

Ordinal-response GARCH models for transaction data: A forecasting exercise

Stefanos Dimitrakopoulos^{1*} and Mike Tsionas^{**}

^{*}Economics Division, Leeds University, UK

^{**}Department of Economics, Lancaster University, UK

Abstract

Using numerous high-frequency transaction data sets, we evaluate the forecasting performance of several dynamic ordinal-response time series models with generalized autoregressive conditional heteroscedasticity (GARCH). The specifications account for three components; leverage effects, in-mean effects and moving average error terms. To estimate the model parameters we develop Markov chain Monte Carlo algorithms. Our empirical analysis showed that the proposed ordinal-response GARCH models achieve better point and density forecasts than standard benchmarks.

JEL CODE: C11, C13, C18, C22, C50, G00

¹Correspondence to: Stefanos Dimitrakopoulos, Economics Division, Leeds University Business School, Leeds University, UK, E-mail: s.dimitrakopoulos@leeds.ac.uk. We are grateful to Marwan Izzeldin and the GOLCER center at Lancaster University for providing us with some of the data in this paper. All computations have been performed with the generous technical support of GOLCER and its staff.

1 Introduction

Since the seminal work of Bollerslev (1986), several extensions of the generalized autoregressive conditional heteroscedastic (GARCH) model have been proposed. Such extensions include, among others, the autoregressive moving average (ARMA) component (Nakatsuma, 2000), the leverage effect (Nelson, 1991; Engle and Ng, 1993; Glosten et al., 1993, among others) and the conditional heteroscedasticity in mean. Incorporating the aforementioned empirical facts of financial time series in the context of GARCH models, leads to well-documented forecast gains (Awartani & Corradi, 2005; Christensen et al., 2010; Corsi & Renò, 2012; Chan and Grant, 2016).

However, financial time series do not always follow a continuous path but there is a discreteness in the responses that can not be captured by continuous-response models. One such series is related to transaction - by - transaction data in security markets, which are the ultimate high-frequency data in finance. A distinguishing feature of trading data is that the price change of a stock (an asset) from one transaction to the next one occurs in integer multiples of a tick size (tick-by-tick data).

High-frequency price changes can not, therefore, be treated as a continuous variable. Ignoring price discreteness can affect severely the empirical distribution of stock returns (Chordia and Subrahmanyam, 1995, Cordella and Foucault, 1999, Ronen and Weaver, 2001, Chakravarty et al., 2004, Münnix et al., 2010, O'Hara et al., 2014, and Ye and Yao, 2014).

The discreteness has motivated the development of models for high-frequency integer price changes that have also accounted for conditional heteroscedasticity (Hausman et al., 1992, Dueker, 1999). For instance, Hausman et al. (1992) proposed an ordered probit model for the analysis of this type of data, which was extended with an autoregressive term by Müller and Czado, (2005). Müller and Czado (2009) defined an ordinal-response stochastic volatility model, Barra and Koopman (2017) presented a negative binomial difference model with stochastic volatility, while Koopman et al. (2017) set up a dynamic Skellam model. Other researchers developed Tobit GARCH models (Calzolari and Fiorentini, 1998), ordered probit GARCH models (Yang and Parwada, 2012) and probit GARCH models (Ahmed, 2015). See, also, Bianco & Renò (2006, 2009).

Yet, the forecasting behaviour of discrete-response time series models with conditional het-

eroscedastic remains an uncharted region, when they are extended to account for additional empirical facts of financial data. In particular, does the forecast ability of these models increase, when they are equipped with (AR)MA components, or/and leverage or/and in-mean effects? These issues have not been examined before. The present paper aspires to fill these gaps.

With these thoughts in mind, our objective is to investigate the forecasting performance of a new class of discrete-response models with conditional heteroscedasticity. Since this paper deals with discrete stock price changes that occur in clusters, the order of which matters, the series of price changes can be regarded as an ordinal-valued high-frequency financial time series. As such, we propose a novel ordinal-response model with a GARCH-type conditional heteroscedasticity.

The novelty of our model is that it accounts for potential microstructure effects that have been disregarded in a discrete-response setting so far. In particular, our model is equipped with leverage, in-mean effects and a moving average error term. It is also enriched with an autoregressive structure in the conditional mean to control for time-dependence that usually characterises discrete (macro)financial data (Müller and Czado, 2005).

The resulting model specification is an autoregressive moving average (ARMA) ordinal-response (OR) GARCH model with leverage (L) and in-mean effects (M). We name this specification the ARMA-OR-GARCHLM model.

