowest average loss. A value of one represents a model passing all the backtests. A score of one is always added for a model with a lowest average loss.

Table 4.9: Model selection and backtesting results for stock returns, $\tau = 0.01$ and $H = 1000$.

Model	DAX				Dow Jones					FTSE 100					EUROSTOXX 50				Score
	BMW	DTE	SAP	BAYN	AAPL	IBM	MSFT	PFE	XOM	BP	HLMA	MKS	ULVR	VOD	AIR	NOKIA	SAN	FP	
QREG-RV	1					1							1				1		22
QREG-BV	1		1			1							1	1			1		24
QREG-MedRV	1				1	1							1		1	1	1		25
QREG-logRV																			18
QREG-logBV	1																1		20
QREG-logMedRV					1												1		19
HARQ-RV			1		2	1	1			1	1				1	1			27
HARQ-BV	1		2		1	1		1		1	1				1	1		1	**29**
HARQ-MedRV					1	1	1			1	1				1	1			25
HARQ-logRV					1		1	1								1			21
HARQ-logBV			1		1		2									1		1	23
HARQ-logMedRV	1				1		1											1	22
QREG-IV	1		1			1									1	2	1		25
QREG-logIV						2	1						2		1	1	1		26
CAViaR-SAV	1		1		1	1	1		1	1					1	1	1	1	28
CAViaR-ASlope	1	1	1		1	1	1		1	1	2		1		1	1	1	2	**34**
CAViaR-iGARCH	1	1	1			1					1		1		1	1	1		25
GARCH-SN	1		1		1					1					2	1	1		25
GARCH-ST	1	1	1		1	1			1	1	1	1	1			1	2		**30**
apARCH-SN			1		1		1				1				1		1		24
apARCH-ST	1	2	1		1	1			1		1				1	1	1		26
GAS-N		1	1	1	1	1	1		1		1			1		1		1	28
GAS-t		1	1				1									1		1	22
GAS-ST		1	1		1		1				1					1	1	1	21

Note: The table summarizes the results for model confidence set procedure and backtesting for the stocks data. The forecasting of value-at-risk was performed for a rolling-window length of 1000 and a quantile level of 1%. Light blue cells mark the models included in the SSM and dark blue cells mark the models with the lowest average loss. A value of one represents a model passing all the backtests. A score of one is always added for a model with a lowest average loss.

Table 4.10: Model selection and backtesting results for stock returns, $\tau = 0.05$ and $H = 1000$.

Model	DAX				Dow Jones					FTSE 100					EUROSTOXX 50				Score
	BMW	DTE	SAP	BAYN	AAPL	IBM	MSFT	PFE	XOM	BP	HLMA	MKS	ULVR	VOD	AIR	NOKIA	SAN	FP	
QREG-RV	1		1	1	1						1				1		1		23
QREG-BV	2		1	1	2						1				1		1		**27**
QREG-MedRV	1		1	1	1						1				1		1		25
QREG-logRV	1																		18
QREG-logBV															1				19
QREG-logMedRV			1														1		20
HARQ-RV	1						1					1			1				22
HARQ-BV	1			1							1	1	1		1				24
HARQ-MedRV	1						1					1	1						22
HARQ-logRV			1												1				19
HARQ-logBV	1						1								1				20
HARQ-logMedRV	1		1												1				20
QREG-IV						1		1				1				1			21
QREG-logIV														1					18
CAViaR-SAV	1		1	1	1	1				1					2			1	**27**
CAViaR-ASlope	1		1	2		1	1		2	1					1			2	**29**
CAViaR-iGARCH	1			1	1				1						1			1	22
GARCH-SN	1			1	1				1	1					1			1	22
GARCH-ST	1			1	1			1	1	2		1			1		1	1	**28**
apARCH-SN	1			1					1	1	1		1						23
apARCH-ST	1	1	1		1	1	1	1	1	1	1				1				**29**
GAS-N									1										16
GAS-t	1		1		1		1		1								1	1	23
GAS-ST																			8

Note: The table summarizes the results for model confidence set procedure and backtesting for the stocks data. The forecasting of value-at-risk was performed for a rolling-window length of 1000 and a quantile level of 5%. Light blue cells mark the models included in the SSM and dark blue cells mark the models with the lowest average loss. A value of one represents a model passing all the backtests. A score of one is always added for a model with a lowest average loss.

Table 4.11: Overall model scores for stock returns data.

Model	$\tau = 0.01$			$\tau = 0.05$			Total score
	250	500	1000	250	500	1000	
QREG-RV	25	27	22	32	31	23	160
QREG-BV	28	31	24	33	34	27	**177**
QREG-MedRV	29	29	25	28	35	25	**171**
QREG-logRV	23	23	18	25	23	18	130
QREG-logBV	27	24	20	23	22	19	135
QREG-logMedRV	27	24	19	25	23	20	138
HARQ-RV	21	30	27	28	26	22	154
HARQ-BV	19	30	29	29	31	24	**162**
HARQ-MedRV	18	32	25	27	29	22	153
HARQ-logRV	21	25	21	26	25	19	137
HARQ-logBV	21	27	23	24	27	20	142
HARQ-logMedRV	19	26	22	20	25	20	132
QREG-IV	22	24	25	22	24	21	138
QREG-logIV	26	24	26	21	24	18	139
CAViaR-SAV	17	28	28	25	28	27	153
CAViaR-ASlope	19	26	34	25	29	29	**162**
CAViaR-iGARCH	12	23	25	15	24	22	121
GARCH-SN	26	29	25	9	18	22	129
GARCH-ST	21	27	30	18	27	28	151
apARCH-SN	30	30	24	13	17	23	137
apARCH-ST	22	28	26	20	25	29	150
GAS-N	15	33	28	9	10	16	111
GAS-t	30	27	22	23	22	23	147
GAS-ST	17	26	21	13	10	8	95

Appendix

4.5.1 Realized measures definitions

Assuming the log price variable $p(t)$ is driven by the stochastic process:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), \quad t \geq 0, \tag{4.10}$$

where the mean $\mu(t)$ is continuous and locally bounded, volatility $\sigma(t)$ is positive and cádlág, and $W(t)$ is standard Brownian motion. The last term represents a discrete jump component of the process with $dq(t) = 1$ when a jump occurs (and being zero otherwise) and $\kappa(t)$ the corresponding size of the jump. The common goal in practice is to estimate and predict the quadratic variation on a daily time scale:

$$\tilde{\sigma}^2(t) = \int_t^{t+1} \sigma^2(s)ds + \sum_{t \leq m_j \leq t+1} \kappa^2(m_j).$$

The first term of this decomposition is the so-called integrated variance, while the second one is the jump variation.

Given a sample of size $T(M+1)$ of T days with $M+1$ intraday observations each, one can define intraday returns as

$$r_{t,j} = p_{t,j} - p_{t,j-1}, \quad j = 1, \ldots, M, \ t = 1, \ldots, T. \tag{4.11}$$

This data can be used to compute various realized measures with different properties. Daily realized volatility, first introduced by Andersen and Bollerslev (1998), is defined as a sum of squared intraday returns,

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2, \quad t = 1, \ldots, T.$$

In order to separate the continuous and the jump components of quadratic variation, Barndorff-Nielsen and Shephard (2004, 2006) introduced bipower variation,

$$BV_t = \mu_1^{-2} \frac{M}{M-(k+1)} \sum_{j=k+2}^{M} |r_{t,j}||r_{t,j-k-1}| \quad t = 1, \ldots, T,$$

where $\mu_1 = \sqrt{2/\pi}$. This measure is a consistent estimate of the integrated volatility and can be used to tease out the jump component.

Finally, integrated volatility can also be consistently estimated by the median realized volatility of Andersen et al. (2012):

$$MedRV_t = \frac{\pi}{6-4\sqrt{3}+\pi}\Big(\frac{M}{M-2}\Big)\sum_{j=0}^{M-3}\mathrm{med}\Big(|r_{t,j}||r_{t,j+1}||r_{t,j+2}|\Big)^2, \quad t=1,\ldots,T.$$

Table 4.12: Quantile regression estimation results for DAX, full sample, $\tau = 1\%$.

	QREG								HARQ					
	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
Intercept	-0.0060*	-0.0043	-0.0072*	-0.0987*	-0.1029*	-0.1072*	-0.0066*	-0.0058*	-0.0066*	-0.0052*	-0.0068*	-0.1307*	-0.1303*	-0.1343*
RV^d	-4.5889*			-0.0111					-0.9680*			-0.0097*		
RV^w	-0.4020			-0.0057*					-0.3713			-0.0129*		
RV^m	-0.1658			-0.0056*					-0.6612*			0.0006		
BV^d		-2.1755*			-0.0022					-1.5464*			-0.0094*	
BV^w		-0.3260			-0.0065*					0.0083			-0.0133*	
BV^m		-0.4509*			-0.0057					-0.7898*			0.0011	
$MedRV^d$			-4.7243*			0.0003					-1.0738*			-0.0138*
$MedRV^w$			-0.3732			-0.0057*					-0.2125			-0.0083
$MedRV^m$			-0.3296			-0.0058*					-0.8101*			0.0001
r_{t-1}^2							0.1900*							
$\mathbb{I}(r_{t-1} < 0)$	0.0004	0.0003	0.0004	-0.0002	0.0014	0.0008	0.0018	0.0013						
$\|r_{t-1}\|$	0.8129*	0.2487*	0.8400*	0.2184*	0.1302*	0.1393*		0.1920*						
$\|r_{t-1}\|\mathbb{I}(r_{t-1} < 0)$	-1.1433*	-0.4152*	-1.1454*	0.0049	-0.2567	-0.1011		-0.3288						
RS^-	4.8860*	1.4766*	4.9743*	-0.3114	-0.0032	-0.0067*								
IV	-0.0007	-0.0015	0.0005	-0.0019	-0.0015	-0.0025	-0.0242*	-0.0243*						
Volume$\times 10^{-8}$	-0.0054*	-0.0036*	-0.0048*	-0.0086*	-0.0083*	-0.0074*	-0.0128*	-0.0133*						
Wednesday	0.0023	0.0028	0.0029	0.0004	0.0000	0.0001	0.0004	0.0002						
Friday	0.0024*	0.0031*	0.0029*	0.0039*	0.0038*	0.0045*	0.0039*	0.0035						

Note: The table summarizes estimated coefficients of the quantile regression specifications. Significance at 10% is marked by an asterisk.

Table 4.13: Quantile regression estimation results for DAX, full sample, $\tau = 5\%$

	QREG								HARQ					
	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
Intercept	0.0010	-0.0006	-0.0007	-0.0785*	-0.0814*	-0.0861*	-0.0020	-0.0023	-0.0026*	-0.0027*	-0.0020	-0.0947*	-0.0951*	-0.0981
RV^d	0.1146			0.0047					-0.7662*			-0.0077*		
RV^w	-0.7504*			-0.0038					-0.4681			-0.0023		
RV^m	-0.3165			-0.0053*					-0.3548			-0.0063*		
BV^d		-0.9605			-0.0010					-0.8816*			-0.0087*	
BV^w		-0.4563			-0.0026*					-0.2628			-0.0008	
BV^m		-0.4434			-0.0060					-0.5126			-0.0066*	
MedRV^d			-0.1026			-0.0017					-0.7305*			-0.0086
MedRV^w			-0.8468*			-0.0037					-0.5678			-0.0019
MedRV^m			-0.3532			-0.0056*					-0.4369			-0.0060
r_{t-1}^2							0.0441							
$\mathbf{1}(r_{t-1} < 0)$	-0.0021	-0.0022	-0.0018	-0.0008	-0.0007	-0.0012	-0.0019	-0.0022						
$\lvert r_{t-1} \rvert$	0.0387	0.1849	0.0936	0.0175	0.1534	0.1248		-0.0115						
$\lvert r_{t-1} \rvert \mathbf{1}(r_{t-1} < 0)$	0.0755	-0.0926	-0.0010	-0.0093	-0.1541	-0.1272		-0.1324						
RS^-	-0.9092	0.1514	-0.7775	-0.0088*	-0.0042	-0.0036								
IV	-0.0002	0.0006	0.0010	-0.0029	-0.0032	-0.0020	-0.0204*	-0.0196						
Volume$\times 10^{-8}$	-0.0022*	-0.0018	-0.0013	-0.0037*	-0.0031*	-0.0027*	-0.0062*	-0.0055						
Wednesday	0.0002	0.0013	0.0010	0.0005	0.0004	0.0004	-0.0011	-0.0001						
Friday	-0.0006	-0.0005	-0.0009	0.0006	0.0006	0.0004	-0.0003	-0.0002						

Notes: The table summarizes estimated coefficients of the quantile regression specifications. Significance at 10% is marked by an asterisk.

Table 4.14: Quantile regression estimation results for Dow Jones, full sample, $\tau = 1\%$

	QREG								HARQ					
	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
Intercept	0.0091*	0.0097*	0.0106*	0.0237*	0.0113	0.0159	-0.0174*	-0.0190*	-0.0037*	-0.0038*	-0.0049*	-0.1228*	-0.1263*	-0.1293*
RV^d	-0.5606			0.0014					-0.9058*			-0.0087*		
RV^w	0.2209			-0.0026					-1.8229*			-0.0125*		
RV^m	0.8195*			0.0098*					0.2903*			0.0015		
BV^d		-0.4029			-0.0013					-1.3870*			-0.0092*	
BV^w		0.2074			-0.0018					-1.7921*			-0.0129*	
BV^m		0.7402*			0.0063*					0.3321			0.0022	
MedRV^d			-0.1506			0.0006					-1.0718*			-0.0115*
MedRV^w			0.0492			-0.0025					-1.8996*			-0.0111*
MedRV^m			0.9099*			0.0066*					0.3950*			0.0022
r_{t-1}^2							-0.0025							
$\mathbf{1}(r_{t-1} < 0)$	-0.0025	-0.0005	-0.0012	-0.0006	-0.0009	-0.0007	-0.0007	-0.0011						
$\lvert r_{t-1} \rvert$	0.2474*	0.2492*	0.2392*	0.0241	0.1137	0.1337		0.1221						
$\lvert r_{t-1} \rvert \mathbf{1}(r_{t-1} < 0)$	0.1733	-0.1514	-0.0309	0.0715	0.0197	0.0098		0.1266						
RS^-	0.7787	0.8192*	0.6616	-0.0004	0.0022	0.0014								
IV	-0.0362*	-0.0355*	-0.0365*	-0.0410*	-0.0380*	-0.0392*	-0.0336*	-0.0348*						
$\mathrm{Volume} \times 10^{-8}$	0.0015*	0.0007	0.0007	-0.0013*	-0.0014*	-0.0018*	-0.0013	-0.0008						
Wednesday	0.0009	0.0015*	0.0008	0.0004	0.0009	0.0010	0.0011	0.0011						
Friday	0.0017*	0.0020*	0.0018*	0.0002	0.0011	0.0009	0.0016*	0.0013						

Notes: The table summarizes estimated coefficients of the quantile regression specifications. Significance at 10% is marked by an asterisk.

Table 4.15: Quantile regression estimation results for Dow Jones, full sample, $\tau = 5\%$

	QREG								HARQ					
	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
Intercept	0.0125*	0.0129*	0.0139*	0.0271*	0.0194*	0.0272*	-0.0125*	-0.0131*	-0.0016	-0.0013	-0.0011	-0.0894*	-0.1002*	-0.1028*
RV^d	-0.2772*			0.0009					-0.6373*			-0.0072*		
RV^w	0.0995			0.0002					-0.8489*			-0.0050*		
RV^m	1.0762*			0.0077*					-0.1931			-0.0025		
BV^d		-0.0874			-0.0012					-0.7240*			-0.0092*	
BV^w		-0.1329			-0.0010					-0.8388*			-0.0039*	
BV^m		1.3927*			0.0076*					-0.4516			-0.0031	
MedRV^d			-0.1749			-0.0011					-0.5037*			-0.0082*
MedRV^w			-0.1528			-0.0002					-1.2798*			-0.0075*
MedRV^m			1.4493*			0.0079*					-0.2521			-0.0011
r_{t-1}^2							0.0983*							
$\mathbf{1}(r_{t-1} < 0)$	-0.0013*	-0.0014*	-0.0015*	0.0002	-0.0001	-0.0002	0.0000	0.2047*						
$\|r_{t-1}\|$	0.2640*	0.2230*	0.2769*	0.2138	0.2880*	0.2629*		-0.0005						
$\|r_{t-1}\|\mathbf{1}(r_{t-1} < 0)$	0.1221	0.1832*	0.0716	-0.1289	-0.2155	-0.1945		-0.1461						
RS^-	0.2486	0.0615	0.2872	-0.0005	0.0010	0.0011								
IV	-0.0340*	-0.0338*	-0.0346*	-0.0373*	-0.0346*	-0.0360*	-0.0264*	-0.0275*						
Volume$\times 10^{-8}$	0.0008*	0.0006*	0.0004*	-0.0002	-0.0005*	-0.0006*	-0.0004*	-0.0005*						
Wednesday	-0.0001	0.0000	-0.0002	-0.0001	-0.0004	-0.0006	0.0000	0.0004						
Friday	-0.0004	-0.0002	-0.0006	-0.0002	-0.0005	-0.0005	0.0000	0.0001						

Notes: The table summarizes estimated coefficients of the quantile regression specifications. Significance at 10% is marked by an asterisk.

Table 4.16: Quantile regression estimation results for FTSE 100, full sample, $\tau = 1\%$

	QREG								HARQ					
	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
Intercept	0.0004	-0.0001	0.0013	-0.0263*	-0.0259*	-0.0186*	-0.0161*	-0.0168*	-0.0058*	-0.0057*	-0.0053*	-0.1093*	-0.1063*	-0.1079*
RV^d	0.4609			-0.0004					-1.2301*			-0.0098*		
RV^w	-0.6798			-0.0079*					-1.0584*			-0.0109*		
RV^m	0.7741			0.0114*					0.3444			0.0032		
BV^d		-1.9137			-0.0102*					-1.4147*			-0.0081*	
BV^w		-0.4244			-0.0071*					-0.9049*			-0.0100*	
BV^m		0.6803			0.0106*					0.2848			0.0012	
MedRV^d			-0.4058			-0.0166*					-1.3094*			-0.0125*
MedRV^w			-0.3967			0.0017					-0.6435			-0.0006
MedRV^m			0.7228			0.0067*					-0.1820			-0.0038
r_{t-1}^2							0.2369*							
$\mathbf{1}(r_{t-1} < 0)$	0.0009	0.0004	0.0003	0.0002	-0.0009	-0.0027*	-0.0005	-0.0003						
$\lvert r_{t-1}\rvert$	0.2225	0.3973	0.2571	0.2620*	0.2823*	0.1859*		0.1743						
$\lvert r_{t-1}\rvert\mathbf{1}(r_{t-1} < 0)$	-0.1689	-0.5611	-0.1581	-0.3795*	-0.4073*	-0.2219		-0.4418						
RS^-	-1.8694	1.8554	-0.8493	-0.0041	0.0051*	0.0077*								
IV	-0.0149*	-0.0181*	-0.0169*	-0.0213*	-0.0212*	-0.0224*	-0.0231*	-0.0239*						
Volume$\times 10^{-8}$	0.0000	0.0002	0.0001	0.0000	0.0000	0.0001	-0.0003*	-0.0002*						
Wednesday	0.0014	0.0008	0.0007	-0.0002	-0.0005	-0.0016*	-0.0011	-0.0005						
Friday	0.0007	0.0001	0.0010	0.0004	0.0004	0.0006	0.0011	0.0009						

Notes: The table summarizes estimated coefficients of the quantile regression specifications. Significance at 10% is marked by an asterisk.

Table 4.17: Quantile regression estimation results for FTSE 100, full sample, $\tau = 5\%$

	QREG								HARQ					
	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
Intercept	0.0020	0.0016	0.0017	-0.0078	-0.0037	-0.0057	-0.0095*	-0.0096*	-0.0013	-0.0013	-0.0014	-0.0754*	-0.0753*	-0.0748*
RV^d	-2.4995*			-0.0009					-0.7839*			-0.0059*		
RV^w	-0.2791			-0.0005					-0.7680*			-0.0059*		
RV^m	0.9712*			0.0046*					-0.0571			-0.0006		
BV^d		-1.5446*			-0.0058*					-0.8040*			-0.0056*	
BV^w		-0.3362			-0.0004					-0.9315*			-0.0060*	
BV^m		1.0080*			0.0051*					0.0649			-0.0008	
MedRV^d			-2.6128*			-0.0124*					-0.8569*			-0.0080*
MedRV^w			-0.2101			0.0016					-0.8253*			-0.0011
MedRV^m			0.8023*			0.0047*					-0.0124			-0.0028
r_{t-1}^2							-0.0414							
$\mathbf{1}(r_{t-1} < 0)$	-0.0008	-0.0005	-0.0005	-0.0008	-0.0017*	-0.0019*	-0.0023*	-0.0021*						
$\lvert r_{t-1}\rvert$	0.4122*	0.3532*	0.4324*	0.0638	0.0633	0.1706*		-0.0471						
$\lvert r_{t-1}\rvert\mathbf{1}(r_{t-1} < 0)$	-0.5665*	-0.4060*	-0.5685*	0.0364	-0.0070	-0.0412		0.0531						
RS^-	2.8029*	1.5180*	2.8865*	-0.0023	0.0024	0.0069*								
IV	-0.0140*	-0.0141*	-0.0126*	-0.0165*	-0.0161*	-0.0167*	-0.0140*	-0.0140*						
$\mathrm{Volume}\times 10^{-8}$	-0.0001	-0.0001*	-0.0001	-0.0001	-0.0001	-0.0002*	-0.0002*	-0.0002*						
Wednesday	0.0003	0.0005	0.0002	0.0008	0.0009	0.0002	0.0013	0.0012						
Friday	0.0000	0.0003	-0.0001	0.0003	0.0003	0.0008	0.0007	0.0006						

Notes: The table summarizes estimated coefficients of the quantile regression specifications. Significance at 10% is marked by an asterisk.

Table 4.18: Quantile regression estimation results for EUROSTOXX 50, full sample, $\tau = 1\%$

	QREG								HARQ					
	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
Intercept	-0.0009	-0.0018	-0.0011	-0.1149*	-0.1007*	-0.1075*	-0.0730*	-0.0712*	-0.0016	0.0009	-0.0022	-0.1523*	-0.1481*	-0.1495*
RV^d	-3.0618			-0.0150*					-1.3219*			-0.0093*		
RV^w	-1.0558*			-0.0136*					-0.9344*			-0.0125*		
RV^m	0.4449			0.0053					-0.3418			-0.0046		
BV^d		-2.0132*			0.0057					-1.2760*			-0.0077*	
BV^w		-1.2499*			-0.0118*					-1.4639*			-0.0168*	
BV^m		0.5194			0.0075*					-0.3055			-0.0006	
MedRV^d			-3.2700*			-0.0033					-1.2457*			-0.0083*
MedRV^w			-1.2667*			-0.0118*					-1.3140*			-0.0168*
MedRV^m			0.4927			0.0072*					-0.1761			0.0001
r_{t-1}^2							-0.1001							
$\mathbf{1}(r_{t-1} < 0)$	-0.0005	0.0015	-0.0008	-0.0017	-0.0013	-0.0012	-0.0020	-0.0025						
$\lvert r_{t-1} \rvert$	0.4220	0.2219	0.4594	0.3161	-0.0096	0.0955		-0.2463						
$\lvert r_{t-1} \rvert \mathbf{1}(r_{t-1} < 0)$	-0.7281*	-0.6382	-0.8210*	-0.3555	0.1438	0.0097		0.2496						
RS^-	3.1747	1.5570*	3.3907	0.0094	-0.0083	-0.0015								
IV	-0.0611*	-0.0449*	-0.0531*	-0.0174*	-0.0262*	-0.0228*	-0.0339	-0.0331*						
Volume$\times 10^{-8}$	0.0046*	0.0019	0.0042*	0.0021	0.0013	0.0032*	-0.0022	-0.0014						
Wednesday	0.0016	0.0023	0.0016	-0.0025*	-0.0015	-0.0022	-0.0009	-0.0013						
Friday	0.0013	0.0003	0.0019*	0.0007	-0.0017	0.0004	0.0001	0.0002						

Notes: The table summarizes estimated coefficients of the quantile regression specifications. Significance at 10% is marked by an asterisk.

Table 4.19: Quantile regression estimation results for EUROSTOXX 50, full sample, $\tau = 5\%$

	QREG								HARQ					
	RV	BV	MedRV	logRV	logBV	logMedRV	IV1	IV2	RV	BV	MedRV	logRV	logBV	logMedRV
Intercept	0.0029*	0.0023	0.0026	-0.0779*	-0.0815*	-0.0787*	-0.0466*	-0.0489*	-0.0016	-0.0006	-0.0019	-0.1114*	-0.1106*	-0.1099*
RV^d	-0.8507			-0.0015					-0.5104*			-0.0057*		
RV^w	-0.4782			-0.0053*					-0.6003			-0.0057*		
RV^m	-0.0177			-0.0041*					-0.4578			-0.0084*		
BV^d		-0.8097*			-0.0009					-0.4965*			-0.0064*	
BV^w		-0.9600*			-0.0052*					-0.8994*			-0.0046	
BV^m		-0.0630			-0.0034					-0.4415			-0.0084*	
MedRV^d			-0.7258			-0.0050					-0.4615*			-0.0071*
MedRV^w			-0.8559*			-0.0052*					-0.8744*			-0.0036
MedRV^m			-0.1280			-0.0028					-0.3936			-0.0081*
r_{t-1}^2							0.1600*							
$\mathbf{1}(r_{t-1} < 0)$	-0.0020*	-0.0016	-0.0026*	-0.0010	-0.0010	-0.0013	-0.0004	-0.0004						
$\|r_{t-1}\|$	0.2735*	0.2701*	0.2372*	0.2513	0.2549	0.2485*		0.2324						
$\|r_{t-1}\|\mathbf{1}(r_{t-1} < 0)$	-0.2434	-0.3141	-0.2012	-0.3660*	-0.0007	-0.3841*		-0.3194						
RS^-	0.8746	0.4256	0.7008	0.0010	-0.3212	0.0030								
IV	-0.0440*	-0.0213	-0.0282*	-0.0110*	-0.0107*	-0.0095*	-0.0214*	-0.0224*						
$\text{Volume}\times 10^{-8}$	-0.0027	-0.0013	-0.0028*	-0.0018	-0.0016	-0.0009	-0.0014	-0.0016						
Wednesday	0.0016*	0.0012*	0.0013*	0.0015*	0.0015*	0.0013	0.0015	0.0019*						
Friday	-0.0016	-0.0003	-0.0014	-0.0013	-0.0011	-0.0014	-0.0010	-0.0003						

Notes: The table summarizes estimated coefficients of the quantile regression specifications. Significance at 10% is marked by an asterisk.

The following Tables 4.20 – 4.43 contain the models included in the SSM, their corresponding average loss values and p-values of the backtests. UC stands for unconditional coverage test; CC for conditional coverage; DQ for dynamic quantile test with three lags; DB_1 and DB_7 represent two specifications of the dynamic binary test described in Section 4.3.2. The tables represent the results for the index dataset.[9]

Table 4.20: MCS and backtesting results for DAX 30, $\tau = 1\%$, $H = 250$

Model	Loss×10^4	UC	CC	DQ	DB_1	DB_7
QREG-RV	1.986	0.278	0.553	0.987	0.563	0.763
QREG-BV	1.937	0.278	0.553	0.992	0.563	0.763
QREG-MedRV	2.003	0.278	0.553	0.991	0.563	0.762
QREG-logRV	1.819	0.380	0.638	0.587	0.657	0.841
QREG-logBV	**1.809**	0.742	0.932	1.000	0.945	0.989
QREG-logMedRV	1.897	0.380	0.638	0.530	0.657	0.183
HARQ-BV	1.978	0.278	0.553	0.994	0.563	0.764
HARQ-logRV	2.040	0.162	0.339	0.000	0.357	0.024
HARQ-logBV	1.972	0.380	0.638	0.000	0.657	0.692
HARQ-logMedRV	1.959	0.380	0.638	0.000	0.657	0.010
QREG-IV	2.236	0.025	0.081	0.928	0.657	0.010
GARCH-SN	2.053	0.278	0.553	0.992	0.563	0.763
apARCH-SN	1.896	0.278	0.553	0.996	0.563	0.765
GAS-ST	2.033	0.278	0.553	0.994	0.563	0.764

Table 4.21: MCS and backtesting results for DAX 30, $\tau = 5\%$, $H = 250$

Model	Loss×10^4	UC	CC	DQ	DB_1	DB_7
QREG-RV	6.387	0.037	0.097	0.359	0.119	0.233
QREG-BV	6.423	0.037	0.097	0.476	0.119	0.234
QREG-MedRV	**6.336**	0.083	0.181	0.451	0.231	0.402
QREG-logRV	6.575	0.669	0.396	0.076	0.844	0.187
QREG-logBV	6.473	0.669	0.886	0.365	0.844	0.173
QREG-logMedRV	6.524	0.079	0.191	0.008	0.181	0.007
HARQ-RV	6.502	0.083	0.181	0.669	0.231	0.401
HARQ-BV	6.499	0.083	0.181	0.636	0.231	0.403
HARQ-MedRV	6.492	0.453	0.530	0.911	0.746	0.900
HARQ-logRV	6.829	0.329	0.621	0.017	0.555	0.042
HARQ-logBV	6.651	0.215	0.457	0.025	0.406	0.052
HARQ-logMedRV	6.523	0.329	0.621	0.143	0.555	0.045
QREG-IV	7.343	0.004	0.016	0.540	0.018	0.046
QREG-logIV	7.331	0.000	0.001	0.246	0.001	0.003
CAViaR-ASlope	6.747	0.001	0.004	0.030	0.005	0.014
GARCH-SN	6.735	0.037	0.097	0.803	0.119	0.234
GARCH-ST	6.581	0.083	0.181	0.865	0.231	0.401
apARCH-SN	6.486	0.083	0.181	0.558	0.231	0.106
apARCH-ST	6.548	0.083	0.181	0.564	0.231	0.004
GAS-N	6.926	0.004	0.016	0.539	0.018	0.046
GAS-t	6.617	0.453	0.530	0.678	0.746	0.225

Table 4.22: MCS and backtesting results for DAX 30, $\tau = 1\%$, $H = 500$

Model	Loss×10^4	UC	CC	DQ	DB_1	DB_7
QREG-logRV	2.305	0.641	0.869	0.944	0.895	0.396
QREG-logBV	2.286	0.331	0.613	0.974	0.628	0.814
QREG-MedRV	2.381	0.028	0.090	0.864	0.092	0.190
QREG-RV	2.397	0.028	0.090	0.864	0.092	0.190
QREG-BV	2.418	0.028	0.090	0.862	0.092	0.189
QREG-logMedRV	**2.266**	0.641	0.869	0.957	0.895	0.377
HARQ-BV	2.394	0.028	0.090	0.860	0.092	0.189
HARQ-MedRV	2.411	0.028	0.090	0.858	0.092	0.189
HARQ-logRV	2.621	0.397	0.000	0.000	0.680	0.035
HARQ-logBV	2.508	0.663	0.845	0.001	0.893	0.006
HARQ-logMedRV	2.471	1.000	0.951	0.001	0.990	0.028
GARCH-SN	2.544	0.331	0.613	0.997	0.628	0.818
apARCH-SN	2.454	0.331	0.613	0.127	0.628	0.817
GAS-t	2.733	0.641	0.869	0.998	0.895	0.971
GAS-N	2.856	0.641	0.869	0.998	0.895	0.731

Table 4.23: MCS and backtesting results for DAX 30, $\tau = 5\%$, $H = 500$

Model	Loss×10^4	UC	CC	DQ	DB_1	DB_7
QREG-RV	8.592	0.027	0.055	0.559	0.090	0.184
QREG-BV	8.593	0.027	0.055	0.561	0.090	0.184
QREG-MedRV	**8.589**	0.131	0.163	0.831	0.323	0.518
QREG-logRV	8.817	0.838	0.234	0.077	0.949	0.075
QREG-logBV	8.799	0.838	0.926	0.285	0.949	0.059
QREG-logMedRV	8.827	0.685	0.842	0.371	0.922	0.198
HARQ-RV	8.784	0.027	0.055	0.585	0.090	0.183
HARQ-BV	8.706	0.027	0.055	0.585	0.090	0.184
HARQ-MedRV	8.712	0.027	0.055	0.596	0.090	0.184
HARQ-logRV	9.045	0.836	0.967	0.115	0.937	0.063
HARQ-logBV	8.955	1.000	0.970	0.067	0.949	0.047
HARQ-logMedRV	8.894	0.836	0.967	0.302	0.885	0.221
CAViaR-ASlope	9.074	0.007	0.018	0.282	0.027	0.066
GARCH-SN	9.034	0.082	0.121	0.476	0.147	0.277
GARCH-ST	8.941	0.199	0.207	0.545	0.323	0.518
apARCH-SN	8.871	0.082	0.121	0.490	0.147	0.193
apARCH-ST	8.879	0.131	0.163	0.627	0.224	0.391
GAS-t	9.406	0.836	0.967	0.576	0.885	0.542

[9] Similar tables for the stocks dataset can be provided upon request.

Table 4.24: MCS and backtesting results for DAX 30, $\tau = 1\%$, $H = 1000$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	3.013	0.139	0.157	0.002	0.325	0.516
QREG-BV	2.998	0.231	0.399	0.084	0.475	0.681
QREG-MedRV	2.989	0.079	0.112	0.085	0.208	0.367
QREG-logRV	3.069	0.043	0.012	0.000	0.124	0.025
QREG-logBV	3.111	0.079	0.016	0.000	0.208	0.016
QREG-logMedRV	2.966	0.079	0.016	0.000	0.208	0.005
HARQ-RV	2.820	0.314	0.573	0.850	0.603	0.798
HARQ-BV	2.819	0.314	0.573	0.849	0.603	0.798
HARQ-MedRV	2.846	0.314	0.573	0.835	0.603	0.797
HARQ-logRV	3.012	0.510	0.755	0.308	0.803	0.370
HARQ-logBV	2.962	0.510	0.755	0.460	0.803	0.523
HARQ-logMedRV	2.983	0.754	0.258	0.023	0.939	0.121
QREG-IV	3.324	0.005	0.013	0.000	0.019	0.016
QREG-logIV	3.397	0.000	0.002	0.000	0.002	0.002
CAViaR-SAV	2.914	0.510	0.117	0.092	0.803	0.932
CAViaR-ASlope	2.859	0.030	0.094	0.690	0.096	0.196
CAViaR-iGARCH	3.035	0.510	0.117	0.092	0.803	0.932
GARCH-SN	2.967	0.538	0.263	0.382	0.812	0.935
GARCH-ST	2.956	0.746	0.875	0.900	0.943	0.989
apARCH-SN	**2.791**	0.510	0.755	0.100	0.803	0.516
apARCH-ST	2.880	0.510	0.755	0.091	0.803	0.456
GAS-N	3.000	0.538	0.715	0.978	0.812	0.932
GAS-t	3.085	0.079	0.112	0.138	0.208	0.130

Table 4.25: MCS and backtesting results for DAX 30, $\tau = 5\%$, $H = 1000$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	10.751	0.885	0.368	0.727	0.936	0.985
QREG-BV	10.728	1.000	0.333	0.656	0.950	0.991
QREG-MedRV	10.796	0.666	0.189	0.310	0.854	0.941
QREG-logRV	10.699	0.159	0.086	0.003	0.338	0.027
QREG-logBV	10.775	0.475	0.209	0.033	0.720	0.145
QREG-logMedRV	10.770	0.320	0.208	0.008	0.562	0.032
HARQ-RV	10.560	0.461	0.132	0.537	0.739	0.879
HARQ-BV	10.572	0.461	0.329	0.827	0.739	0.887
HARQ-MedRV	10.599	0.660	0.208	0.643	0.873	0.942
HARQ-logRV	10.643	0.773	0.173	0.065	0.903	0.039
HARQ-logBV	10.595	0.885	0.154	0.053	0.936	0.045
HARQ-logMedRV	10.672	0.257	0.092	0.012	0.483	0.005
QREG-IV	11.432	0.093	0.032	0.000	0.220	0.000
QREG-logIV	11.455	0.051	0.055	0.000	0.135	0.000
CAViaR-SAV	10.662	0.178	0.254	0.765	0.399	0.606
CAViaR-ASlope	10.527	0.033	0.042	0.395	0.103	0.206
CAViaR-iGARCH	10.781	0.375	0.263	0.335	0.657	0.836
GARCH-SN	10.693	0.885	0.154	0.266	0.936	0.971
GARCH-ST	10.707	0.475	0.209	0.468	0.720	0.857
apARCH-SN	**10.403**	0.375	0.511	0.840	0.657	0.397
apARCH-ST	10.439	0.770	0.862	0.966	0.918	0.785
GAS-N	10.825	0.557	0.702	0.739	0.812	0.912
GAS-t	11.022	0.204	0.428	0.112	0.407	0.017

Table 4.26: MCS and backtesting results for Dow Jones, $\tau = 1\%$, $H = 250$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	1.059	0.742	0.932	1.000	0.945	0.009
QREG-BV	1.039	0.742	0.932	1.000	0.945	0.979
QREG-MedRV	**1.044**	0.758	0.919	0.998	0.936	0.073
QREG-logRV	1.800	0.005	0.010	0.000	0.019	0.000
QREG-logBV	1.842	0.005	0.016	0.000	0.019	0.000
QREG-logMedRV	1.711	0.019	0.025	0.000	0.059	0.000
HARQ-RV	1.364	0.278	0.553	0.996	0.563	0.765
HARQ-logRV	2.333	0.000	0.000	0.000	0.000	0.000
HARQ-logBV	1.873	0.001	0.004	0.000	0.005	0.000
HARQ-logMedRV	1.784	0.162	0.339	0.099	0.357	0.102
QREG-IV	1.709	0.000	0.000	0.000	0.000	0.000
QREG-logIV	1.716	0.000	0.000	0.000	0.000	0.000
GARCH-SN	1.420	0.758	0.919	0.932	0.936	0.981
apARCH-SN	1.209	0.742	0.932	1.000	0.945	0.990
apARCH-ST	1.294	0.742	0.932	1.000	0.945	0.990
GAS-N	1.599	0.758	0.919	0.888	0.936	0.979
GAS-t	1.765	0.380	0.638	0.636	0.657	0.049
GAS-ST	1.946	0.758	0.919	0.857	0.936	0.056

Table 4.27: MCS and backtesting results for Dow Jones, $\tau = 5\%$, $H = 250$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	3.441	0.215	0.005	0.000	0.406	0.000
QREG-BV	**3.399**	0.215	0.032	0.001	0.406	0.000
QREG-MedRV	3.462	0.024	0.001	0.000	0.064	0.000
QREG-IV	4.275	0.000	0.000	0.000	0.009	0.000
QREG-logIV	4.514	0.000	0.000	0.000	0.000	0.000

Table 4.28: MCS and backtesting results for Dow Jones, $\tau = 1\%$, $H = 500$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	1.452	0.397	0.149	0.039	0.680	0.276
QREG-BV	**1.437**	0.397	0.149	0.039	0.680	0.285
QREG-MedRV	1.443	0.215	0.113	0.003	0.448	0.029
QREG-logRV	1.665	0.215	0.407	0.000	0.448	0.003
QREG-logBV	1.698	0.106	0.230	0.000	0.260	0.002
QREG-logMedRV	1.686	0.215	0.128	0.000	0.448	0.000
HARQ-RV	1.989	0.331	0.613	0.997	0.628	0.817
HARQ-BV	2.059	0.641	0.869	1.000	0.895	0.973
HARQ-MedRV	2.025	0.331	0.613	0.996	0.628	0.817
HARQ-logRV	2.549	0.000	0.000	0.000	0.000	0.000
HARQ-logBV	2.467	0.008	0.021	0.000	0.027	0.000
QREG-IV	1.661	0.048	0.115	0.000	0.134	0.000
QREG-logIV	1.662	0.008	0.016	0.000	0.027	0.000
CAViaR-SAV	2.207	0.125	0.306	0.963	0.313	0.506
CAViaR-ASlope	2.180	0.125	0.306	0.966	0.313	0.507
CAViaR-iGARCH	2.237	0.125	0.306	0.962	0.313	0.506
GARCH-SN	2.119	1.000	0.951	0.943	0.990	0.170
GARCH-ST	2.208	1.000	0.951	0.936	0.990	0.180
apARCH-SN	2.056	0.641	0.869	0.998	0.895	0.972
apARCH-ST	2.162	0.641	0.869	0.997	0.895	0.971
GAS-N	2.412	1.000	0.951	0.986	0.989	0.056
GAS-t	2.433	0.397	0.632	0.591	0.680	0.015
GAS-ST	2.583	0.663	0.845	0.906	0.893	0.045