To control for leverage effects, we exploit the threshold GARCH (TGARCH) approach by Glosten et al. (1993). Moreover, the presence of the moving average part in our model tackles a problem related to transaction data according to which prices suffer from bid-ask bounce (Roll, 1984, and Campbell et al., 1997). Depending on whether or not the last trade was a buy or sell, the registered trade price difference will be almost surely positive or negative. As a result, a lag-1 serial correlation is usually observed in the series of price changes. A moving average term could potentially rectify this.

Our contribution is twofold. First of all, using a large number of transaction data sets (ten) from the Trades and Quotes (TAQ) database of the New York Stock Exchange (NYSE), we perform a substantive forecasting exercise that lacks in the literature on discrete-response time series models. In particular, the proposed model is compared against several nested versions of it. In this way, we examine the forecasting behaviour of ordinal-response GARCH variants

(with AR, MA, leverage, and/or in mean effects).

The forecasting ability of the competing models is evaluated by conducting a recursive out-of-sample forecasting analysis, where point and density forecasts are computed. Moreover, we exploit the conditional predictive ability (CPA) test of Giacomini and White (2006), which is a pairwise test of the accuracy of competing forecasting results, while we also perform multiple testings, using the Model Confidence Set (MCS) approach (Hansen et al., 2011). In addition, we conduct optimal prediction pooling, proposed by Geweke and Amisano (2011) that combines predictive densities from various models, with the aim of selecting the optimal model combination (pool). Aside from the forecast comparison metrics, we also simulated a simple trading rule for buying or selling a stock based on the forecasts, as these metrics are hard to tell if the improvements are large or small. Then, we calculated the trading profits generated by the forecasts.

Second, the ARMA-OR-GARCHLM model entails estimation challenges due to its latent nature, its autoregressive structure, the inclusion of the leverage and moving average components and the presence of conditional heteroscedasticity both in its conditional mean and its conditional variance. Therefore, we resort to Bayesian methods and design a Markov chain Monte Carlo (MCMC) method in order to estimate all parameters of our main specification. Our estimation approach exploits the Riemann manifold Metropolis adjusted Langevin algorithm (RMMALA) of Girolami and Calderhead (2011).

At this point it is worth mentioning that we excluded from the present analysis additional market microstructure effects, such as irregularly spacing in time. We believe that it is important to investigate first how the proposed models of this paper behave in terms of forecasting performance, before equipping them with additional characteristics of trading data.

The paper is structured as follows. In section 2 we set up the model and in section 3 we describe the posterior analysis. In section 4 we carry out our empirical study. Section 5 concludes. An Online Appendix accompanies this paper.

2 Econometric set up: The main specification

Suppose that $\{y_t, t = 1, \dots, T\}$ is a time series of ordinal responses, where each y_t takes on any one of the J ordered values in the range $1, \dots, J$ such that $P(y_t = j) = p_{tj}$ for $j = 1, \dots, J$ and

$$\sum_{j=1}^J p_{tj} = 1 \text{ for } t = 1, \dots, T.$$

The underlying mechanism that induces the ordinal structure in the observed process y_t lies in the existence of a latent variable y_t^* , which is connected to y_t are according to the following mapping specification

$$y_t = j \Leftrightarrow \zeta_{j-1} < y_t^* \leq \zeta_j, \quad 1 \leq j \leq J. \quad (1)$$

Relationship (1) implies that y_t is observed in category j if y_t^* lies in the interval demarcated by the cutpoints ζ_{j-1} and ζ_j . In order to ensure that the cumulative distribution function for y_t is properly defined, we require that $\zeta_j > \zeta_{j-1}, \forall j$, with $\zeta_0 = -\infty$ and $\zeta_J = +\infty$. A typical restriction in this type of models is to set a cutpoint equal to zero (usually the second one, $\zeta_1 = 0$) and another one equal to 1 (usually the second to last, $\zeta_{J-1} = 1$); see, for example, Müller and Czado (2009) and Dimitrakopoulos and Dey (2017).

For the latent variable y_t^* we consider the following novel latent autoregressive moving average model with conditional heteroscedasticity σ_t^2 ,

$$y_t^* = \rho y_{t-1}^* + \mathbf{z}_t' \boldsymbol{\delta} + \tau_1 \sigma_t^2 + \varepsilon_t + u_t + \gamma u_{t-1}, t = 1, \dots, T, \quad (2)$$

$$\varepsilon_t \sim N(0, \sigma_t^2), \quad u_t \sim N(0, 1), \quad (3)$$

$$\sigma_t^2 = \alpha_0 + (\alpha_1 + \tau_2 \mathbf{1}(\varepsilon_{t-1} < 0)) \varepsilon_{t-1}^2 + \alpha_2 \sigma_{t-1}^2. \quad (4)$$

In equation (2), y_{t-1}^* is the single one-period lagged latent dependent variable that captures time dependence in the observed sequence of y_t , \mathbf{z}_t is a vector of covariates and $\boldsymbol{\delta}$ is a constant coefficient vector of dimension $k \times 1$. Also, expression (2) allows for a moving average error process, where the error term u_t follows a normal distribution with mean zero and a variance which is normalized to one for identification. This process satisfies the invertibility condition; that is, $|\gamma| < 1$. The stochastic disturbance ε_t is normally distributed with time-varying conditional variance σ_t^2 .

Conditional heteroscedasticity is captured by a GARCH(1,1) specification with leverage and is described by (4). Expression (4) was proposed by Glosten et al. (1993), where $\mathbf{1}(\cdot)$ is an indicator function and τ_2 controls for asymmetric (leverage) effects; if $\tau_2 > 0$ the volatility increases much more after a negative shock than after a positive shock. To guarantee that the

variance process is always positive we impose the restrictions $\alpha_0 > 0$, $\alpha_1 + \tau_2 \geq 0$ and $\alpha_2 \geq 0$. The GARCH process is initialized with $\varepsilon_0 = 0$ and a known constant σ_0^2 .

As can be seen in (2), the conditional variance σ_t^2 is included as an additional regressor in the conditional mean equation of y_t^* , where the coefficient τ_1 is known in financial literature as the risk premium parameter. Provided that $\tau_1 \neq 0$, the in-mean effects imply that there is an additional source of serial correlation in the underlying continuous latent process $\{y_t^*\}$ (apart from the serial correlation in y_t^* s that is induced by the ARMA component) that stems from the serial correlation in the GARCH-type volatility process $\{\sigma_t^2\}$.

The model, given by (1)-(4), is an autoregressive moving average ordinal-response GARCH model with leverage and conditional heteroscedasticity in mean. We name this model the ARMA-OR-GARCHLM model¹.

Note that the ARMA-OR-GARCHLM model encompasses a large number of submodels. For instance, when $\tau_1 = 0$, the ARMA-OR-GARCHLM model reduces to the ARMA-OR-GARCH model with leverage (ARMA-OR-GARCHL). When $\tau_2 = 0$, the AROR-GARCHLM model is equivalent to the ARMA-OR-GARCH model with in-mean effects (ARMA-OR-GARCHM). For both $\tau_1 = 0$ and $\tau_2 = 0$, we obtain the ARMA-OR-GARCH model².

3 Posterior analysis

3.1 The MCMC algorithm for the ARMA-OR-GARCHLM model

Define:

$$\mathbf{x}_t = (y_{t-1}^*, \mathbf{z}_t')', \boldsymbol{\beta} = (\rho, \boldsymbol{\delta}')', \boldsymbol{\alpha} = [\alpha_0, \alpha_1, \alpha_2, \tau_2]'$$

¹Note that this model does not account for the fact that the trades are not necessarily equispaced. Putting a Δ operator in (2) could potentially alleviate the problem. However, this is not a proper way of dealing with this issue as there are other more suitable models for this purpose, such as the autoregressive conditional duration model (ACD) proposed by Engle and Russell (1998). Similar analysis holds for potential diurnal effects and microstructure noises. These issues will be examined in a future paper.

²In time series econometrics literature, ARMA-GARCH models in a Bayesian setting have been analyzed by Nakatsuma (2000). ARMA models in a non-Gaussian setting have been considered by Benjamin et al. (2003), Startz (2008), and Zheng et al., (2015), among others.

The likelihood for the latent model is:

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_1, \gamma, \{u_t\}; \mathbf{y}^*, \mathbf{X}) = \left\{ \prod_{t=1}^T (2\pi\sigma_t^2(\boldsymbol{\alpha}))^{-1/2} \right\} \times \exp \left\{ - \sum_{t=1}^T \frac{(y_t^* - \mathbf{x}_t' \boldsymbol{\beta} - \tau_1 \sigma_t^2(\boldsymbol{\alpha}) - u_t - \gamma u_{t-1})^2}{2\sigma_t^2(\boldsymbol{\alpha})} \right\}, \quad (5)$$

where $\mathbf{y}^* = [y_1^*, \dots, y_T^*]'$, \mathbf{X} is the $T \times k$ matrix of explanatory variables, $\sigma_t^2(\boldsymbol{\alpha}) = \alpha_0 + (\alpha_1 + \tau_2 \mathbf{1}(\varepsilon_{t-1} < 0))\varepsilon_{t-1}^2 + \alpha_2 \sigma_{t-1}^2(\boldsymbol{\alpha})$, and $\varepsilon_t = y_t^* - \mathbf{x}_t' \boldsymbol{\beta} - \tau_1 \sigma_t^2(\boldsymbol{\alpha}) - u_t - \gamma u_{t-1}$.