Table 4.29: MCS and backtesting results for Dow Jones, $\tau = 5\%$, $H = 500$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	**4.989**	0.168	0.001	0.000	0.345	0.006
QREG-BV	5.010	0.117	0.001	0.000	0.259	0.001
QREG-MedRV	5.007	0.052	0.000	0.000	0.132	0.000
QREG-logRV	5.539	0.530	0.000	0.000	0.799	0.003
QREG-logBV	5.462	0.546	0.000	0.000	0.771	0.000
QREG-logMedRV	5.527	0.235	0.000	0.000	0.445	0.000
QREG-IV	5.374	0.003	0.000	0.000	0.009	0.000
QREG-logIV	5.574	0.000	0.000	0.000	0.002	0.000

Table 4.30: MCS and backtesting results for Dow Jones, $\tau = 1\%$, $H = 1000$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	1.621	0.538	0.715	0.980	0.812	0.921
QREG-BV	1.604	0.139	0.157	0.278	0.325	0.521
QREG-MedRV	1.627	0.362	0.556	0.214	0.646	0.290
QREG-logRV	1.747	0.139	0.020	0.000	0.325	0.001
QREG-logBV	1.733	0.005	0.000	0.000	0.019	0.000
QREG-logMedRV	1.746	0.005	0.014	0.000	0.019	0.000
HARQ-RV	2.298	0.538	0.263	0.059	0.812	0.932
HARQ-BV	2.270	0.754	0.258	0.039	0.939	0.986
HARQ-MedRV	2.273	0.754	0.258	0.038	0.939	0.986
QREG-IV	**1.716**	0.000	0.000	0.000	0.000	0.000
QREG-logIV	1.717	0.001	0.004	0.000	0.004	0.000
GARCH-SN	2.352	0.139	0.020	0.000	0.325	0.061
apARCH-SN	2.141	0.314	0.573	0.987	0.603	0.789

Table 4.31: MCS and backtesting results for Dow Jones, $\tau = 5\%$, $H = 1000$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	6.094	0.009	0.000	0	0.028	0.000
QREG-BV	5.985	0.009	0.001	0	0.028	0.000
QREG-MedRV	**5.815**	0.069	0.001	0	0.173	0.000
QREG-logRV	6.172	0.122	0.000	0	0.275	0.000
QREG-logMedRV	6.229	0.003	0.000	0	0.010	0.000

Table 4.32: MCS and backtesting results for FTSE 100, $\tau = 1\%$, $H = 250$

Model	Loss×10^4	UC	CC	DQ	DB_1	DB_7
QREG-RV	1.324	0.278	0.553	0.954	0.563	0.753
QREG-MedRV	1.312	0.278	0.553	0.957	0.563	0.759
QREG-logRV	**1.227**	0.380	0.638	1.000	0.936	0.987
QREG-logBV	1.288	0.380	0.638	0.780	0.657	0.836
QREG-logMedRV	1.327	0.019	0.052	0.053	0.157	0.292
HARQ-BV	1.494	0.278	0.553	0.997	0.563	0.765
HARQ-logRV	1.572	0.380	0.638	0.432	0.657	0.820
HARQ-logBV	1.511	0.758	0.919	0.919	0.936	0.981
HARQ-logMedRV	1.697	0.380	0.638	0.109	0.657	0.323
QREG-IV	1.372	0.005	0.010	0.000	0.059	0.010
QREG-logIv	1.362	0.162	0.078	0.000	0.357	0.543
CAViaR-iGARCH	1.543	0.278	0.553	0.995	0.563	0.765
GARCH-SN	1.519	0.758	0.919	0.613	0.936	0.008
GARCH-ST	1.522	0.742	0.932	0.955	0.945	0.984
apARCH-SN	1.471	0.380	0.638	0.444	0.657	0.104
apARCH-ST	1.551	0.742	0.932	0.939	0.945	0.981
GAS-N	1.609	0.758	0.919	0.672	0.936	0.006
GAS-t	1.687	0.380	0.638	0.268	0.657	0.044
GAS-ST	1.774	0.380	0.638	0.545	0.657	0.011

Table 4.33: MCS and backtesting results for FTSE 100, $\tau = 5\%$, $H = 250$

Model	Loss×10^4	UC	CC	DQ	DB_1	DB_7
QREG-RV	4.931	0.286	0.342	0.710	0.387	0.057
QREG-BV	4.832	0.453	0.530	0.895	0.569	0.129
QREG-MedRV	4.933	0.286	0.342	0.699	0.387	0.095
QREG-logRV	5.013	0.669	0.422	0.475	0.931	0.152
QREG-logBV	5.086	0.329	0.406	0.138	0.709	0.149
QREG-logMedRV	4.969	0.669	0.422	0.381	0.931	0.141
HARQ-RV	4.886	0.083	0.181	0.530	0.119	0.038
HARQ-BV	4.871	0.163	0.290	0.649	0.231	0.007
HARQ-MedRV	4.868	0.163	0.290	0.711	0.231	0.161
HARQ-logRV	4.859	0.453	0.496	0.821	0.569	0.426
HARQ-logBV	**4.820**	0.453	0.496	0.812	0.569	0.059
HARQ-logMedRV	4.936	0.163	0.290	0.577	0.387	0.209
QREG-IV	4.838	0.215	0.344	0.375	0.555	0.059
QREG-logIV	4.838	0.481	0.776	0.557	0.844	0.260
GARCH-ST	5.058	0.163	0.290	0.640	0.231	0.029
apARCH-SN	5.139	0.163	0.290	0.483	0.231	0.007
apARCH-ST	5.126	0.163	0.290	0.473	0.231	0.010
GAS-t	5.251	0.884	0.539	0.543	0.882	0.103

Table 4.34: MCS and backtesting results for FTSE 100, $\tau = 1\%$, $H = 500$

Model	Loss×10^4	UC	CC	DQ	DB_1	DB_7
QREG-RV	1.873	1.000	0.951	0.859	0.990	0.063
QREG-BV	1.756	0.641	0.869	0.934	0.895	0.075
QREG-MedRV	1.870	1.000	0.951	0.898	0.990	0.134
QREG-logRV	1.795	0.215	0.407	0.901	0.448	0.201
QREG-logBV	1.801	0.048	0.115	0.326	0.134	0.259
QREG-logMedRV	**1.701**	0.397	0.632	0.653	0.680	0.853
HARQ-RV	1.761	0.028	0.090	0.860	0.092	0.189
HARQ-BV	1.752	0.028	0.090	0.857	0.092	0.189
HARQ-MedRV	1.768	0.028	0.090	0.857	0.092	0.189
HARQ-logRV	1.851	0.641	0.869	0.704	0.895	0.040
HARQ-logBV	1.843	0.641	0.869	0.651	0.895	0.043
HARQ-logMedRV	1.897	1.000	0.951	0.216	0.990	0.011
QREG-IV	1.759	0.020	0.033	0.014	0.063	0.136
QREG-logIV	1.731	0.048	0.059	0.049	0.134	0.260
CAViaR-ASlope	1.848	0.125	0.306	0.970	0.313	0.508
CAViaR-iGARCH	1.904	0.125	0.306	0.960	0.313	0.507
GARCH-SN	1.857	0.663	0.845	0.999	0.893	0.972
GARCH-ST	1.862	0.331	0.613	0.997	0.628	0.818
apARCH-SN	8.952	0.000	0.000	0.000	0.000	0.000
apARCH-ST	1.872	0.641	0.869	1.000	0.895	0.973
GAS-N	2.037	0.663	0.845	0.995	0.893	0.529
GAS-t	1.974	0.397	0.632	0.935	0.680	0.353
GAS-ST	2.255	0.215	0.407	0.621	0.448	0.079

Table 4.35: MCS and backtesting results for FTSE 100, $\tau = 5\%$, $H = 500$

Model	Loss×10^4	UC	CC	DQ	DB_1	DB_7
QREG-RV	6.470	0.836	0.967	0.982	0.937	0.986
QREG-BV	6.392	0.685	0.833	0.912	0.860	0.959
QREG-MedRV	6.403	0.530	0.513	0.867	0.799	0.915
QREG-logRV	6.516	0.546	0.784	0.924	0.771	0.906
QREG-logBV	6.632	0.319	0.415	0.322	0.554	0.432
QREG-logMedRV	6.355	0.838	0.842	0.981	0.922	0.804
HARQ-RV	6.276	0.049	0.084	0.672	0.147	0.279
HARQ-BV	6.283	0.049	0.084	0.662	0.147	0.273
HARQ-MedRV	6.277	0.049	0.084	0.672	0.147	0.279
HARQ-logRV	6.333	0.131	0.163	0.688	0.323	0.325
HARQ-logBV	6.388	0.131	0.163	0.730	0.323	0.506
HARQ-logMedRV	6.289	0.027	0.055	0.571	0.090	0.179
QREG-IV	**6.242**	0.423	0.605	0.909	0.666	0.840
QREG-logIV	6.252	0.423	0.703	0.928	0.666	0.829
apARCH-ST	6.693	0.288	0.247	0.861	0.564	0.736

Table 4.36: MCS and backtesting results for FTSE 100, $\tau = 1\%$, $H = 1000$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	2.306	0.139	0.157	0.355	0.325	0.503
QREG-BV	2.297	0.079	0.166	0.291	0.208	0.369
QREG-MedRV	2.277	0.139	0.266	0.735	0.325	0.499
QREG-logRV	2.125	0.079	0.166	0.275	0.208	0.369
QREG-logBV	2.061	0.139	0.157	0.073	0.325	0.518
QREG-logMedRV	**1.894**	0.231	0.399	0.839	0.475	0.682
HARQ-RV	1.994	0.746	0.875	0.999	0.943	0.990
HARQ-BV	2.027	1.000	0.904	0.999	0.990	0.999
HARQ-MedRV	2.030	0.538	0.263	0.396	0.812	0.933
HARQ-logRV	2.118	0.754	0.258	0.002	0.939	0.000
HARQ-logBV	2.127	0.754	0.258	0.022	0.939	0.007
HARQ-logMedRV	2.018	0.231	0.204	0.001	0.475	0.001
QREG-IV	2.010	0.011	0.006	0.000	0.038	0.087
QREG-logIV	2.047	0.005	0.003	0.000	0.019	0.046
CAViaR-SAV	2.072	0.362	0.556	0.959	0.646	0.831
CAViaR-ASlope	1.988	0.170	0.376	0.964	0.392	0.598
CAViaR-iGARCH	2.041	1.000	0.226	0.271	0.990	0.999
GARCH-SN	2.171	0.005	0.003	0.000	0.019	0.047
GARCH-ST	2.070	0.231	0.204	0.402	0.475	0.685
apARCH-SN	4.755	0.000	0.000	0.000	0.000	0.000
apARCH-ST	2.013	1.000	0.904	0.997	0.990	0.997
GAS-N	2.235	0.043	0.096	0.348	0.124	0.238
GAS-t	2.288	0.005	0.013	0.024	0.019	0.007
GAS-ST	2.483	0.022	0.053	0.070	0.070	0.031

Table 4.37: MCS and backtesting results for FTSE 100, $\tau = 5\%$, $H = 1000$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	7.016	0.257	0.355	0.208	0.483	0.693
QREG-BV	6.966	0.393	0.691	0.863	0.642	0.817
QREG-MedRV	7.021	0.257	0.494	0.955	0.483	0.692
QREG-logRV	7.113	0.159	0.349	0.476	0.338	0.199
QREG-logBV	7.118	0.204	0.428	0.439	0.407	0.371
QREG-logMedRV	**6.944**	0.204	0.428	0.889	0.407	0.595
HARQ-RV	7.097	0.660	0.789	0.698	0.873	0.965
HARQ-BV	7.153	0.660	0.789	0.751	0.873	0.952
HARQ-MedRV	7.117	0.233	0.483	0.882	0.483	0.692
HARQ-logRV	7.182	0.566	0.693	0.674	0.792	0.367
HARQ-logBV	7.192	0.885	0.688	0.606	0.936	0.300
HARQ-logMedRV	7.075	0.557	0.399	0.808	0.812	0.224
QREG-IV	7.005	0.204	0.428	0.562	0.407	0.614
QREG-logIV	7.011	0.257	0.494	0.873	0.483	0.691
CAViaR-SAV	7.455	0.022	0.057	0.486	0.073	0.154
CAViaR-ASlope	7.387	0.022	0.057	0.476	0.073	0.155
GARCH-SN	7.462	1.000	0.949	0.507	0.950	0.049
GARCH-ST	7.444	1.000	0.949	0.515	0.950	0.054
apARCH-SN	10.300	0.000	0.000	0.000	0.000	0.000
apARCH-ST	7.353	0.557	0.838	0.807	0.812	0.889
GAS-N	7.501	0.660	0.789	0.648	0.873	0.056
GAS-t	7.627	0.320	0.603	0.223	0.562	0.003

Table 4.38: MCS and backtesting results for EUROSTOXX 50, $\tau = 1\%$, $H = 250$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	2.100	0.742	0.932	0.971	0.945	0.989
QREG-BV	1.967	0.742	0.932	0.914	0.945	0.987
QREG-MedRV	2.053	0.742	0.932	0.988	0.945	0.989
QREG-logRV	2.755	0.000	0.000	0.000	0.000	0.000
QREG-logBV	2.851	0.001	0.004	0.000	0.005	0.000
QREG-logMedRV	3.016	0.001	0.004	0.000	0.005	0.000
HARQ-BV	**1.638**	0.278	0.553	0.976	0.563	0.765
HARQ-MedRV	1.731	0.025	0.081	0.928	0.084	0.175
HARQ-logRV	2.158	0.162	0.339	0.000	0.357	0.013
HARQ-logBV	2.313	0.005	0.016	0.000	0.019	0.000
HARQ-logMedRV	2.182	0.380	0.638	0.000	0.657	0.045
QREG-IV	2.477	0.000	0.001	0.000	0.001	0.000
QREG-logIV	2.735	0.000	0.001	0.000	0.001	0.000
CAViaR-ASlope	1.829	0.278	0.553	0.996	0.563	0.765
GARCH-SN	1.950	0.278	0.553	0.990	0.563	0.763
apARCH-SN	1.819	0.278	0.553	0.996	0.563	0.765
GAS-t	1.986	0.278	0.553	0.991	0.563	0.763

Table 4.39: MCS and backtesting results for EUROSTOXX 50, $\tau = 5\%$, $H = 250$

Model	Loss$\times 10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	6.573	0.037	0.097	0.512	0.119	0.234
QREG-BV	6.272	0.037	0.097	0.741	0.119	0.232
QREG-MedRV	6.463	0.083	0.181	0.602	0.231	0.394
QREG-logRV	7.183	0.133	0.080	0.000	0.279	0.003
QREG-logBV	7.217	0.079	0.044	0.000	0.181	0.001
QREG-logMedRV	7.436	0.044	0.114	0.000	0.111	0.001
HARQ-RV	6.329	0.014	0.043	0.669	0.052	0.115
HARQ-BV	**6.126**	0.083	0.181	0.894	0.231	0.397
HARQ-MedRV	6.184	0.014	0.043	0.688	0.052	0.114
HARQ-logRV	6.777	0.006	0.019	0.000	0.018	0.000
HARQ-logBV	6.929	0.012	0.032	0.000	0.035	0.000
HARQ-logMedRV	7.017	0.012	0.044	0.000	0.035	0.000
QREG-IV	7.202	0.215	0.133	0.000	0.406	0.005
QREG-logIV	7.181	0.215	0.133	0.000	0.406	0.005
GARCH-ST	6.489	0.037	0.097	0.744	0.119	0.229
apARCH-ST	6.332	0.014	0.043	0.689	0.052	0.114
GAS-t	6.630	0.083	0.181	0.812	0.231	0.092

Table 4.40: MCS and backtesting results for EUROSTOXX 50, $\tau = 1\%$, $H = 500$

Model	Loss$\times10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	3.973	0.663	0.142	0.000	0.893	0.690
QREG-BV	4.181	0.663	0.142	0.000	0.893	0.733
QREG-MedRV	4.012	0.663	0.142	0.000	0.893	0.664
QREG-logRV	4.140	0.000	0.001	0.000	0.001	0.000
QREG-logBV	3.960	0.001	0.003	0.000	0.004	0.000
QREG-logMedRV	3.961	0.003	0.007	0.000	0.010	0.000
HARQ-RV	3.594	0.331	0.613	0.988	0.628	0.817
HARQ-BV	**3.502**	0.641	0.869	0.998	0.895	0.974
HARQ-MedRV	3.608	0.331	0.613	0.991	0.628	0.816
HARQ-logRV	3.751	0.106	0.230	0.000	0.260	0.002
HARQ-logBV	3.846	0.003	0.007	0.000	0.010	0.000
HARQ-logMedRV	3.724	0.106	0.230	0.000	0.260	0.003
QREG-IV	3.810	0.003	0.008	0.000	0.010	0.000
QREG-logIV	3.967	0.000	0.001	0.000	0.001	0.000
CAViaR-SAV	3.542	0.641	0.869	0.999	0.895	0.973
CAViaR-ASlope	3.681	0.641	0.869	1.000	0.895	0.974
CAViaR-iGARCH	3.639	0.641	0.869	0.999	0.895	0.974
GARCH-SN	3.618	1.000	0.951	1.000	0.990	0.999
GARCH-ST	3.645	0.331	0.613	0.995	0.628	0.817
apARCH-SN	3.783	0.641	0.869	1.000	0.895	0.974
apARCH-ST	3.826	0.641	0.869	1.000	0.895	0.974
GAS-N	4.127	0.663	0.845	0.999	0.893	0.973
GAS-t	3.747	0.397	0.632	0.967	0.680	0.803
GAS-ST	4.179	0.397	0.632	0.984	0.680	0.856

Table 4.41: MCS and backtesting results for EUROSTOXX 50, $\tau = 5\%$, $H = 500$

Model	Loss$\times10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	10.442	0.007	0.016	0.001	0.027	0.017
QREG-BV	10.405	0.082	0.193	0.005	0.224	0.171
QREG-MedRV	10.409	0.007	0.016	0.001	0.027	0.018
QREG-logRV	10.710	0.838	0.234	0.048	0.922	0.053
QREG-logBV	10.691	0.319	0.089	0.012	0.554	0.001
QREG-logMedRV	10.739	0.423	0.605	0.025	0.666	0.013
HARQ-RV	10.363	0.027	0.055	0.478	0.090	0.109
HARQ-BV	**10.129**	0.027	0.055	0.596	0.090	0.185
HARQ-MedRV	10.211	0.082	0.121	0.681	0.224	0.323
HARQ-logRV	10.765	0.117	0.249	0.000	0.259	0.000
HARQ-logBV	10.737	0.013	0.010	0.000	0.039	0.000
HARQ-logMedRV	10.709	0.013	0.010	0.000	0.039	0.000
QREG-IV	10.557	0.838	0.234	0.126	0.922	0.005
QREG-logIV	10.581	0.838	0.234	0.083	0.922	0.002
GARCH-SN	10.720	0.082	0.121	0.448	0.224	0.294
GARCH-ST	10.662	0.131	0.163	0.534	0.323	0.407
apARCH-ST	10.423	0.003	0.009	0.370	0.013	0.035
GAS-N	12.320	0.027	0.055	0.279	0.090	0.185
GAS-t	11.037	0.199	0.207	0.647	0.438	0.648

Table 4.42: MCS and backtesting results for EUROSTOXX 50, $\tau = 1\%$, $H = 1000$

Model	Loss$\times10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	4.184	0.022	0.046	0.000	0.070	0.149
QREG-BV	4.021	0.011	0.026	0.000	0.038	0.086
QREG-MedRV	4.053	0.043	0.074	0.000	0.124	0.132
QREG-logRV	3.696	0.000	0.001	0.000	0.001	0.000
QREG-logBV	3.748	0.000	0.002	0.000	0.002	0.001
QREG-logMedRV	3.721	0.043	0.074	0.000	0.124	0.081
HARQ-RV	3.607	0.139	0.266	0.662	0.325	0.517
HARQ-BV	3.577	0.362	0.556	0.935	0.646	0.829
HARQ-MedRV	3.662	0.079	0.166	0.321	0.208	0.368
HARQ-logRV	3.562	0.139	0.266	0.003	0.325	0.007
HARQ-logBV	3.468	0.362	0.556	0.117	0.646	0.335
HARQ-logMedRV	3.487	0.538	0.715	0.412	0.812	0.041
QREG-IV	3.394	0.002	0.007	0.000	0.009	0.001
QREG-logIV	3.508	0.022	0.046	0.000	0.070	0.039
CAViaR-SAV	3.384	0.746	0.875	0.999	0.943	0.989
CAViaR-ASlope	**3.368**	0.754	0.842	0.994	0.939	0.988
CAViaR-iGARCH	3.433	0.754	0.842	0.997	0.939	0.989
GARCH-SN	3.504	0.079	0.166	0.568	0.208	0.369
GARCH-ST	3.429	1.000	0.904	0.997	0.990	0.999
apARCH-SN	3.433	0.754	0.842	0.996	0.939	0.988
apARCH-ST	3.456	0.746	0.875	0.985	0.943	0.986
GAS-N	3.735	0.362	0.556	0.941	0.646	0.827
GAS-t	3.630	0.139	0.266	0.308	0.325	0.522

Table 4.43: MCS and backtesting results for EUROSTOXX 50, $\tau = 5\%$, $H = 1000$

Model	Loss$\times10^4$	UC	CC	DQ	DB_1	DB_7
QREG-RV	11.591	0.666	0.709	0.025	0.854	0.952
QREG-BV	11.766	0.885	0.368	0.009	0.936	0.975
QREG-MedRV	11.525	0.557	0.399	0.023	0.812	0.932
QREG-logRV	11.592	0.666	0.189	0.001	0.854	0.201
QREG-logBV	11.530	0.773	0.395	0.004	0.903	0.198
QREG-logMedRV	11.566	1.000	0.333	0.005	0.950	0.234
HARQ-RV	11.581	0.178	0.036	0.084	0.399	0.587
HARQ-BV	11.461	0.375	0.100	0.025	0.657	0.829
HARQ-MedRV	11.501	0.233	0.153	0.026	0.483	0.685
HARQ-logRV	11.672	0.884	0.293	0.041	0.944	0.054
HARQ-logBV	11.623	1.000	0.133	0.000	0.950	0.058
HARQ-logMedRV	11.575	1.000	0.333	0.000	0.950	0.029
QREG-IV	11.359	0.773	0.708	0.005	0.903	0.559
QREG-logIV	11.477	0.566	0.422	0.003	0.792	0.441
CAViaR-SAV	11.534	0.178	0.339	0.885	0.400	0.603
CAViaR-ASlope	11.343	0.033	0.086	0.424	0.103	0.205
CAViaR-iGARCH	11.586	0.178	0.339	0.832	0.400	0.600
GARCH-SN	11.569	0.884	0.951	0.999	0.944	0.989
GARCH-ST	11.562	0.666	0.709	0.550	0.854	0.955
apARCH-SN	11.299	0.299	0.418	0.730	0.570	0.557
apARCH-ST	**11.270**	0.375	0.511	0.825	0.657	0.657
GAS-N	12.046	0.299	0.579	0.820	0.570	0.626
GAS-t	11.785	0.566	0.847	0.823	0.792	0.289

Chapter 5

Bias Corrections for Exponentially Transformed Forecasts: Are they worth the effort?

Coauthored by: Matei Demetrescu and Vasyl Golosnoy

5.1 Motivation

Taking logs is widely used in applied (time series) econometrics to linearize relations or to stabilize variances. It has become a standard transformation for time series in numerous economic and financial applications; see among others Andersen et al. (2011), Bauer and Vorkink (2011), Hautsch (2012), Lütkepohl and Xu (2012), Golosnoy et al. (2012), Mayr and Ulbricht (2015), Brechmann et al. (2018), or Gribisch (2018). In fact, models in logs often turn out to be better suited for both estimation and forecasting. Then, the forecast in logs should be transformed back in order to predict the original variable of interest. Such reverse exponential transformations, however, introduce a point forecast bias into the procedure, as already emphasized by Granger and Newbold (1976).

The practically relevant question is how to deal with such bias in finite samples for various types of distributions for log-model errors. Of course, ignoring the bias and simply transforming the forecast in logs through the exponential function is one possible course of action, even if a naïve one, since ignoring the bias may lead to substantial losses in forecasting precision (cf. Lütkepohl and Xu, 2012; Proietti and Lütkepohl, 2013). At the other end of the spectrum of possibilities, one finds numerical bootstrap-based corrections (cf. Thombs and Schucany, 1990) – which are, however, computationally demanding and not always easy to implement. As for interval forecasts, note that the required quantiles remain unbiased after any monotone transformation, so that bias correction in point forecasts is indeed the essential problem to be resolved.

In this paper we compare several bias correction procedures which are of most practical applicability (cf. Stock and Watson, 2012, p. 314–315). Concretely, we consider a popular method that exploits the residual variance for the bias correction, as well as one that relies on computing the sample mean of exponentially transformed log-model residuals. Whereas the variance-based correction is optimal for normally distributed innovations, the mean-based correction only requires the existence of the relevant expectation. We additionally examine a semiparametric approach based on estimation of the model in logs under the Linex loss (Varian, 1975), which we show to provide asymptotically unbiased forecasts of the original (untransformed) series so that no correction is necessary in this case. The Linex-based approach exploits a non-linear estimation procedure that could, however, cause losses in estimation efficiency compared to maximum likelihood estimation; such efficiency losses may seriously impact the behavior in finite samples.

We study settings with autoregressive (AR) data generating processes in logs with model errors following different types of distributions. The interest lies here in making a one step ahead forecast of the original variable, but our analysis could easily be applied for the task of predicting at longer horizons. We compare the effectiveness of the above-mentioned bias correction methods with those for the naïve approach without any adjustment. The forecasting performance of different correction methods has been already studied for several settings, e.g. for a family of data generating processes with Markov switching (cf. Patton and Timmermann, 2007b). We extend this strand of literature by specifically focusing on error distributions that exhibit deviations from normality which are of high empirical relevance. In particular, we study the effect of skew-normal (Azzalini, 1985), mixture normal (Everitt and Hand, 1981; McLachlan and Peel, 2004), contaminated normal (Seidel, 2011) as well as t-distributed innovations (cf. Tarami and Pourahmadi, 2003). Since we investigate several AR and ARMA models with different degrees of persistence, our setup covers a broad class of practically important situations.

We find, first, that the variance-based correction appears to be the preferable approach in smaller samples, even for various deviations from normality; the expectation-based correction of the residual exponent is a close competitor. Second, despite being attractive from a theoretical point of view, the Linex-based approach requiring no specific correction shows losses in estimation efficiency. It appears to be dominated by two above-mentioned alternatives in terms of the considered forecasting loss functions; but, as the estimation error diminishes with increasing sample size, the Linex-based approach becomes competitive. Third, a naïve prediction without bias correction is found to be suitable for highly persistent AR processes in logs with the AR(1) coefficient ≥ 0.9. This perhaps surprising finding could be explained by difficulties with variance estimation for bias correction factors when the process gets closer to the unit root.

Our paper is focused on taking logs which is the most frequently used in practice non-linear transformation. A more general Box-Cox transformation may be of empirical importance as well, see e.g. Taylor (2017) and the literature cited therein. The price to pay in this more flexible case is, however, that one cannot easily obtain corrections for the induced prediction bias. We address the bias correction for the Box-Cox procedure in Appendix B, where we show that it is not a trivial task at all, so that one should rely on some type of bootstrap-based corrections there.

Summarizing, the variance-based correction performs well for different cases when the normality assumption is violated, but the Linex-based method providing unbiased forecast or even naïve no-correction approach may outperform the variance-based method for persistent series and larger samples. These findings are supported by the empirical results of using a log heterogenous autoregressive (HAR) model (Corsi et al., 2012) for the purpose of predicting daily realized volatility for highly liquid U.S. stocks.

The remainder of the paper is structured as follows. In section 5.2 we discuss the necessity of making a bias correction, give a summary of established methods suitable for this purpose, and discuss the approach based on estimation under Linex loss. The extensive simulation study covering various types of linear processes and error distributions is presented in section 5.3, whereas section 5.4 contrasts the behavior of the different bias corrections in an empirical application. The final section 5 concludes the paper, while the Appendix collects some technical arguments.

5.2 Problem Setting and Bias Correction Techniques

5.2.1 The model

Let the strictly positive original untransformed process of interest be given as y_t and assume that its log-transformation $x_t = \log(y_t)$ follows a stationary AR(p) process with *iid* innovations ϵ_t:

$$y_t = \exp(x_t), \qquad \rho(L)x_t = \mu + \epsilon_t, \qquad \epsilon_t \sim iid\,(0, \sigma^2), \tag{5.1}$$

where $\rho(L) = 1 - \sum_{j=1}^{p} \rho_j L^j$ is an invertible lag polynomial of order p. This setting could be directly generalized for ARMA(p,q) time series models, since linear processes may be approximated by means of AR(p) processes with a sufficiently large order p; see Berk (1974); Bhansali (1978). We investigate both AR(1) and ARMA(1,1) settings in the Monte Carlo simulation study in section 3.

We are interested in one step ahead mean squared error (MSE) optimal forecasts of y_{T+1} given $y_T, y_{T-1}, \ldots$, and hence search for the conditional expectation of the examined series:

$$y_T(1) = \mathrm{E}[y_{T+1}|\mathcal{F}_T], \qquad \mathcal{F}_T = \sigma\{y_T, y_{T-1}, \ldots, y_1\} = \sigma\{x_T, x_{T-1}, \ldots, x_1\}$$

with $\sigma\{a_1, a_2, \ldots\}$ denoting the σ-algebra generated by a_1, a_2 etc. Denote by $x_T(1) = \mathrm{E}[x_{T+1}|\mathcal{F}_T]$ the one step head MSE-optimal forecast of the log series x_t so that it holds $x_T(1) = \mu + (1 - \rho(L))x_{T+1}$.

Note that $\mathrm{E}[y_{T+1}]$ must be finite, otherwise no MSE-optimal forecast exists. We therefore require that the distribution of ϵ_t has thin tails. We take ϵ_t to have thin tails if $\mathrm{E}[|\epsilon_t|^k] \leq Ca^k$ for some $C > 0$, $a > 1$ and all $k \in \mathbb{R}^+$, which implies e.g. that $\mathrm{E}[\exp(\epsilon_t)] < \infty$. Moreover, given the stable finite-order autoregressive model assumed, an application of the Minkowski's inequality shows the moments of x_t to satisfy the same conditions as those of ϵ_t, hence, x_t has thin tails as well.

The representation (5.1) in logs is often quite useful for modelling and estimation purposes. In this paper, however, our interest is to make the out-of-sample one step ahead MSE-optimal forecast $y_T(1) = \mathrm{E}[\exp(x_{T+1})|\mathcal{F}_T]$ of the original variable y_{T+1} which is then given as

$$y_T(1) = \exp\Big(\mathrm{E}[x_{T+1}|\mathcal{F}_T]\Big)\ \mathrm{E}\Big[\exp(\epsilon_{T+1})\Big] = \exp\Big(x_T(1)\Big)\ \mathrm{E}\Big[\exp(\epsilon_{T+1})\Big], \tag{5.2}$$

because of the representation

$$x_{T+1} = \mu + \sum_{j=1}^{p} \rho_j x_{T+1-j} + \epsilon_{T+1} = x_T(1) + \epsilon_{T+1} \qquad \text{with} \qquad \mathrm{E}(\epsilon_{T+1}) = 0.$$

As in general $\mathrm{E}[\exp(\epsilon_{T+1})] > 1$ due to Jensen's inequality, a naïve (uncorrected) forecast is

$$y_T(1) = \exp\Big(x_T(1)\Big). \tag{5.3}$$

Clearly, the naïve forecast in (5.3) is not MSE-optimal and has a *downward* bias given by

$$\mathrm{E}\Big[y_T(1) - y_{T+1}|\mathcal{F}_T\Big] = \exp\Big(x_T(1)\Big)\Big(1 - \mathrm{E}[\exp(\epsilon_{T+1})]\Big).$$

The magnitude of the bias depends on the unknown distribution of the shocks ϵ_t, but also on the conditional expectation of x_t. In practice, one should estimate these forecast functions by plugging in consistent estimators $\hat{\mu}$ and $\hat{\rho}_j$, leading to $\hat{x}_T(1)$, so that the issue of forecasting bias could be addressed subsequently. The sample $x_1, ..., x_T$ is used in order to estimate the model parameters in Equation (5.1) and any of the bias correction terms.

To summarize, the MSE of the forecast $y_T(1)$ has three main sources: the bias, the volatility of shocks, and the model estimation error. Note that the estimation error could play a substantial role, since it is not negligible in smaller samples. Hence, lack of efficiency in parameter estimation may become an issue when the sample is not large enough. E.g. for Gaussian errors ϵ_t, a least squares (LS) estimation is maximum likelihood (ML) and, thus, asymptotically efficient, while estimation under the Linex loss could be advantageous under skewed error distributions as it may be approximately proportional to the logarithm of the errors' density for a suitable skewness.

5.2.2 Variance-based bias corrections

When the model innovations ϵ_t are iid normally distributed, the optimal forecast $y_T(1)$ can be obtained as an explicit function of the error variance (Granger and Newbold, 1976):

$$y_T(1) = \mathrm{E}[y_{T+1}|\mathcal{F}_T] = \exp\Big(x_T(1)\Big)\ \mathrm{E}\Big[\exp(\epsilon_{T+1})\Big] = \exp\Big(x_T(1)\Big) \cdot \exp\Big(\frac{1}{2}\sigma^2\Big).$$

Then a feasible variance-corrected forecast is given by

$$\hat{y}_T(1) = \exp\Big(\hat{x}_T(1)\Big)\exp\Big(\frac{1}{2}\hat{\sigma}^2\Big), \tag{5.4}$$

where $\hat{\sigma}^2$ denotes a consistent estimator of the error variance σ^2 and $\hat{x}_T(1)$ the estimated forecast from the log model in (5.1). For large T, where estimation noise is negligible, this correction is exact for normally distributed model innovations. Pronounced empirical deviations from normality are rather frequent in practice, however. For this reason we now examine bias corrections which place fewer restrictions on the distribution of ϵ_t.

5.2.3 Mean-based bias correction

One could estimate the expectation $\mathrm{E}\Big[\exp(\epsilon_{T+1})\Big]$ in (5.2) directly from the sample, e.g. as the sample average of transformed residuals,

$$\widehat{\mathrm{E}}\Big[\exp(\epsilon_{T+1})\Big] = \frac{1}{T}\sum_{t=1}^{T}\exp(\hat{\epsilon}_t), \tag{5.5}$$

where $\hat{\epsilon}_t$ are the in-sample model residuals. Since we assumed thin-tailed innovations, the expectation is finite. The mean-corrected forecast would then be

$$\hat{y}_T(1) = \exp\Big(\hat{x}_T(1)\Big) \cdot \widehat{\mathrm{E}}\Big[\exp(\epsilon_{T+1})\Big]. \tag{5.6}$$

Of course, more robust estimates of the central tendency (e.g., median or truncated mean) could be applied here as well. In practice we resort to residuals to compute an estimate of $\mathrm{E}[\exp(\epsilon_{T+1})]$ so that estimation error plays a role in this step as well.

5.2.4 Forecasts based on the Linex loss

The considered above variance-based and mean-based bias corrections are two step procedures, as in the first step one should estimate the AR model in (5.1) and in the second step compute the bias correction factor. We now consider a distribution-free approach which enables unbiased forecasts in a single step for exponentially transformed values.

To obtain such a single step forecast let

$$m_{t+1} = \log \mathrm{E}\left[y_{t+1}|\mathcal{F}_t\right],$$

such that $\exp(m_{t+1}) = \mathrm{E}\left[y_{t+1}|\mathcal{F}_t\right] = \mathrm{E}\left[\exp(x_{t+1})|\mathcal{F}_t\right]$, or, equivalently,

$$\mathrm{E}\Big[\, e^{x_{t+1}-m_{t+1}} - 1 \Big|\, x_t, x_{t-1}, \dots \Big] = 0. \tag{5.7}$$

Note that this equality holds irrespectively of the distribution of forecast errors.

Then, rather than predicting x_{T+1} and correcting the bias introduced by a non-linear transformation of $x_T(1)$, the idea is to estimate the conditional quantity m_{T+1} by directly imposing the moment condition (5.7). The latter is simply a transformed version of the required MSE-optimal forecast for y_t, as we have $e^{m_{t+1}} = \mathrm{E}\left[y_{t+1}|\mathcal{F}_t\right]$ for all t by definition.

Note that since the forecast m_{T+1} is not the conditional expectation of x_{T+1} given $\mathcal{F}_T$, it delivers a biased prediction of log-transformed variables x_{T+1}. However, this bias in the log series forecasts is such that the exponent transformation to the original variable of interest y_{T+1}

provides an unbiased forecast. This is different from conventional procedures with first doing unbiased forecasts for log series and then making some bias corrections in order to predict original variables in an unbiased manner.

The generalized method of moments (GMM) estimation is a natural choice to impose the condition (5.7); for our case a particular selection of instruments leads to the following estimator with a nice interpretation. Namely, we obtain from (5.1) that

$$m_{t+1} = \mu + \sum_{j=1}^{p} \rho_j \, x_{t+1-j} + \log\Big(\mathrm{E}\Big[\exp(\epsilon_{T+1})\Big]\Big) := \tilde{\mu} + \sum_{j=1}^{p} \rho_j \, x_{t+1-j},$$

and consider for $\theta = (\tilde{\mu}, \rho_1, \ldots, \rho_p)'$ the vector of moment conditions

$$\mathrm{E}\left[\left(e^{x_{t+1}-m_{t+1}} - 1\right) \frac{\partial m_{t+1}}{\partial \theta}\right] = 0.$$

For $\mathcal{L}(u) = e^u - u - 1$, these are the first-order conditions for the optimization problem

$$\min_{\theta} \mathrm{E}\Big[\mathcal{L}\Big(x_{t+1} - m_{t+1}(\theta)\Big)\Big],$$

where $\mathcal{L}(u)$ is recognized to be the linear-exponential (Linex) loss function introduced by Varian (1975) with parameters $a_1 = a_2 = 1$ in $\mathcal{L}_{a_1,a_2}(u) = e^{a_1 u} - a_2 u - 1$.[1]

Hence, we may estimate the model in logs under the Linex loss instead of using least squares by minimizing the average empirical loss

$$\hat{\theta} = \arg\min \sum_{t=p+1}^{T} \mathcal{L}\Big(x_{t+1} - m_{t+1}(\theta)\Big),$$

and by computing

$$\hat{y}_T(1) = \exp\left(\hat{m}_{T+1}\right) = \exp\left(m_{T+1}(\hat{\theta})\right).$$

We establish consistency of this forecasting procedure – in the sense that $\hat{y}_T(1)$ converges to the conditional expectation of y_{T+1} – by means of standard extremum estimator theory; the details are provided in the Appendix. In particular, we show in Appendix A the Linex-based estimators of $\rho_1, \ldots, \rho_p$ to be consistent; note that the intercept estimated via Linex is asymptotically biased, as it to converges a.s. to $\tilde{\mu}$ where $\tilde{\mu} \neq \mu$ in general. Fortunately, as argued above, this feature delivers the desired $\mathrm{E}\left[\exp\left(x_{T+1}\right)\right]$ in the limit, particularly because of the asymptotic bias of $\hat{\mu}$. This guarantees the MSE-optimality of the forecasts for y_t for large T values. Additionally, in Appendix B we discuss the application of bias corrections for the non-linear Box-Cox procedure which is a generalization of the log transformation.