Given \mathbf{y}^* , the likelihood can be maximized using standard techniques and is amenable to Markov Chain Monte Carlo (MCMC) simulation through the posterior:

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_1, \gamma, \{u_t\} | \mathbf{y}^*, \mathbf{X}) \propto L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_1, \gamma, \{u_t\}; \mathbf{y}^*, \mathbf{X}) \cdot p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_1, \gamma), \quad (6)$$

where $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \tau_1, \gamma)$ denotes the joint prior of the model parameters.

MCMC in the complete data posterior is implemented using a Riemannian manifold technique due to Girolami and Calderhead (2011). This permits jointly updating all elements of the parameter vector $\boldsymbol{\theta} = [\boldsymbol{\alpha}', \boldsymbol{\beta}', \tau_1]'$, conditional on \mathbf{y}^* and $\{u_t\}$.

As \mathbf{y}^* is latent, we have to consider the augmented posterior:

$$p(\boldsymbol{\theta}, \boldsymbol{\zeta}, \gamma, \{u_t\}, \mathbf{y}^* | \mathbf{y}, \mathbf{X}) \propto \left\{ \prod_{t=1}^T \sigma_t^2(\boldsymbol{\alpha})^{-1/2} \right\} \exp \left\{ - \sum_{t=1}^T \frac{(y_t^* - \mathbf{x}_t' \boldsymbol{\beta} - \tau_1 \sigma_t^2(\boldsymbol{\alpha}) - u_t - \gamma u_{t-1})^2}{2\sigma_t^2(\boldsymbol{\alpha})} \right\} \times p(\boldsymbol{\theta}) \left\{ \prod_{t=1}^T \mathbf{1}(\zeta_{j-1} < y_t^* \leq \zeta_j) \right\} p(\boldsymbol{\zeta}) p(\gamma), \quad (7)$$

where $\mathbf{y} = [y_1, \dots, y_T]'$ and $p(\boldsymbol{\zeta})$ is the joint prior for the vector of cutoff points $\boldsymbol{\zeta} = [\zeta_2, \dots, \zeta_{J-2}]'$.

The MCMC algorithm for the ARMA-OR-GARCHLM model is described in the Online Appendix.

3.2 Average marginal effects

It is desirable in nonlinear models to understand how small changes in the regressors affect the mean responses. Since we are dealing with such models, interpreting directly the coefficients may be ambiguous. To this end, we calculate partial effects, which can be obtained as a by-product of our MCMC sampler. In this way, we can estimate the effect of a covariate change

on the probability of y being equal to an ordered value.

Let $x_{k,t}$ be a continuous regressor. The partial effect (pe) of $x_{k,t}$ on the probability of y_t being equal to j , after marginalizing out all the unknown parameters, is defined as

$$E(pe_{ktj}|\mathbf{y}) = \int \left(\frac{\partial P(y_t = j|\boldsymbol{\theta}, \mathbf{y}_{t-1}^*, \gamma, u_t, u_{t-1}, \zeta_{j-1}, \zeta_j)}{\partial x_{k,t}} \right) dp(\boldsymbol{\theta}, \mathbf{y}_{t-1}^*, \gamma, u_t, u_{t-1}, \zeta_{j-1}, \zeta_j|\mathbf{y}). \quad (8)$$

The derivative $\frac{\partial P(y_t=j|\boldsymbol{\theta}, \mathbf{y}_{t-1}^*, \gamma, u_t, u_{t-1}, \zeta_{j-1}, \zeta_j)}{\partial x_{k,t}}$ is calculated as

$$\begin{aligned} \frac{\partial P(y_t=j|\boldsymbol{\theta}, \mathbf{y}_{t-1}^*, \gamma, u_t, u_{t-1}, \zeta_{j-1}, \zeta_j)}{\partial x_{k,t}} &= \phi\left(\frac{\zeta_{j-1} - \mathbf{x}'_t \boldsymbol{\beta} - \tau_1 \sigma_t^2(\boldsymbol{\alpha}) - u_t - \gamma u_{t-1}}{\sigma_t^2(\boldsymbol{\alpha})}\right) \\ &\quad - \phi\left(\frac{\zeta_j - \mathbf{x}'_t \boldsymbol{\beta} - \tau_1 \sigma_t^2(\boldsymbol{\alpha}) - u_t - \gamma u_{t-1}}{\sigma_t^2(\boldsymbol{\alpha})}\right), \end{aligned} \quad (9)$$

where ϕ denotes the density of the standard normal distribution.