[1] The existence of the expectation $E[\mathcal{L}_{a_1,a_2}]$ is easy to establish given the thin tails of x_t and the linear-exponential shape of $\mathcal{L}_{a_1,a_2}$.

5.3 Monte Carlo analysis

We now examine how the shape of the error distribution affects the MSE of the forecasts under consideration in finite samples. In order to contrast various bias corrections, we concentrate both on a simple AR(1) as well as on a more sophisticated ARMA(1,1) processes for the log-transformed values because of their immense practical importance. The stationary AR(1) model is given as

$$x_t = \mu + \rho x_{t-1} + \epsilon_t, \qquad \epsilon_t \sim iid(0, 1/4), \qquad |\rho| < 1,$$

and we experiment with $\rho \in \{0.2, 0.5, 0.9\}$. We set error variance equal to 1/4 to be in line with our empirical study; see Table 5.1.

Additionally, we specify the stationary ARMA(1,1) model by

$$x_t = \mu + \rho x_{t-1} + \phi\epsilon_{t-1} + \epsilon_t, \qquad \epsilon_t \sim iid(0, 1/4),$$

where we use the values $\rho = 0.8$ and $\phi = -0.5$. This choice of ARMA(1,1) parameters leads to an autocorrelation function similar to that of the HAR model of Corsi (2009) which is investigated in the empirical study in section 5. For both AR(1) and ARMA(1,1) we set $\mu = 0$ without loss of generality but estimate it from the data.

We are interested in forecasting $y_{T+1} = \exp(x_{T+1})$ given the information set $\mathcal{F}_T$. In case of normally distributed innovations ϵ_t the variance-based bias correction would be optimal. In the following analysis we investigate different types of innovation distributions and compare forecasting losses from the competing bias correction methods.

With all simulations are performed in R (R-Core-Team, 2014), the estimation of these models in logs is conducted based on samples of size $T \in \{200, 500, 1000\}$ with 10^4 Monte Carlo replications. We have also conducted some simulations with 10^5 repetitions, however, the form and, apparently, non-smoothness of the resulting curves remain essentially the same. A possible reason for this behavior is a rather high variance of the observed MSE of forecasts of *exponentially* transformed series.

5.3.1 Distribution of innovations

We consider four different types of deviations from normality that are next discussed as Cases I–IV. In Case I with skew-normal innovations we assume ϵ_t to follow a standardized skew normal distribution (SND) which is of vast importance in the current literature (Bondon, 2009; Sharafi and Nematollahi, 2016). A SND-distributed random variable u_t is characterized (Azzalini, 1985) by three parameters (ξ, ω, β) such that its mean and variance are given with the parameter $\delta = \beta/\sqrt{1+\beta^2}$ as

$$\mathrm{E}[u_t] = \xi + \omega \cdot \delta\sqrt{2/\pi}, \quad \text{and} \quad \mathrm{Var}[u_t] = \omega^2(1 - 2\delta^2/\pi).$$

We set $\xi = 0$, $\omega = 1$ and compute the innovations calibrated to zero mean and variance equal 1/4 for various values of the skewness parameter β:

$$\epsilon_t = \frac{u_t - \mathrm{E}[u_t]}{2\sqrt{\mathrm{Var}[u_t]}}.$$

In Case II we assume that ϵ_t follows a symmetric normal mixture distribution (NMD). This is an another popular deviation from normality (cf. McLachlan and Peel, 2004). NMD random variables u_t are given as

$$u_t \quad \sim \quad \begin{cases} \mathcal{N}(0, \sigma_1^2), & \text{if } B_t = 1, \text{ i.e. with probability } \pi, \\ \mathcal{N}(0, \sigma_2^2), & \text{if } B_t = 0, \text{ i.e. with probability } 1 - \pi, \end{cases}$$

with an *iid* mixture variable B_t drawn from the Bernoulli distribution with the success probability $\pi \in (0, 1)$. Thus, the mixture distribution is characterized by three parameters $(\sigma_1^2, \sigma_2^2, \pi)$ with the variance $\mathrm{Var}[u_t] = \pi\sigma_1^2 + (1 - \pi)\sigma_2^2$. We set the mixture probability $\pi = 1/2$, $\sigma_1^2 = 1$ and vary only the second variance σ_2^2. We model innovations as

$$\epsilon_t = \frac{u_t}{2\sqrt{\pi\sigma_1^2 + (1 - \pi)\sigma_2^2}}.$$

Note that the benchmark case of standard normal distribution is implicitly included in Case I with $\beta = 0$.

In Case III we assume the innovations to follow a contaminated normal distribution which allows for higher kurtosis values (cf. Seidel, 2011). This case is rather similar to the previous Case II specification. The difference lies in the mixture probability that is now set to $\pi = 0.95$ and the second variance is of larger magnitude. We, again, use innovations ϵ_t calibrated to zero mean and variance of 1/4 for model estimation and evaluation.

Next, in Case IV we assume u_t follows a central t-distribution with $v \in [5, 30]$ degrees of freedom. The adjusted errors ϵ_t are obtained by $\epsilon_t = u_t \cdot 0.5\ (v/(v - 2))^{-1/2}$. This choice is motivated as a robustness check, since the t-distribution has fat tails, and therefore y_t would not have a finite expectation.

Finally, we investigate Cases I, II, and IV for the ARMA(1,1) model with $\rho = 0.8$ and $\phi = -0.5$ for different sample sizes T, whereas Case III is omitted as the results there are very similar to those in Case II.

5.3.2 Methods for bias correction

The following methods are considered for making one step ahead forecasts of y_{T+1}.

1. Naïve forecast ignoring bias corrections $\hat{y}_T(1) = \exp(\hat{x}_T(1))$.
2. Variance-based correction with $\hat{y}_T(1) = \exp(\hat{x}_T(1)) \exp\left(\frac{1}{2}\hat{\sigma}^2\right)$, where the variance σ^2 is estimated from sample residuals $\hat{\epsilon}_T, ..., \hat{\epsilon}_1$ as $\hat{\sigma}^2 = (1/T)\sum_{t=1}^{T} \hat{\epsilon}_t^2$.

3. Mean-based bias correction with $\hat{y}_T(1) = \exp(\hat{x}_T(1)) \cdot (1/T)\sum_{t=1}^{T}\exp(\hat{\epsilon}_t)$.

4. Linex-based forecast with $\hat{y}_T(1) = \exp(\hat{m}_{T+1})$.

5. A simple average of the mean-based and variance-based forecasts.

6. Untransformed forecasts as in Mayr and Ulbricht (2015): i.e. fitting either AR(1) or ARMA(1,1) models directly to the untransformed variables y_t.

The plug-in estimates for the autoregressive parameters were obtained by means of OLS. As for the estimation of the Linex-based forecast function, we make use of a Newton-Raphson non-linear optimization algorithm (see e.g. Hamilton, 1994, p. 138) having as starting values the OLS estimates obtained for the other three corrections.

For the AR(1) model, we plot in Figures 5.1-5.3 corresponding to Cases I, II, IV the log MSE differences of naïve, mean-based, average, Linex-based, and untransformed forecasts to the baseline variance-based forecast correction which is optimal in case of normal innovations. Moreover, in Figure 5.4 we show the log MSE differences for the ARMA(1,1) model. As the results for Case III are very similar to those for Case II, we decide to skip the Case III plots due to space considerations. For better visualization, we do not show the method which is completely dominated by the variance-based correction in any of the plots.

5.3.3 Monte Carlo results

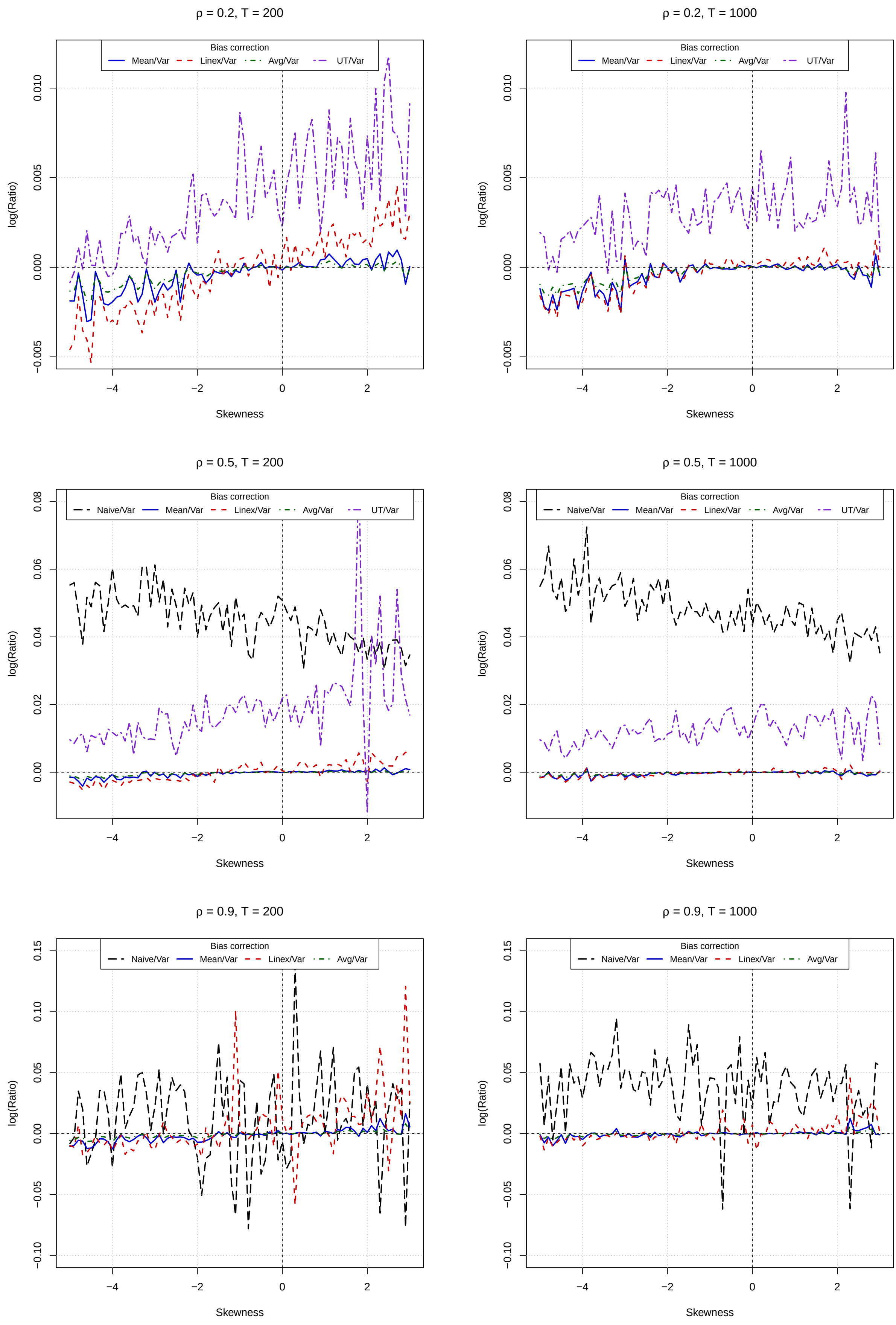
For all cases we consider both small samples with $T = 200$ and large samples with $T = 1000$. The results for Case I with skew-normal distribution of innovations is shown in Figure 5.1, where we plot the log MSE differences depending on the value of the skewness parameter β. The obtained evidence is quite similar for all $T = 200$ and $T = 1000$, but it varies with respect to the autocorrelation parameter value ρ.

For weak and medium autocorrelations $\rho = 0.2$ and $\rho = 0.5$, the Linex-based correction is the best for a pronounced negative skewness with parameter values $\beta < -2$. It is closely followed by mean-based and average-based corrections. For positive skewness, however, the variance-based correction appears to be mostly appropriate. This may be explained by how the parameters are estimated: for negative skewness, the Linex loss function mimics the negative log-density of innovations and, hence, the estimation under Linex is more efficient than OLS. Both naÿe and untransformed forecasts are much worse than the other procedures.

For strong autocorrelation $\rho = 0.9$ the mean-, variance-, and average-based forecasts are close to each other. The Linex is slightly better for negative and slightly worse for positive skewness values. For $T = 200$ the naÿe forecast is to some extent worse than the alternatives, however, it appears to be the worst for large sample size $T = 1000$. Again, this is most likely due to the nature of the estimation error.

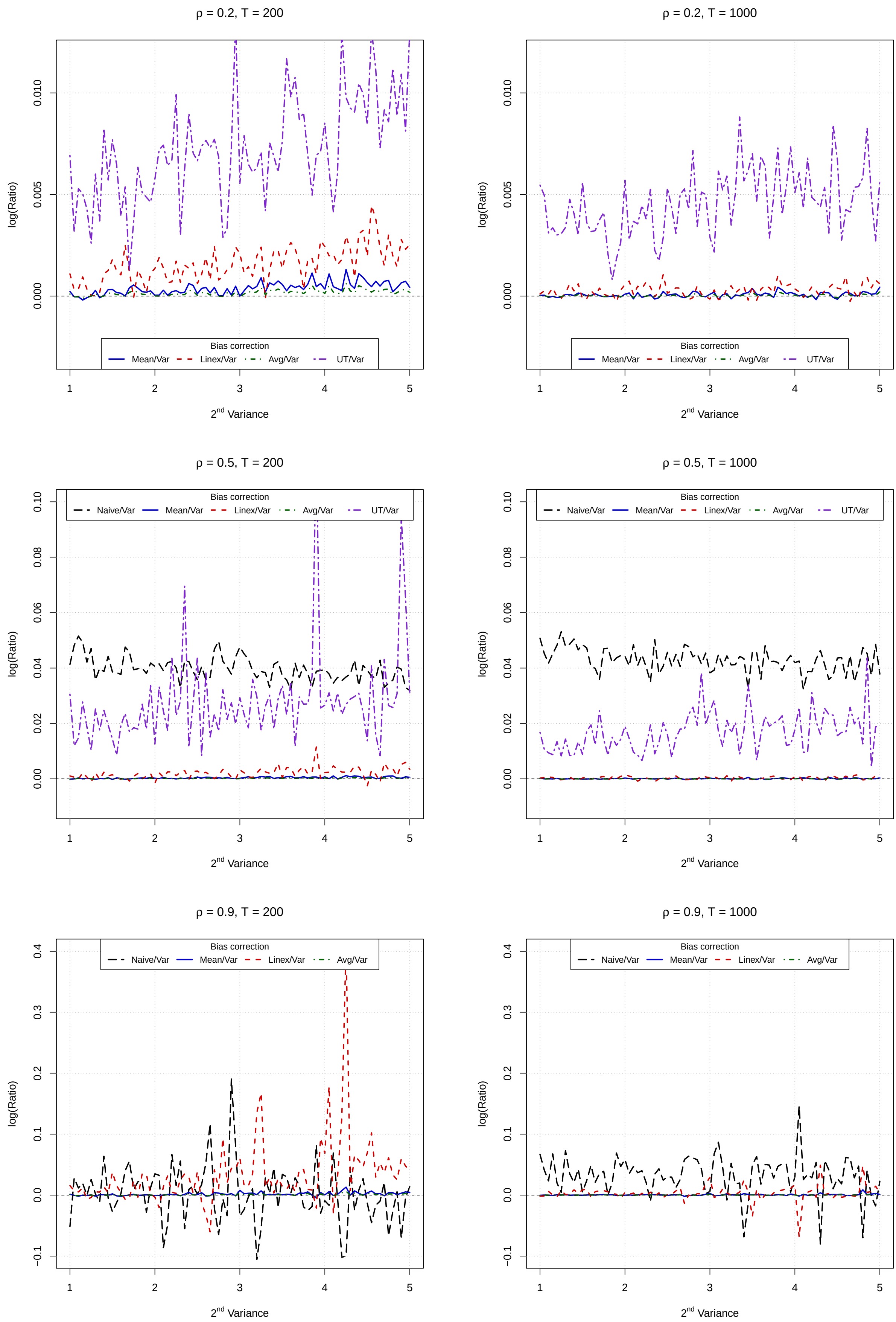
The log differences of MSEs in Case II with normal mixture are presented in Figures 5.2 for different values of σ_2^2. For AR(1) parameters $\rho = 0.2$ and $\rho = 0.5$ the variance correction method appears to be the best one for $T = 200$, whereas for $T = 1000$ the mean- and average-based corrections are quite close to it. The Linex-based forecast is dominated for $T = 200$ but gets close to the variance-based alternative for $T = 1000$. Again, the naïve uncorrected forecast appears to be reasonable for strong autocorrelation $\rho = 0.9$. The results for Case III with contaminated distribution are quantitatively similar to those for Case II so that we do not show them here.

Figure 5.1: Log MSE ratios for Case I



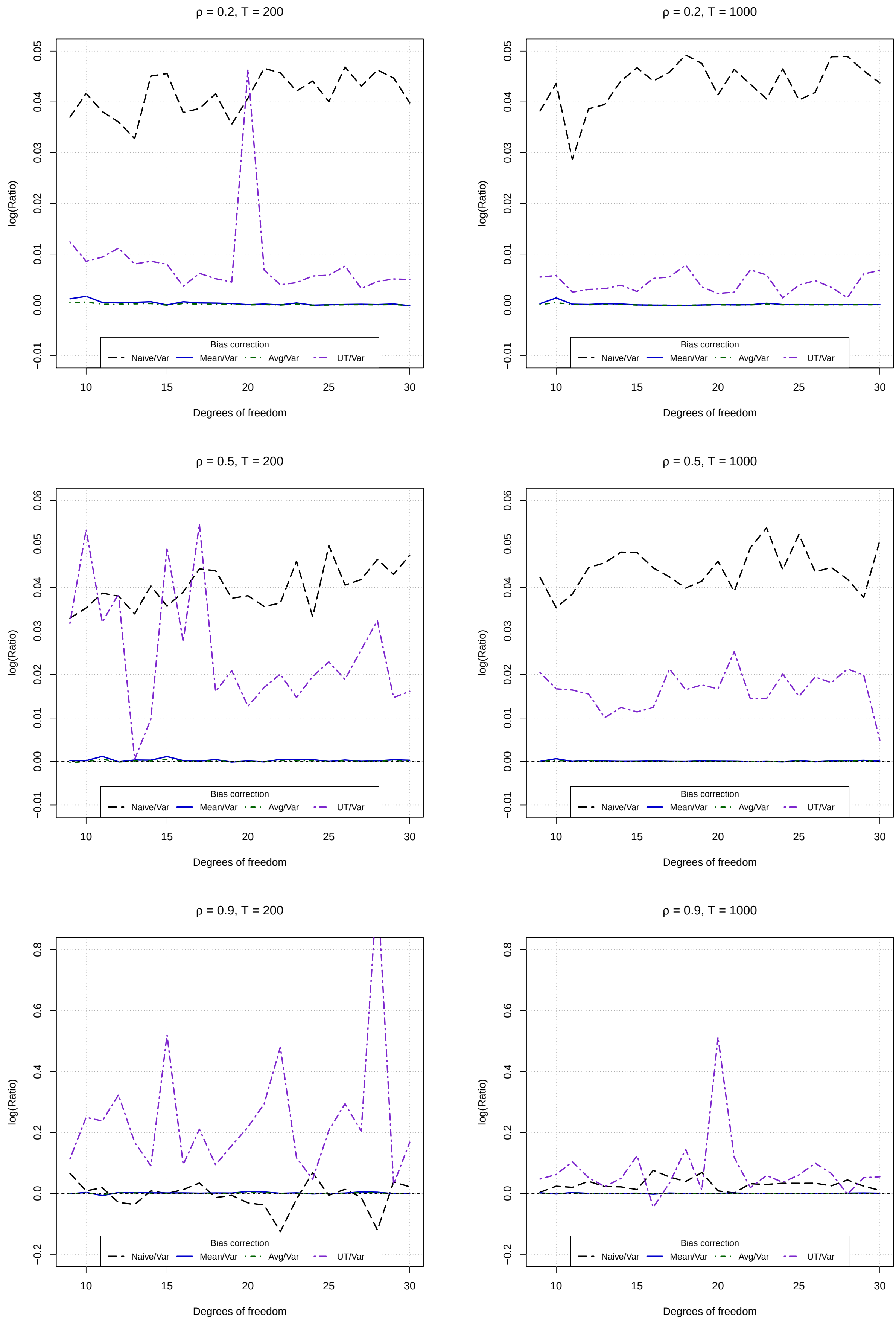
Notes: $\rho \in \{0.2, 0.5, 0.9\}$ from above to below, $T \in \{200, 1000\}$ from left to right, skewness parameter $\beta \in [-5,\ 3]$ with $\beta = 0$ for a symmetric distribution.

Figure 5.2: Log MSE ratios for Case II



Notes: $\rho \in \{0.2, 0.5, 0.9\}$ from above to below, $T \in \{200, 1000\}$ from left to right, mixture of $\mathcal{N}(0,1)$ and $\mathcal{N}(0,\sigma_2^2)$ with probability $\pi = 0.5$ and $\sigma_2^2 \in [1,\ 5]$.

Figure 5.3: Log MSE ratios for Case IV



Notes: $\rho \in \{0.2, 0.5, 0.9\}$ from above to below, $T \in \{200, 1000\}$ from left to right, t-distribution with $v \in [5,\ 30]$ degrees of freedom.

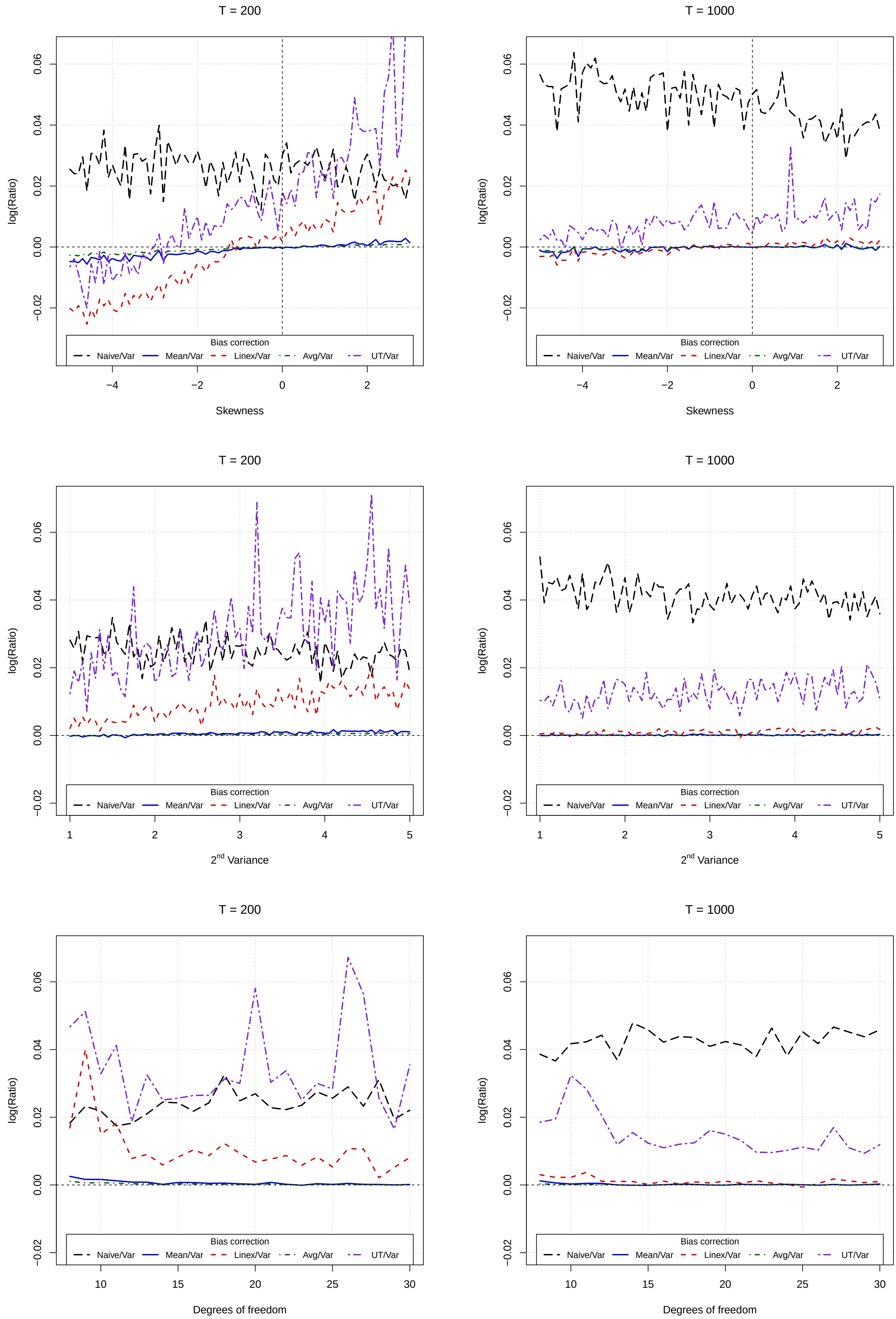
Next, in Case IV with t-distributed innovations shown in Figure 5.3 we observe that the variance-based and mean-based forecasts are the best for $\rho = 0.2$ and $\rho = 0.5$. As the Linex-based MSE appears to be numerically unstable, we do not recommend it for correction in case of t-distributed innovations and do not report Linex-based results for Case IV. As earlier, the advantages of the naïve forecast decrease with the increase of the sample size T, but for $\rho = 0.9$ and $T = 200$ the naïve uncorrected forecast should be used.

The log MSE ratios are shown in Figure 5.4 for the ARMA(1,1) model specification. The graphs in the top line correspond to Case I, in the medium line to Case II, and in the bottom line to Case IV, respectively. For Case I, the Linex-based correction is to use for negative skewness and to avoid for positive skewness values. In Cases II and IV, the Linex is dominated by variance-based corrections. The naïve and untransformed forecasts are worse than variance-based for all settings, whereas the mean-based and average corrections are mostly close to the variance-based benchmark. Thus, the results for AR(1) and ARMA(1,1) models are quite similar.

Finally, for Case IV with t-distributed innovations for the small estimation window $T = 200$ we investigate the performance of naïve forecasts with respect to the AR(1) coefficient ρ. The corresponding plots for $\rho \in \{0.75, 0.8, 0.85, 0.9, 0.95, 0.98\}$ are shown in Figure 5.5. As one could observe, the naïve uncorrected forecast is very close to the variance-based correction for $\rho = 0.9$, and gets much better for $\rho = 0.95$ and $\rho = 0.98$. Hence, we conclude that more persistent in the time series behavior speaks for a possible usage of naïve predictors.

Hence, our major findings for weak and medium autocorrelation coefficients $\rho \in \{0.2, 0.5\}$ are as follows. In Case I negative skewness β is in favor of the Linex-based method, whereas for positive β values this method becomes unstable and variance-based correction is preferable. In Cases II and III variance-based correction is slightly better than mean-based correction in case of normal mixture distribution. Case IV: variance correction is suitable for t-distributed innovations. However, higher values of the autoregressive parameter $|\rho|$ lead to more instability for all studied bias correction methods, so that for e.g. $\rho \geq 0.9$ and $T = 200$ no bias correction appears to be preferable.

Figure 5.4: Log MSE ratios for ARMA(1,1)



Notes: Innovations are skewed normal, normal mixture and Student-t innovations from above to below, $T \in \{200, 1000\}$ from left to right.

5.4 Empirical Illustration

The availability of intraday data allows us to estimate the true daily volatility σ_t^2 consistently by its realized measure y_t (cf. Andersen et al., 2007) which serves as the time series of interest $\{y_t\}$ in our study. We focus on the autoregressive model for realized volatility in logs in order to make forecasts of y_{t+1} conditional on the information set $\mathcal{F}_t$. For this purpose we contrast naïve uncorrected forecasts with those from the variance-based, mean-based, the average of variance- and mean-based, as well as Linex-based methods for the bias correction with the purpose of one step ahead prediction of daily realized volatility.

5.4.1 HAR model for daily realized volatility

The heterogenous autoregressive (HAR) model of Corsi (2009) appears to be rather successful for modelling and forecasting daily realized volatility. In order to assess the complex autoregressive structure of the process $\{y_t\}$, we exploit the HAR model which includes daily, weekly, monthly, and quarterly components (cf. Andersen et al., 2011):

$$y_{t+1} = \alpha_0 + \alpha_1 y_t + \alpha_2 y_t^{(w)} + \alpha_3 y_t^{(m)} + \alpha_4 y_t^{(q)} + \varepsilon_{t+1}, \tag{5.8}$$

with $y_t^{(w)} = (1/5) \cdot \sum_{i=0}^{4} y_{t-i}$, $y_t^{(m)} = (1/22) \cdot \sum_{i=0}^{21} y_{t-i}$, and $y_t^{(q)} = (1/65) \cdot \sum_{i=0}^{64} y_{t-i}$. Here the lag orders 5, 22, and 65 are the average number of weekly, monthly, and quarterly trading days, respectively.The HAR model (5.8) for daily volatility predictions based directly on the non-transformed realized volatility measures stands for the untransformed approach.

A considerable disadvantage of the specification in (5.8) is that the symmetry assumption for the distribution of ε_t is obviously violated due to pronounced impact discrepancies of positive and negative volatility shocks (cf. Tsay, 2010). For this reason a log transformation $x_t = \log y_t$ is commonly applied for the purpose of modelling (Andersen et al., 2007), since it levels out the asymmetries in the innovations. Then the corresponding HAR model in logs (cf. Corsi et al., 2012, Golosnoy et al., 2014) is:

$$x_{t+1} = \beta_0 + \beta_1 x_t + \beta_2 x_t^{(w)} + \beta_3 x_t^{(m)} + \beta_4 x_t^{(q)} + \epsilon_{t+1}, \tag{5.9}$$

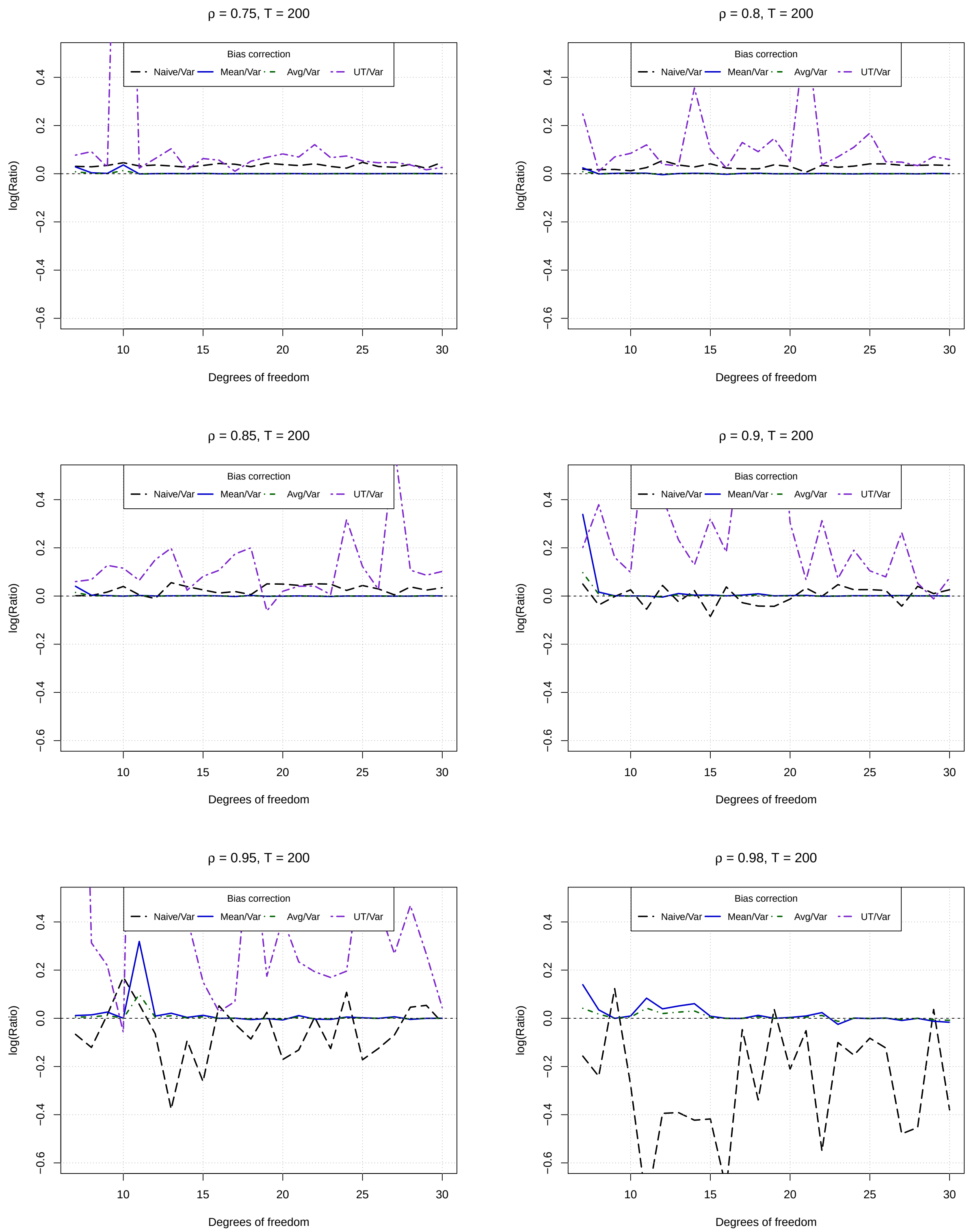
where $x_t^{(\cdot)}$ is defined analogously to $y_t^{(\cdot)}$.

As it holds $\exp[\mathrm{E}(x_{t+1}|\mathcal{F}_t)] \neq \mathrm{E}[\exp(x_{t+1})|\mathcal{F}_t]$, one requires a bias correction for the volatility forecasts. Hence, we estimate the model in (5.9) and make a forecast of y_{t+1} given the information set $\mathcal{F}_t$ by applying various types of bias corrections. Note that the non-normality of the innovations ϵ_{t+1} in log volatility processes is well-documented (cf. Lanne, 2006).

5.4.2 Data and descriptive statistics

We investigate daily realized volatilities for S&P500 index and three highly liquid US stocks, namely American Express, Exxon Mobil, and Microsoft which represent different sectors of

Figure 5.5: Log MSE ratios for AR(1) with Student-t innovations and $T = 200$ and increasing persistence.



the US economy. The realized volatility series of S&P500 is obtained from the Oxford Man Library whereas the daily volatility series for individual stocks are computed from 1 minute intraday returns taken from QuantQuote.com as realized kernel measures with the Parzen kernel (Barndorff-Nielsen et al., 2011).

Our time series cover the period ranging from December 31, 2001 until December 31, 2014 which results in 3255 daily realized variances for each asset. The considered time series are depicted in Figure 5.6 such that both calm and turmoil periods on U.S. financial markets are observed during the investigated period.

In order to investigate the properties of residuals from the log-HAR model in (5.9), we first estimate this model by the OLS based on the full-sample information. The parameter estimates are given in Table 5.1 where we also provide the estimates of the residual variance, skewness, and kurtosis. All regressor coefficients are significantly larger than zero supporting the selected HAR specification. The estimated models for all considered series show no unit root behavior as $\sum_{i=1}^{4} \hat{\beta}_i < 1$. The R^2 measures for all assets are quite high, between 0.6 and 0.8. For all four series the residuals appear to be right-skewed and exhibit kurtosis around four, i.e. the excess kurtosis around one. The autocorrelation functions (ACFs) for the original data and the residuals from the untransformed HAR models are shown in Figure 5.7, whereas the ACFs for the log-transformed data and the residuals from the log HAR model in Figure 5.8.

The hyperbolic decay of the ACF is observed for both original (untransformed) data and for the log-transformed series. The remaining residual autocorrelation is more pronounced for the untransformed HAR in Figure 5.7 than for the log HAR in Figure 5.8. This evidence could be seen as a support of modelling log realized measures by the HAR specifications. To summarize, these HAR models appear to provide a reasonable time series specifications for the log realized volatilities.

5.4.3 Comparison of bias correction methods

For making one step ahead volatility predictions, the log-HAR model in (5.9) is re-estimated based on the moving window of size $T \in \{200, 500, 750, 1000\}$ days. We set variance-based correction as a benchmark and compare it to the naïve, mean-based and Linex-based methods. Additionally, we consider the average of the mean- and variance-based forecasts as well as the untransformed forecasts from the model without logs in (5.8). The corresponding logs of MSE ratios for all assets – i.e., MSE increase in % compared to the variance-based corrections – are presented in Table 5.2.

Figure 5.6: Realized kernel time series.

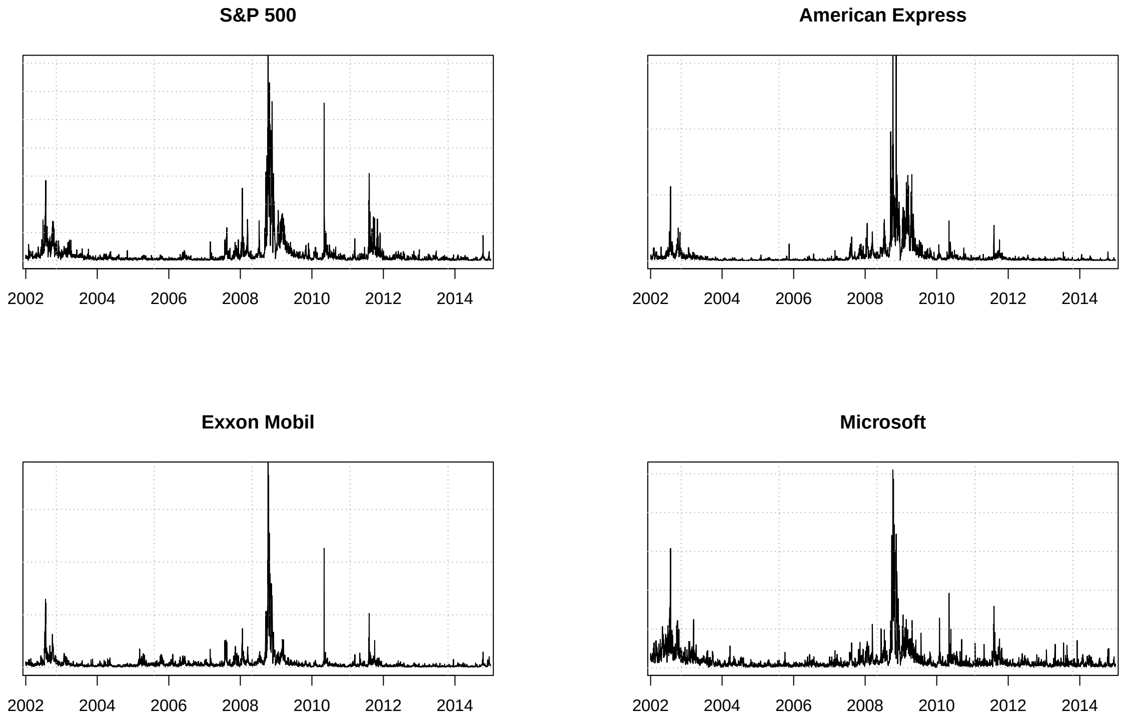


Table 5.1: Parameter estimates (st. errors) and descriptive statistics of residuals for the full sample log-HAR model in (5.9).