The average partial effect (APE) is the mean of the partial effects:

$$APE = \sum_{t=1}^T E(pe_{ktj}|\mathbf{y}). \quad (10)$$

Using draws from the MCMC chain, expression (10) is estimated by taking the average of (8) over all $t = 1, \dots, T$ and over all iterations.

3.3 Forecast evaluation

In order to evaluate the forecasting performance of the ARMA-OR-GARCHLM model we conduct a recursive out-of-sample forecasting exercise. In particular, we produce density and point r -step-ahead iterated forecasts with $r = 1, 5$ and 10 . This process works as follows. Given data up to time t , we implement the MCMC algorithm of section 3.1 for a large number of iterations (after a burn-in period) and obtain pairs of posterior draws (one pair at each iteration) for both the predictive mean and the predictive density, which are used as point and density forecasts, respectively. By taking the average over all posterior draws, we obtain estimates for the predictive mean and the predictive density for y_{t+r} . Next, we move one period ahead and repeat the same forecasting exercise and so forth.

In particular, let $\boldsymbol{\Theta}$ denote the vector of all parameters in the model and $\boldsymbol{\Theta}^{(m)}$ be an MCMC sample of $\boldsymbol{\Theta}$ at iteration $m = 1 \dots M$, after the burn-in period. The conditional predictive density

for the r -step ahead y_{t+r} given $\Theta^{(m)}$ and the data $\Omega_t = (\mathbf{y}_t, \mathbf{X}_t)$, where $\mathbf{y}_t = (y_1, \dots, y_t)$ and $\mathbf{X}_t = (\mathbf{x}_1, \dots, \mathbf{x}_t)$ is given by

$$p(y_{t+r}|\Omega_t, \Theta^{(m)}) = \Phi\left(\frac{\zeta_j^{(m)} - \mathbf{x}'_{t+r}\boldsymbol{\beta}^{(m)} - \tau_1^{(m)}\sigma_{t+r}^2(\boldsymbol{\alpha}^{(m)}) - u_{t+r}^{(m)} - \gamma^{(m)}u_{t+r-1}^{(m)}}{\sigma_{t+r}^2(\boldsymbol{\alpha}^{(m)})}\right) \\ - \Phi\left(\frac{\zeta_{j-1}^{(m)} - \mathbf{x}'_{t+r}\boldsymbol{\beta}^{(m)} - \tau_1^{(m)}\sigma_{t+r}^2(\boldsymbol{\alpha}^{(m)}) - u_{t+r}^{(m)} - \gamma^{(m)}u_{t+r-1}^{(m)}}{\sigma_{t+r}^2(\boldsymbol{\alpha}^{(m)})}\right). \quad (11)$$

By taking the average over the MCMC samples we can integrate out Θ to obtain the predictive density defined as

$$p(y_{t+r}|\Omega_t) = \frac{1}{M} \sum_{m=1}^M p(y_{t+r}|\Omega_t, \Theta^{(m)}). \quad (12)$$

Replacing y_{t+r} by the observed value y_{t+r}^o , we obtain the value $p(y_{t+r} = y_{t+r}^o|\Omega_t)$, which is called the predictive likelihood of y_{t+r} . Next, we move one period ahead and repeat the same forecasting exercise using Ω_{t+1} data. The log predictive score (LPS) of the model for the evaluation period $t = t_0 + 1, \dots, T - r$ is the sum of the log predictive likelihoods

$$LPS = \sum_{t=t_0}^{T-r} \log p(y_{t+r} = y_{t+r}^o|\Omega_t). \quad (13)$$

Higher values indicate better (out-of-sample) forecasting ability of the model.