Company	β_0	$\beta^{(d)}$	$\beta^{(w)}$	$\beta^{(m)}$	$\beta^{(q)}$	R^2	$\widehat{\sigma}^2$	$\widehat{\xi}$	$\widehat{\kappa}$
S&P 500	−0.325 (0.114)	0.452 (0.017)	0.378 (0.028)	0.050 (0.037)	0.087 (0.029)	0.74	0.302	0.202	3.996
American Express	0.003 (0.010)	0.407 (0.017)	0.372 (0.029)	0.068 (0.041)	0.136 (0.031)	0.80	0.298	0.197	4.046
Microsoft	0.004 (0.010)	0.372 (0.017)	0.356 (0.031)	0.068 (0.045)	0.164 (0.037)	0.62	0.306	0.130	4.145
Exxon Mobil	0.001 (0.009)	0.437 (0.017)	0.400 (0.028)	0.013 (0.037)	0.107 (0.030)	0.68	0.246	0.237	4.229

The major findings for MSE are summarized as follows. The untransformed forecast is the worst one for all constellations. For $T = 200$, the naïve approach leads to the smallest MSE for single stocks, whereas the Linex correction is the best for the index S&P500. However, both naïve and Linex approaches get worse than the variance-based correction for larger values of estimation window T. The mean-based approach is slightly worse than the variance-based correction as well as the average of mean-based and variance-based approaches.

Figure 5.7: ACF of the data and corresponding HAR residuals.

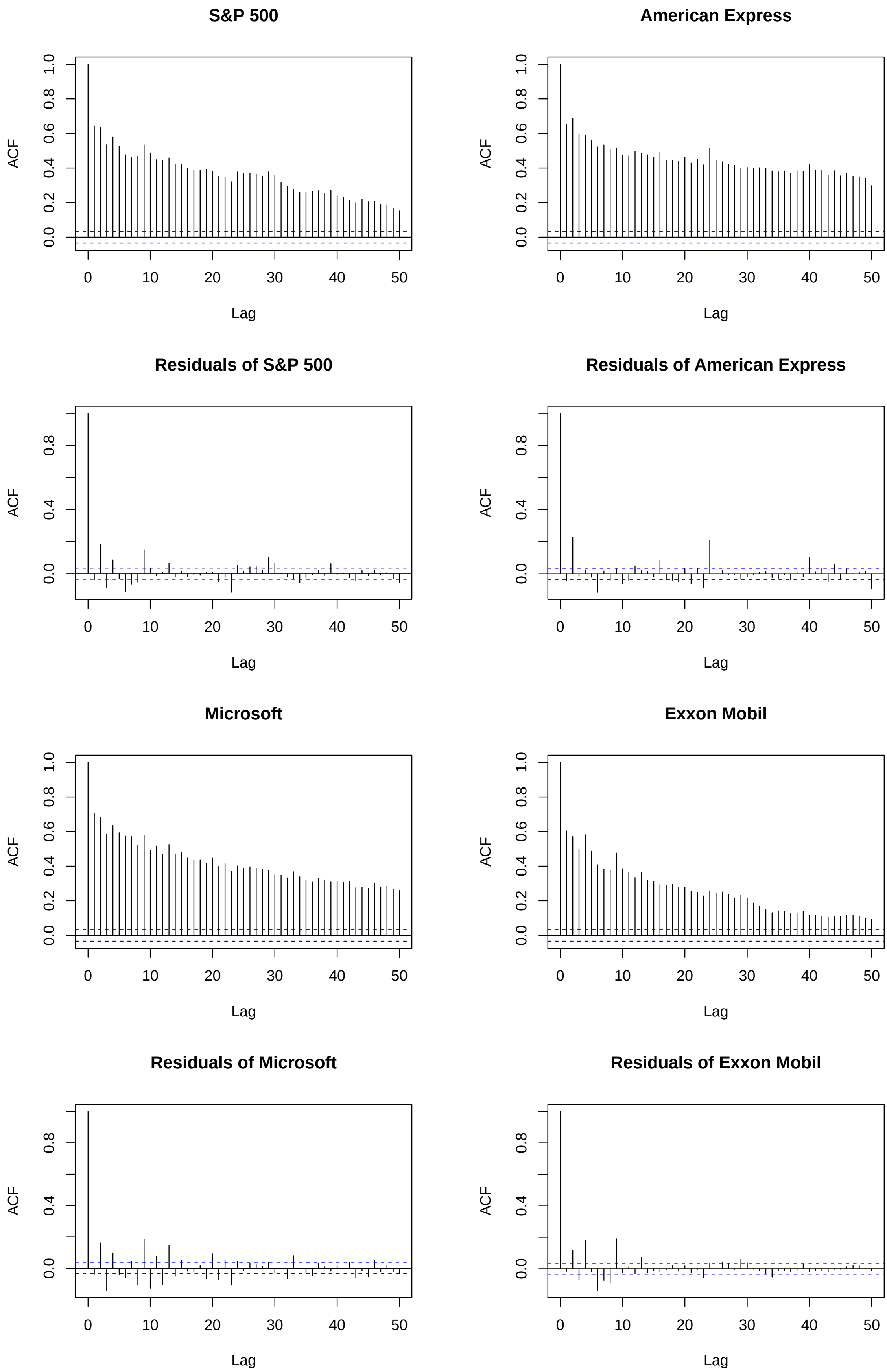


Figure 5.8: ACF of the data in logs and corresponding HAR residuals.

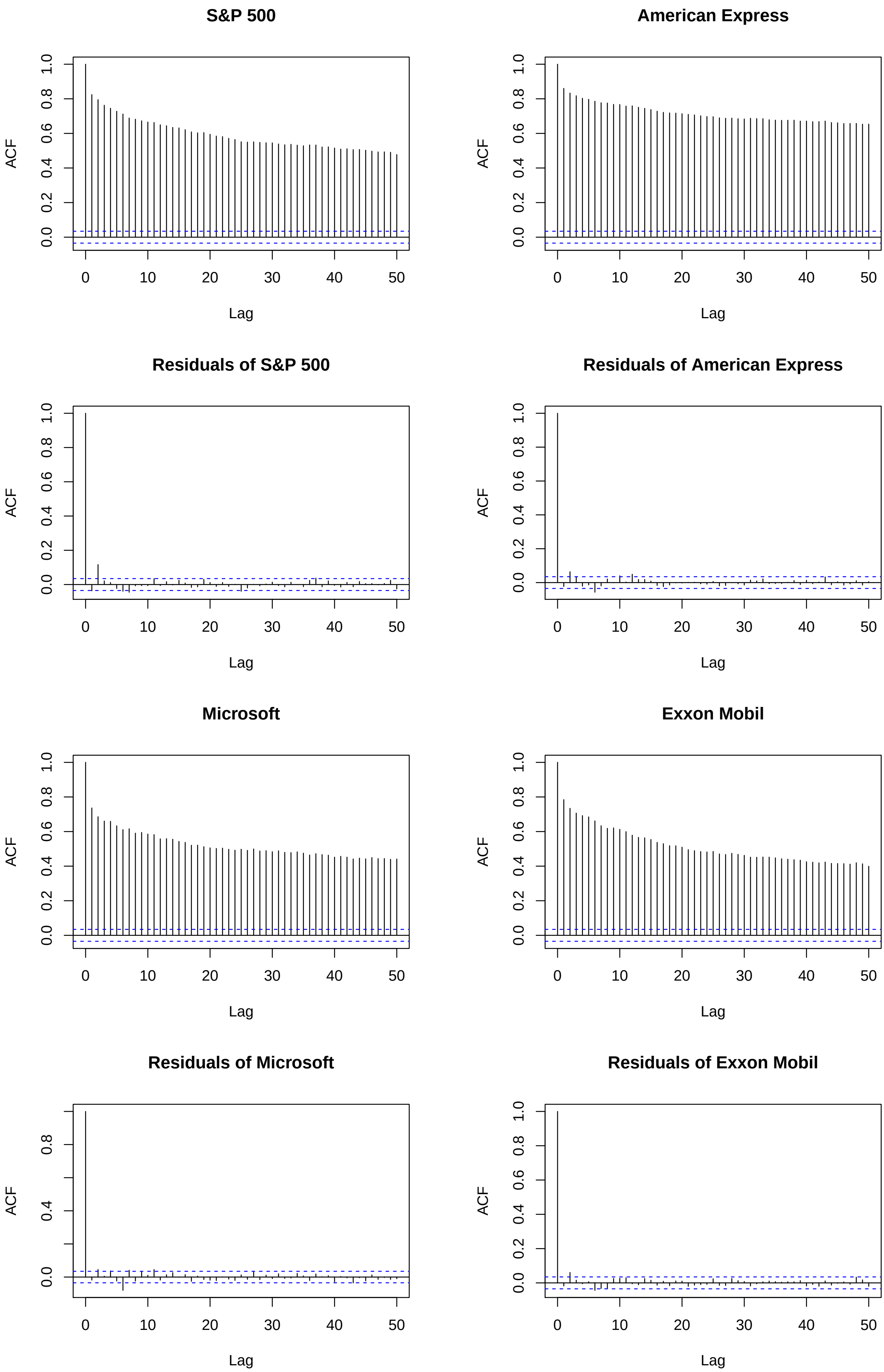


Table 5.2: Log of MSE ratios for the log-HAR model forecasts

			S&P 500		
T	naïve/var	mean/var	Linex/var	Avg/var	UT/var
200	−0.025	0.004	−0.033	0.002	0.441
500	0.008	0.003	0.009	0.001	0.457
750	0.017	0.002	0.021	0.001	0.467
1000	0.025	0	0.024	0	0.463
			American Express		
T	naïve/var	mean/var	Linex/var	Avg/var	UT/var
200	−0.034	0.002	−0.032	0.001	0.303
500	0.001	0.002	0.092	0.001	0.318
750	0.001	0.003	0	0.001	0.318
1000	0.011	0.003	0.01	0.002	0.323
			Exxon Mobil		
T	naïve/var	mean/var	Linex/var	Avg/var	UT/var
200	−0.038	0.004	−0.012	0.002	0.599
500	0.017	0.001	0.017	0	0.601
750	0.029	0	0.031	0	0.579
1000	0.037	−0.001	0.037	−0.001	0.553
			Microsoft		
T	naïve/var	mean/var	Linex/var	Avg/var	UT/var
200	−0.025	0.009	−0.024	0.004	0.028
500	0.034	0.001	0.054	0.001	0.071
750	0.039	0.001	0.035	0.001	0.074
1000	0.047	0.001	0.049	0	0.062

Notes: MSE increase in % compared to variance-based corrections for one day ahead volatility forecasts from the log-HAR model in (5.9) estimated based on moving windows of size T.

Note that although the numerical differences in the MSE in Table 5.2 are not so large, looking for the best point volatility forecast is still of much economic relevance. E.g., since volatility prediction is of importance for pricing derivative financial instruments, such as European and American options (cf. Tsay, 2010), even a small improvement of daily volatility forecasts could lead to substantial economic gains or losses. However, trying to use our approach in order to construct a profitable trading strategy clearly remains beyond the scope of our paper.

To assess the statistical significance of our results, we consider the popular Diebold-Mariano tests for equal predictive accuracy in order to compare the competing correction approaches. We compare the approaches under consideration pairwise and report in Table 5.3 the rejects of the benchmark in columns by '+' and the non-reject by '–' at the 5% significance level.

The forecasts from the untransformed model (UT) are statistically rejected in all settings. For the small estimation window $T = 200$, the Linex is significantly the best approach for the index S&P500 whereas the naïve forecast is the best for all three stocks. The variance-based correction appears to be statistically the best approach for almost all settings for $T \in \{500, 750, 1000\}$, with the exception of three cases where the Linex and mean-based forecasts are better. Summarizing our evidence, for small estimation windows one should rely on the Linex or even on naïve predictions, while for larger windows the variance-based correction is mostly appropriate.

5.5 Summary

Making forecasts with an autoregressive model for log-transformed variables is a convenient option in numerous applications. A reverse transformation in order to get the forecast of the original variable, would, however, introduce a bias that should be accounted for. For normally distributed innovations in the log-autoregressive models, the variance-based correction appears to be optimal. The alternative mean-based and Linex-based correction approaches require no distributional assumptions.

In this paper we investigate a finite sample MSE forecasting performances of several bias correction methods. Namely, we contrast a naïve no-correction approach, the variance-based correction with the mean-based and Linex-based corrections under empirically relevant deviations from normality of the error distribution.

We find that the sample size and the degree of autoregressive persistence are of most importance for the choice of the optimal correction strategy. For large samples where the estimation risk gets negligible, the Linex-based correction shows decent performance, however, in finite samples it is subject to numerical instabilities. The variance-based correction seems to be the best approach in finite samples, closely followed by the mean-based correction. The untransformed forecasts appear to be not reasonable when the model in logs is the correct one. Finally, in case of small samples and highly persistent autoregression, no correction at all appears to be a reasonable alternative.

Acknowledgements

This research has been in part financially supported by the Collaborative Research Center “Statistical modelling of nonlinear dynamic processes” (SFB 823, Teilprojekt A1) of the German Research Foundation (DFG).

Table 5.3: Diebold-Mariano test results for daily realized volatilities of S&P 500, American Express, Exxon, and Microsoft.

benchmark⟶	Naïve				Variance Corrected				Mean Corrected				Linex				Average				Untransformed			
window size	200	500	750	1000	200	500	750	1000	200	500	750	1000	200	500	750	1000	200	500	750	1000	200	500	750	1000
competitor ↓	S&P 500																							
Naïve					−	+	+	+	−	+	+	+	+	−	−	+	−	+	+	+	−	−	−	−
Var	+	−	−	−					−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−
Mean	+	−	−	−	+	+	+	+					+	−	−	−	+	+	+	+	−	−	−	−
Linex	−	+	+	−	−	+	+	+	−	+	+	+					−	+	+	+	−	−	−	−
Avg	+	−	−	−	+	+	+	+	−	−	−	−	+	−	−	−					−	−	−	−
UT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
competitor ↓	American Express																							
Naïve					−	+	+	+	−	−	−	+	−	−	+	+	−	−	−	+	−	−	−	−
Var	+	−	−	−					−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−
Mean	+	−	+	−	+	+	+	+					+	−	+	−	+	+	+	+	−	−	−	−
Linex	+	+	−	−	−	+	−	+	−	+	−	+					−	+	−	−	−	−	−	−
Avg	+	−	−	−	+	+	+	+	−	−	−	−	+	−	+	−					−	−	−	−
UT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
competitor ↓	Exxon Mobil																							
Naïve					−	+	+	+	−	+	+	+	−	+	−	+	−	+	+	+	−	−	−	−
Var	+	−	−	−					−	−	+	+	+	−	−	−	−	−	+	+	−	−	−	−
Mean	+	−	−	−	+	+	−	−					+	−	−	−	+	+	−	−	−	−	−	−
Linex	+	−	+	−	−	+	+	+	−	+	+	+					−	+	+	+	−	−	−	−
Avg	+	−	−	−	+	+	−	−	−	−	+	+	+	−	−	−					−	−	−	−
UT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
competitor ↓	Microsoft																							
Naïve					−	+	+	+	−	+	+	+	−	−	+	−	−	+	+	+	−	−	−	−
Var	+	−	−	−					−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−
Mean	+	−	−	−	+	+	+	+					+	−	−	−	+	+	+	+	−	−	−	−
Linex	+	+	−	+	−	+	+	+	−	+	+	+					−	+	+	+	−	−	−	−
Avg	+	−	−	−	+	+	+	+	−	−	−	−	+	−	−	−					−	−	−	−
UT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				

Notes: Columns represent benchmark models with rolling window sizes, $t \in \{200, 500, 750, 1000\}$; rows represent competitors. Under H_0 benchmark and competitor have equal predictive accuracy; under the alternative, benchmark is more accurate. Rejection at 5% level in favor of benchmark is shown by '+'; not rejecting by '–'.

Appendix

Consistency of Linex-based approach for log-transformation

We take the optimization to be conducted over a compact subset Θ of the parameter space guaranteeing stable autoregressions. Then, given the fact that the innovations ϵ_t are *iid*, the process x_t (which has a causal moving average representation in terms of ϵ_t with absolutely summable coefficients) exists a.s., and $(x_t, \epsilon_t)'$ is a jointly strictly stationary and ergodic process. Define now

$$b = \arg\min_{b^*} \mathrm{E}\left[\mathcal{L}\left(\epsilon_t - b^*\right)\right],$$

i.e. the M-measure of location of ϵ_t under $\mathcal{L}$. Recall that ϵ_t (and thus x_t) have thin tails, and the above expectation is therefore finite given the linear-exponential behavior of $\mathcal{L}$.

Note that $b = \log\left(\mathrm{E}\left[\exp\left(\epsilon_t\right)\right]\right)$, which is seen to be true since b must satisfy the f.o.c.

$$\mathrm{E}\left[\mathcal{L}'\left(\epsilon_t - b\right)\right] = 0,$$

i.e. $\mathrm{E}\left[\exp\left(\epsilon_t - b\right) - 1\right] = 0$ or $\mathrm{E}\left[\exp\left(\epsilon_t\right)\right] = \exp\left(b\right)$ as required.

Now, the empirical loss to be minimized is

$$\frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(y_t - \mu^* - \sum_{j=1}^{p} \rho_j^* x_{t-j}\right) = \frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(\epsilon_t - b - \left(\mu^* - \left(\mu + b\right)\right) - \sum_{j=1}^{p}\left(\rho_j^* - \rho_j\right) x_{t-j}\right).$$

Since b is such that the expected loss of $\mathcal{L}\left(\epsilon_t - b\right)$ is smallest, minimizing the empirical loss will result in estimators consistent for $\mu + b = \tilde{\mu}$ and ρ_j as we show below.

Since $\mathcal{L}$ is a strictly convex function of its argument, it follows that

$$\mathrm{E}\left[\mathcal{L}(\cdot)\right] = \mathrm{E}\left[\mathcal{L}\left(\epsilon_t - b - \left(\mu^* - \left(\mu + b\right)\right) - \sum_{j=1}^{p}\left(\rho_j^* - \rho_j\right) x_{t-j}\right)\right]$$

is a strictly convex function of θ^*. Note also that any linear combination of ϵ_t and (lags of) x_t must have thin tails, as an application of the Minkowski's inequality shows, therefore the above expectation is finite and the ergodic theorem indicates that

$$\frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(y_t - \mu^* - \sum_{j=1}^{p} \rho_j^* x_{t-j}\right) \xrightarrow{a.s.} \mathrm{E}\left[\mathcal{L}(\cdot)\right]$$

pointwise in θ^*. Compactness of Θ and convexity of allow us to use Thm. 10.8 in Rockafellar (1970) to conclude that pointwise a.s. convergence implies uniform convergence,

$$\sup_{\theta^* \in \Theta}\left|\frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(y_t - \mu^* - \sum_{j=1}^{p} \rho_j^* x_{t-j}\right) - \mathrm{E}\left[\mathcal{L}(\cdot)\right]\right| \xrightarrow{a.s.} 0;$$

furthermore, e.g. Thm. 4.1.1 in Amemiya (1985) indicates that

$$\arg\min_{\theta^*} \frac{1}{T} \sum_{t=p+1}^{T} \mathcal{L}\left(y_t - \mu^* - \sum_{j=1}^{p} \rho_j^* x_{t-j} \right) \xrightarrow{a.s.} \arg\min_{\theta^*} \; \mathrm{E}\left[\mathcal{L}(\cdot)\right].$$

Given that $\mathrm{E}\left[\mathcal{L}\left(\epsilon_t - b^*\right)\right]$ is minimized at $b^* = b$, it follows that $\mathrm{E}\left[\mathcal{L}(\cdot)\right]$ is minimized at $\tilde{\mu}$ and ρ_j as required. □

Box-Cox transformation

Here we discuss the possibility of implementing the forecast bias correction methodologies for a Box-Cox (BC) transformation given as

$$BC\left(y\right) = \frac{y^\lambda - 1}{\lambda} \quad \text{for} \quad y \geq 0, \quad \lambda \neq 0,$$

where the log transformation is obtained as the limit for $\lambda \to 0$. First note that a simple multiplicative decomposition of the optimal forecast like in Equation (5.2) is not available for any $\lambda \in (0,1)$.

Under the simplifying assumption that the distribution of $x_t = BC_\lambda(y_t)$ is approximately normal (which could reasonably be made for $0 < \lambda \ll 1$), it can be shown that (cf. Freeman and Modarres, 2006, Lemma 1)

$$\mathrm{E}\left[y_{T+1}|\mathcal{F}_T\right] \approx \left(\lambda \mathrm{E}\left[x_{T+1}|\mathcal{F}_T\right] + 1\right)^{1/\lambda} + \sum_{k\geq 1} \frac{\sigma^{2k}}{2^k k!} \left(\lambda \mathrm{E}\left[x_{T+1}|\mathcal{F}_T\right] + 1\right)^{1/\lambda - 2k} \left(\prod_{j=0}^{2k-1} (1 - j\lambda) \right).$$

For the special case of $1/\lambda \in \mathbb{N}$ it simplifies to

$$\mathrm{E}\left[y_{T+1}|\mathcal{F}_T\right] \approx \sum_{i=0}^{1/\lambda} \binom{1/\lambda}{i} \lambda^i \left(\lambda \mathrm{E}\left[x_{T+1}|\mathcal{F}_T\right] + 1\right)^{1/\lambda - i} \mathrm{E}\left[\epsilon_{T+1}^i\right],$$

which is hardly tractable in practice, even if one would truncate the sum on the right hand side for computational reasons. For this reason in case of BC-transformed series we would recommend to rely on bootstrap-based bias correction methods.

Still, one may obtain an analog to the Linex-based correction when $\lambda \ll 1$: write

$$y_{T+1}(1) = \mathrm{E}\left[\left(\lambda x_{T+1} + 1\right)^{1/\lambda} |\mathcal{F}_T\right];$$

if requiring the point forecast of x_{T+1}, m_{T+1}, to be transformed back for forecasting y_{T+1} using the inverse of the BC transformation, we arrive like for the case $\lambda = 0$ at the moment condition

$$\mathrm{E}\left[\left(\frac{\lambda x_{T+1} + 1}{\lambda m_{T+1} + 1} \right)^{1/\lambda} - 1 \middle| \mathcal{F}_T \right] = 0.$$

This is a legitimate GMM condition, which we may employ for estimating any parameters of the model for m_{t+1} the same way as in the case of the log transformation, but optimization is numerically more demanding than for the Linex loss. For $\lambda \ll 1$, we may write approximately

$$\mathrm{E}\left[\left(\frac{\lambda x_{t+1}+1}{\lambda m_{t+1}+1}\right)^{1/\lambda} - 1 \middle| \mathcal{F}_T\right] \approx \mathrm{E}\left[\left(1+\lambda\left(x_{t+1}-m_{t+1}\right)\right)^{(\lambda+1)/\lambda} - 1|\mathcal{F}_T\right] = 0$$

which may again be written as an extremum estimator minimizing the observed loss under the loss function

$$\mathcal{L}_\lambda(u) = \frac{1}{\lambda+1}\left(1+\lambda u\right)^{(\lambda+1)/\lambda} - u - \frac{1}{\lambda+1}.$$

This loss has the advantage of being in difference form, and is therefore less difficult numerically. Furthermore, $\mathcal{L}_\lambda$ converges to the Linex function for $\lambda \to 0$ and is actually the squared-error loss function for $\lambda = 1$.

Chapter 6

Concluding Remarks

This thesis offers various contributions to the existing literature on asymmetric loss functions in time series econometrics. First, we provide a theoretical basis for modelling and forecasting long autoregressions under a generalized asymmetric loss, when the latter is given exogenously. We support our findings with extensive statistical simulations and elaborate proofs. Additionally, we derive a Wold-type decomposition of the linear process into a regular and a predictable component. Here, an investigation of data generating processes exhibiting structural breaks is of further interest.

Second, we reverse the problem and perform inference on the parameters characterizing the loss function itself. To this end, we reexamine the forecast preferences of the European Commission. We replicate and extend the study of Christodoulakis and Mamatzakis (2009), where we provide a more robust inference methodology, as well as expand the dataset. We find that the authors' findings remain mostly confirmed, but observe more of a symmetric tendency in the EU Commission preferences. The next step in this analysis is to perform a rolling-window estimation in order to uncover dynamic changes in the asymmetries during calm and tumultuous economic periods.

Third, I combine existing model validation methods in order to provide a thorough routine for selecting the best performing model for value-at-risk forecasting. Here, I propose an extension of quantile autoregressions based on HAR model of Corsi (2009) and additional weakly exogenous regressors that have proven valid in related literature. These models show excellent and robust results on a large variety of data, which, to my knowledge has not yet been done to such an extent. Additionally, I propose a simple scoring system that has shown helpful in picking the best model according to the relevant criteria. Also, it is intriguing to investigate the lower tail of the conditional distribution of returns beyond the customary quantiles of 5% and 1% in terms of their dependence on external information.

Finally, we examine bias correction techniques for the reverse transformation of log transformed time series. We compare four correction methods, namely, a no-correction naïve method, variance- and mean-based approaches, and obtaining a forecast through a Linex loss function, in simulations as well as empirically. The simulation part has its main focus on deviations from normality in the data generating processes. We find that the variance-based correction approach is the best in finite samples when the persistence of the underlying series is rather moderate. On the other hand, when the process is highly persistent, no correction at all is the reasonable approach. These findings may show useful for practitioners in cases when the normality assumption is unreliable.

Bibliography

Amemiya, T. (1985). *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

Andersen, T. G. and T. Bollerslev (1998). Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *the Journal of Finance 53*(1), 219–265.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics 89*(4), 701–720.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review 93*(1), 38–62.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of international Economics 73*(2), 251–277.

Andersen, T. G., T. Bollerslev, and X. Huang (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics 160*(1), 176–189.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*(1), 75–93.

Andrews, D. W. K. (1992). Generic uniform convergence. *Econometric Theory 8*(2), 241–257.

Ardia, D., K. Boudt, and L. Catania (2019). Generalized autoregressive score models in r: The gas package. *Journal of Statistical Software 88*(6), 1–28.

Aretz, K., S. M. Bartram, and P. F. Pope (2011). Asymmetric loss functions and the rationality of expected stock returns. *International Journal of Forecasting 27*(2), 413–437.

Artis, M. and M. Marcellino (2001). Fiscal forecasting: The track record of the IMF, OECD and EC. *The Econometrics Journal 4*(1), 20–36.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*(2), 171–178.

Baillie, R. and T. Bollerslev (2000). The forward premium anomaly is not as bad as you think. *Journal International Money and Finance 19*(4), 471–488.

Bams, D., G. Blanchard, and T. Lehnert (2017). Volatility measures and value-at-risk. *International Journal of Forecasting 33*(4), 848–863.

Barndorff-Nielsen, O., P. R. Hansen, A. Lunde, and N. Shephard (2011). Subsampling realized kernels. *Journal of Econometrics 160*(1), 204–219.

Barndorff-Nielsen, O. E., S. Kinnebrock, and N. Shephard (2008). Measuring downside risk-realised semivariance.

Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of financial econometrics 2*(1), 1–37.

Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics 4*(1), 1–30.

Bauer, G. H. and K. Vorkink (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics 160*(1), 93–101.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics 2*(3), 489–502.

Berkowitz, J., P. Christoffersen, and D. Pelletier (2011). Evaluating value-at-risk models with desk-level data. *Management Science 57*(12), 2213–2227.

Bernardi, M. and L. Catania (2016). Comparison of value-at-risk models using the mcs approach. *Computational Statistics 31*(2), 579–608.

Bernardi, M. and L. Catania (2018). The model confidence set package for r. *International Journal of Computational Economics and Econometrics 8*(2), 144–158.

Bhansali, R. (1978). Linear prediction by autoregressive model fitting in the time domain. *The Annals of Statistics 6*(1), 224–231.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics 20*(1), 105–134.

Boero, G., J. Smith, and K. Wallis (2008). Evaluating a three-dimensional panel of point forecasts: the bank of england survey of external forecasters. *International Journal of Forecasting 24*(3), 354–367.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The Review of Economics and Statistics 72*, 498–505.

Bondon, P. (2009). Estimation of autoregressive models with epsilon-skew-normal innovations. *Journal of Multivariate Analysis 100*(8), 1761–1776.

Brechmann, E. C., M. Heiden, and Y. Okhrin (2018). A multivariate volatility vine copula model. *Econometric Reviews 37*(4), 281–308.

Brillinger, D. R. (1975). *Time series: Data analysis and theory.* Holt, Rinehart and Winstan, New York.

Busch, T., B. J. Christensen, and M. Ø. Nielsen (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics 160*(1), 48–57.

Campbell, J. Y. and M. Yogo (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics 81*(1), 27–60.

Capistrán, C. (2008). Bias in Federal Reserve inflation forecasts: Is the Federal Reserve irrational or just cautious? *Journal of Monetary Economics 55*(8), 1415–1427.

Christodoulakis, G. and E. Mamatzakis (2008). An assessment of the EU growth forecasts under asymmetric preferences. *Journal of Forecasting 27*(6), 483–492.

Christodoulakis, G. and E. Mamatzakis (2009). Assessing the prudence of economic forecasts in the EU. *Journal of Applied Econometrics 24*(4), 583–606.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841–862.

Christoffersen, P. F. and F. X. Diebold (1997). Optimal prediction under asymmetric loss. *Econometric Theory 13*(6), 808–817.

Clatworthy, M. A., D. A. Peel, and P. F. Pope (2012). Are analysts' loss functions asymmetric? *Journal of Forecasting 31*(8), 736–756.

Clements, M., F. Joutz, and H. Stekler (2007). An evaluation of the forecasts of the Federal Reserve: A pooled approach. *Journal of Applied Econometrics 22*(1), 121–136.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Corsi, F., F. Audrino, and R. Renò (2012). HAR Modeling for Realized Volatility Forecasting. In *Handbook of Volatility Models and Their Applications*, pp. 363–382. Hoboken, NJ: John Wiley & Sons.

Corsi, F., D. Pirino, and R. Reno (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics 159*(2), 276–288.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Danıelsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance 26*(7), 1273–1296.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press.

Demetrescu, M. (2006). An extension of the Gauss–Newton algorithm for estimation under asymmetric loss. *Computational Statistics & Data Analysis 50*(2), 379–401.

Demetrescu, M. and U. Hassler (2016). (when) do long autoregressions account for neglected changes in parameters? *Econometric Theory 32*(6), 1317–1348.

Demetrescu, M. and S. H. Hoke (2019). Predictive regressions under asymmetric loss: factor augmentation and model selection. *International Journal of Forecasting 35*(1), 80–99.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance 1*(1), 83–106.

Douc, R., E. Moulines, and D. Stoffer (2014). *Nonlinear time series: theory, methods and applications with R examples.* CRC Press.

Dowd, K. (2007). *Measuring market risk.* John Wiley & Sons.

Dumitrescu, E.-I. and P. R. Hansen (2016). Parameter estimation with out-of-sample objective. Technical report.

Dumitrescu, E.-I., C. Hurlin, and V. Pham (2012). Backtesting value-at-risk: from dynamic quantile to dynamic binary tests. *Finance 33*(1), 79–112.

Elliott, G., I. Komunjer, and A. Timmermann (2005). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies 72*(4), 1107–1125.

Elliott, G., I. Komunjer, and A. Timmermann (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association 6*(1), 122–157.

Embrechts, P., C. Klüppelberg, and T. Mikosch (2013). *Modelling extremal events: for insurance and finance*, Volume 33. Springer Science & Business Media.

Engle, R., D. Lilien, and R. Robins (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica 55*(2), 391–407.

Engle, R. F. and S. Manganelli (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*(4), 367–381.

Everitt, B. S. and D. J. Hand (1981). *Finite Mixture Distributions.* London: Chapman and Hall.

Fernández, C. and M. F. Steel (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association 93*(441), 359–371.

Freeman, J. and R. Modarres (2006). Inverse Box–Cox: the power-normal distribution. *Statistics & Probability Letters 76*(8), 764–772.

Fritsche, U., C. Pierdzioch, J.-C. Rülke, and G. Stadtmann (2015). Forecasting the Brazilian real and the Mexican peso: Asymmetric loss, forecast rationality, and forecaster herding. *International Journal of Forecasting 31*(1), 130–139.

Gaglianone, W. P., L. R. Lima, O. Linton, and D. R. Smith (2011). Evaluating value-at-risk models via quantile regression. *Journal of Business & Economic Statistics 29*(1), 150–160.

Ghalanos, A. (2019). *rugarch: Univariate GARCH models.* R package version 1.4-1.

Giot, P. and S. Laurent (2004). Modelling daily value-at-risk using realized volatility and arch type models. *journal of Empirical Finance 11*(3), 379–398.

Giot, P. and S. Laurent (2007). The information content of implied volatility in light of the jump/continuous decomposition of realized volatility. *Journal of Futures Markets: Futures, Options, and Other Derivative Products 27*(4), 337–359.

Golosnoy, V., A. Hamid, and Y. Okhrin (2014). The empirical similarity approach for volatility prediction. *Journal of Banking and Finance 40*, 321–329.

Golosnoy, V., I. Okhrin, and W. Schmid (2012). Statistical surveillance of volatility forecasting models. *Journal of Financial Econometrics 10*(3), 513–543.

Gonçalves, S. and L. Kilian (2007). Asymptotic and bootstrap inference for ar (infinity) processes with conditional heteroskedasticity. *Econometric Reviews 26*(6), 609–641.

González-Rivera, G., T.-H. Lee, and S. Mishra (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of forecasting 20*(4), 629–645.

Gospodinov, N. (2009). A new look at the forward premium puzzle. *Journal of Financial Econometrics 7*(3), 312–338.

Granger, C. (1999). Outline of forecast theory using generalized cost functions. *Spanish Economic Review 1*(2), 161–173.

Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operational Research Society 20*(2), 199–207.

Granger, C. W. J. and P. Newbold (1976). Forecasting transformed series. *Journal of the Royal Statistical Society Series B 38*(2), 189–203.

Gribisch, B. (2018). A latent dynamic factor approach to forecasting multivariate stock market volatility. *Empirical Economics 55*(2), 621–651.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory 8*(4), 489–500.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Haugom, E., R. Ray, C. J. Ullrich, S. Veka, and S. Westgaard (2016). A parsimonious quantile regression model to forecast day-ahead value-at-risk. *Finance Research Letters 16*, 196–207.

Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data.* Heidelberg: Springer.

Heber, G., A. Lunde, N. Shephard, and K. Sheppard (2009). Oxford-man institutes realized library, version 0.3.

Jeon, J. and J. W. Taylor (2013). Using CAViaR models with implied volatility for value-at-risk estimation. *Journal of Forecasting 32*(1), 62–74.

Koenker, R. (2012). Quantile regression in r: A vignette. *Retrieved November*.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.

Koenker, R. and Z. Xiao (2006). Quantile autoregression. *Journal of the American Statistical Association 101*(475), 980–990.

Kuester, K., S. Mittnik, and M. S. Paolella (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics 4*(1), 53–89.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *FEDS Paper* (95-24).

Lanne, M. (2006). A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics 4*(4), 594–616.

Laurent, J.-P. (2017). The knowns and the known unknowns of capital requirements for market risks. In *Financial Regulation in the EU*, pp. 277–307. Springer.

Liu, W. and A. Maynard (2005). Testing forward rate unbiasedness allowing for persistent regressors. *Journal of Empirical Finance 12*(5), 613–628.

Louzis, D. P., S. Xanthopoulos-Sisinis, and A. P. Refenes (2014). Realized volatility models and alternative value-at-risk prediction strategies. *Economic Modelling 40*, 101–116.

Lütkepohl, H. and F. Xu (2012). The role of the log transformation in forecasting economic variables. *Empirical Economics 42*(3), 619–638.

Martens, M., D. Van Dijk, and M. De Pooter (2009). Forecasting s&p 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of forecasting 25*(2), 282–303.

Maynard, A. and P. C. B. Phillips (2001). Rethinking an old empirical puzzle: Econometric evidence on the forward discount anomaly. *Journal of Applied Econometrics 16*(6), 671–708.

Mayr, J. and D. Ulbricht (2015). Log versus level in VAR forecasting: 42 million empirical answers – Expect the unexpected. *Economics Letters 126*, 40–42.

McCullough, B. D. (2000). Optimal prediction with a general loss function. *Journal of Combinatorics, Information and System Sciences 25*(1–4), 207–221.

McLachlan, G. and D. Peel (2004). *Finite Mixture Models.* New York, NY: John Wiley & Sons.

Müller, U. K. and G. Elliott (2003). Tests for unit roots and the initial condition. *Econometrica 71*(4), 1269–1286.

Müller, U. K. and M. W. Watson (2008). Testing models of low-frequency variability. *Econometrica 76*(5), 979–1016.

Nieto, M. R. and E. Ruiz (2016). Frontiers in var forecasting and backtesting. *International Journal of Forecasting 32*(2), 475–501.

Patton, A. and A. Timmermann (2007a). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association 102*(480), 1172–1184.

Patton, A. J. and K. Sheppard (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics 97*(3), 683–697.

Patton, A. J. and A. Timmermann (2007b). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics 140*(2), 884–918.

Phillips, P. C. B. (1987). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika 74*(3), 535–547.

Pierdzioch, C., J. Rülke, and G. Stadtmann (2011). On the loss function of the Bank of Canada: A note. *Economics Letters 115*, 155–159.

Poskitt, D. S. (2007). Autoregressive approximation in nonstandard situations: The fractionally integrated and non-invertible cases. *Annals of the Institute of Statistical Mathematics 59*(4), 697–725.

Proietti, T. and H. Lütkepohl (2013). Does the Box–Cox transformation help in forecasting macroeconomic time series? *International Journal of Forecasting 29*(1), 88–99.

R-Core-Team (2014). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation of Statistical Computing.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton, NJ: Princeton University Press.

Rubia, A. and L. Sanchis-Marco (2013). On downside risk predictability through liquidity and trading activity: A dynamic quantile approach. *International Journal of Forecasting 29*(1), 202–219.

Seidel, W. (2011). Mixture Models. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science*, pp. 827–829. Springer Berlin Heidelberg.

Sharafi, M. and A. R. Nematollahi (2016). AR(1) model with skew-normal innovations. *Metrika 79*(8), 1011–1029.

Stock, J. H. and M. W. Watson (2012). *Introduction to Econometrics* (3 ed.). Boston, MA: Pearson.

Tarami, B. and M. Pourahmadi (2003). Multivariate t autoregressions: Innovations, prediction variances and exact likelihood equations. *Journal of Time Series Analysis 24*(6), 739–754.

Taylor, N. (2017). Realised variance forecasting under Box-Cox transformation. *International Journal of Forecasting 33*, 770–785.

Thombs, L. A. and W. R. Schucany (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association 85*(410), 486–492.

Torres, V. A. (1986). A note on teaching infinite moving averages. *The American Statistician 40*(1), 40–41.

Tsay, R. S. (2010). *Analysis of Financial Time Series* (3 ed.). Hoboken, NJ: John Wiley & Sons.

Tsuchiya, Y. (2016). Assessing macroeconomic forecasts for Japan under an asymmetric loss function. *International Journal of Forecasting 32*(2), 233 – 242.

Varian, H. R. (1975). A Bayesian Approach to Real Estate Assessment. In S. E. Fienberg and A. Zellner (Eds.), *Studies in Bayesian Econometric and Statistics in Honor of Leonard J. Savage*, pp. 195–208. Amsterdam: North Holland.

Wang, Y. and T.-H. Lee (2014). Asymmetric loss in the Greenbook and the Survey of Professional Forecasters. *International Journal of Forecasting 30*(2), 235 – 245.

Weiss, A. A. (1996). Estimating time series models using the relevant cost function. *Journal of Applied Econometrics 11*(5), 539–560.

Weiss, A. A. and A. P. Andersen (1984). Estimating time series models using the relevant forecast evaluation criterion. *Journal of the Royal Statistical Society. Series A (General) 147*(3), 484–487.

White, H. (2001). *Asymptotic theory for econometricians.* Academic Press New York.

Wong, Z. Y., W. C. Chin, and S. H. Tan (2016). Daily value-at-risk modeling and forecast evaluation: The realized volatility approach. *The Journal of Finance and Data Science 2*(3), 171 – 187.

Zernov, S., V. Zinde-Walsh, and J. W. Galbraith (2009). Asymptotics for estimation of quantile regressions with truncated infinite-dimensional processes. *Journal of Multivariate Analysis 100*(3), 497–508.

Žikeš, F. and J. Baruník (2014). Semi-parametric conditional quantile models for financial returns and realized volatility. *Journal of Financial Econometrics 14*(1), 185–226.