A natural measure to evaluate the density forecast $p(y_{t+r}|\Omega_t)$ is the predictive likelihood $p(y_{t+r} = y_{t+r}^o|\Omega_t)$. We can also obtain the point forecast for y_{t+r} by producing an estimate for the predictive mean $E(y_{t+r}|\Omega_t)$. For the evaluation of point forecasts a usual metric is the root mean squared forecast error (RMSFE) defined as

$$RMSFE = \sqrt{\frac{\sum_{t=t_0}^{T-r} (y_{t+r}^o - E(y_{t+r}|\Omega_t))^2}{T - r - t_0 + 1}}. \quad (14)$$

Lower values of the RMSFE indicate better point forecasts.

In our forecasting comparison of the competing ordinal-response models, both the log predictive score values and the root mean squared forecast error values are reported. We produce r -step ahead point and density forecasts, with $r = 1, 5$ and 10 .

3.4 Tests of equal predictability

As a formal test of equal predictive ability of two competing forecasting results, we consider the Conditional Predictive Ability (CPA) test of Giacomini and White (2006). It is a pairwise test that assesses the significance of differences in the out-of-sample forecasting performance between two models, accounting also for parameter uncertainty.

For the one-step CPA test and given a sample of size T , let N be the size of the estimation window. Based on rolling samples of size N , we, then, obtain $n = T - N$ one-period ahead forecasts. This sequence of out-of-sample forecasts is evaluated by a sequence of n loss functions $\{L_{t+1}(y_{\mathcal{M}_i,t+1}, \hat{y}_{\mathcal{M}_i,t+1})\}_{t=N}^{T-1}$, where $y_{\mathcal{M}_i,t+1}$ is the observed value and $\hat{y}_{\mathcal{M}_i,t+1}$ is the predicted value of model \mathcal{M}_i . The difference in loss of model \mathcal{M}_i relative to a baseline model \mathcal{M}_o is defined as $d_{\mathcal{M}_i,t+1} = L_{t+1}(y_{\mathcal{M}_i,t+1}, \hat{y}_{\mathcal{M}_i,t+1}) - L_{t+1}(y_{\mathcal{M}_o,t+1}, \hat{y}_{\mathcal{M}_o,t+1})$. Under the null hypothesis of equal predictive ability $H_o : E(d_{\mathcal{M}_i,t+1}|\mathcal{F}) = 0$, where \mathcal{F} is some information set, the CPA test statistic is a Wald-type test statistic of the form

$$W_{\mathcal{M}_i} = n \left(n^{-1} \sum_{t=N}^{T-1} h_{\mathcal{M}_i,t} d_{\mathcal{M}_i,t+1} \right)' \hat{\Omega}_n^{-1} \left(n^{-1} \sum_{t=N}^{T-1} h_{\mathcal{M}_i,t} d_{\mathcal{M}_i,t+1} \right), \quad (15)$$

where $h_{\mathcal{M}_i,t}$ is a test function and $\hat{\Omega}_n^{-1}$ is a consistent estimator of the covariance of $h_{\mathcal{M}_i,t} d_{\mathcal{M}_i,t+1}$. A similar analysis holds for a multistep CPA test; see equation (6) in the Giacomini and White (2006) paper.

To test significant differences of multiple competing models, we use the Model Confidence Set (MCS) approach of Hansen et al., (2011). The MCS approach does not require the specification of a benchmark model. Starting from the full set of models, \mathcal{M} , this method sequentially eliminates the worst performing models, until it ends up with a set of models $\mathcal{M}^* \subset \mathcal{M}$, where the null hypothesis of equal predictive ability is not rejected at a certain confidence level α . The resulting set \mathcal{M}^* is the MCS. Model i belongs to \mathcal{M}^* only if the estimated p -value is greater than or equal to α , namely $\hat{p}_{\mathcal{M}_i} \geq \alpha$.

Given the forecast horizon r , the MCS method repeatedly tests the following null hypothesis of equal forecasting ability

$$H_{o,\mathcal{M}} : E \left(L_{t+r}(y_{\mathcal{M}_i,t+r}, \hat{y}_{\mathcal{M}_i,t+r}) - L_{t+r}(y_{\mathcal{M}_j,t+r}, \hat{y}_{\mathcal{M}_j,t+r}) \right) = 0 \text{ for all } i, j \in \mathcal{M}, \quad (16)$$

where L is a loss function. The alternative hypothesis is that the models have different forecast accuracy. To test the above null hypothesis the following statistic is constructed

$$\mathbf{T} = \max_{i,j \in \mathcal{M}} t_{i,j}, \quad (17)$$

where $t_{i,j}$ is the Diebold-Mariano (1995) test statistic. If we fail to reject the null hypothesis, all models in \mathcal{M} are equally accurate. If we reject the null hypothesis, we eliminate the worst model that has the highest average loss and repeat the test in (16) with a different set of models, until no model is eliminated. Following Hansen et al. (2011), we calculate the p -values using block bootstrap methods as the distribution under the null is not identified.

3.5 Optimal pooling

For the comparison of multiple forecasting models, we adopt the approach of model pooling by Geweke and Amisano (2011). This approach assumes that none of the competing models corresponds to the true data generating process, but instead considers a linear prediction pool based on the predictive likelihood (log score function) from a set of competing models.

Given a set of models $\{\mathcal{M}_i\}_{i=1}^M$ and a set of predictive densities $\{p(y_t|y_1, \dots, y_{t-1}, \mathcal{M}_i)\}_{i=1}^M$, we consider the following form of combined predictive densities

$$\sum_{i=1}^M w_i p(y_t|y_1, \dots, y_{t-1}, \mathcal{M}_i), \text{ where } \sum_{i=1}^M w_i = 1, w_i \geq 0, i = 1, \dots, M. \quad (18)$$

The optimal weight vector \mathbf{w}^* is chosen to maximise the log pooled predictive score function; that is,

$$\operatorname{argmax}_{w_i, i=1, \dots, M} \sum_{t=\tau_1}^{\tau_2} \log \left(\sum_{i=1}^M w_i p(y_t|y_1, \dots, y_{t-1}, \mathcal{M}_i) \right), \quad (19)$$

where the predictive density is evaluated at the realised value y_t .

Conditional on the data up to time $t-1$, i.e., y_1, \dots, y_{t-1} , we obtain a large number of posterior draws for the parameters (by applying the MCMC algorithm of section 3.1 for a large number of iterations), which are then used to evaluate the predictive likelihood $p(y_t = y_t^o|y_1, \dots, y_{t-1}, \mathcal{M}_i)$. Based on the entire history of predictive likelihood values we can estimate the weights in expression (19). For optimization, we have used a standard nonlinear solver (Nash, 1984).

4 Empirical examples

4.1 Data sets and modelling strategies

To illustrate our methods we use ten transactions data sets of different liquidity and tick size that were obtained from the Trades and Quotes (TAQ) database of the New York Stock Exchange (NYSE); see Table 1. For each data set, we consider several competing ordinal-response time series models for comparison purposes. In particular, we employ the main model (ARMA-OR-GARCHLM) without covariates, which is compared against its following nested versions:

- ARMA-OR-GARCH model
- ARMA-OR-GARCH model that controls only for leverage (ARMA-OR-GARCHL)
- ARMA-OR-GARCH model that controls only for in mean effects (ARMA-OR-GARCHM)
- AR-OR-GARCH model
- AR-OR-GARCH model that controls only for leverage (AR-OR-GARCHL)
- AR-OR-GARCH model that controls only for in mean effects (AR-OR-GARCHM)
- AR-OR-GARCH model that controls for leverage and in mean effects (AR-OR-GARCHLM)
- the AR-OR model with constant variance (no conditional heteroscedasticity)

The above list is augmented to include also the continuous counterpart of the main model, the ARMA-GARCHLM model.

To make the exposition of the forecasting results more readable, we report, throughout the paper, the ratio of the LPS of the baseline model to that of a given model, with the baseline model being the AR-OR model. Values smaller than one indicate better forecasting ability than the baseline model. Similarly, for ease of exposition, the root mean squared forecast error (RMSFEs) of the baseline model is divided by that of a given model. Values greater than one indicate better forecasting power than the AR-OR model.

Regarding the implementation of the optimal prediction pooling of Geweke and Amisano (2011), we have used the last 500 data points. As far as the CPA tests are concerned, they are all

two-sided. We produce sequences of r -step ahead forecasts for $r = 1, 5, 10$ over rolling windows of 100 observations and report the p -values (in the Online Appendix). The test function for the CPA tests is $h_{\mathcal{M}_i,t} = (1, d_{\mathcal{M}_i,t})'$ for each model \mathcal{M}_i and t , while the squared error loss is used as a loss function. The same loss function was used for the MCS approach, with the block bootstrap length being equal to ten. The MCS p -values are also given in the Online Appendix.

All our estimation results (posterior means and posterior standard deviations) are based on 150000 iterations of the samplers, using a burn-in of 50000 cycles. Also, throughout our empirical analysis we set the ζ_3 cutpoint equal to 1. Below we analyse the estimation and forecasting results for the first data set (Caterpillar stock) of Table 1³.

4.2 Data set I: Caterpillar stock

We first consider the tick-by-tick trading data of Caterpillar stock on January 4, 2010. We focus on the intraday price changes of Caterpillar stock between 9:30 AM and 4:00 PM Eastern time. During this time period there were recorded 37716 transactions and 37715 price changes which are displayed graphically in Figure 1. These price differences tend to be concentrated (clustered) on multiples of 1 cent, as their histogram shows in Figure 2, with 98.7% of these differences lying between -2 cents and $+2$ cents. So, the price change of Caterpillar stock in consecutive trades can be considered as a discrete-valued variable y_t . Furthermore, since the intraday price changes of less than 2 cents and more than 2 cents are few, we group them as price changes of < -2 and > 2 , respectively.

Taking these considerations into account, the dependent variable y_t in our analysis takes on 6 categories of price changes, where a value of “1” is assigned to price changes that fall more than two cents (< -2), a value of “2” is assigned to price changes that fall in the interval $[-2, -1)$ and so on up to price changes that are above 2 cents (> 2), which are assigned a value of “6”. Table 2 displays the intervals of price differences and the corresponding response categories along with their corresponding observed frequencies for the Caterpillar stock data.

The sample autocorrelation functions (ACF) for the price differences (Figure 3) and their squared values (Figure 4) seem a good motivation for the inclusion of the moving average component (likely due to bid ask bounce) and for accounting for time-varying second moments

³Due to the large number of models and data sets in question, we do not report the estimation results for the rest of the empirical data sets. These results are available upon request.

(conditional heteroscedasticity). Also, the negative skewness of the unconditional distribution of the stock price changes (the histogram has skewness equal to -0.1838) prompts for accounting for in-mean effects as the GARCH-M model, generally speaking, implies postulating a marginal distribution for the returns, which is skewed.

4.2.1 Estimation results

The ρ and γ parameters

Table 3 presents the results. The estimated autoregressive parameters ρ are significant and positive across all models. The significance of the AR(1) parameters signals the presence of time dependence in the observed responses of the data set in question, through the dependence of the latent variable y_t^* at time t on its level at $t-1$. The positive sign implies positive dynamic persistence in the discrete price changes.

The magnitude of ρ appears to be largest in the AR-OR-GARCHLM and ARMA-OR-GARCHLM models that control for both leverage and in mean effects (0.414 and 0.410, respectively). The standard models, which have been encountered in the relevant empirical literature (AR-OR and AR-OR-GARCH), downgrade the strength of the dependence of the current transaction price on the previous transaction price, as they produce the smallest values of ρ (0.188 and 0.217, respectively).

It is worth noting that ρ was also found to have a positive sign in the paper by Müller and Czado (2005). Müller and Czado (2005) estimated an autoregressive ordinal-response (AR-OR) model and the coefficient on y_{t-1}^* was equal to 0.1362. In our empirical data, the respective estimated value of ρ for our AR-OR model is similar (0.188); see Table 3.

However, since we are dealing with discrete-response probability models, we also calculated the average partial effects (APEs) for the regressors y_{t-1}^* and σ_t^2 . The results are presented in Table 4. The APE of y_{t-1}^* on the expected price change is positive and significant across all models except for the ARMA-OR-GARCHM model. Therefore, a big (small) price change in the previous transaction increases the probability of observing another big (small) price change in the next transaction.

For the AR-OR-GARCHLM and ARMA-OR-GARCHLM models, the (significant) APEs for y_{t-1}^* are equal to 0.388 and smaller in magnitude than the (significant) APEs obtained

from the ordinal-response models with only leverage or only in mean effects. In any case, the statistically significant effects of the lagged dependent variables on the expected price changes are lower in the standard ordinal-response models (AR-OR and AR-OR-GARCH) than in the rest of the models, as Table 4 shows. Put differently, when additional microstructure effects of the series of prices changes (moving average, leverage or in mean effects) are disregarded, there seems to be a downward bias in the estimated value of the APE for y_{t-1}^* (dynamic effects).

The posterior mean of γ was found to be between 0.02 and 0.182 in the ordinal-response models that incorporate the moving average component (Table 3). The serial error correlation, although not strong in magnitude, verifies that the data behaviour may suffer from the bid-ask bounce (as the data plots also showed). This is a finding that had been ignored by previous empirical studies. Also, the estimated value of γ is the highest for the main model (ARMA-OR-GARCHLM) and the lowest for the ARMA-OR-GARCH model.

The GARCH and in-mean parameters

There is a positive feedback from the conditional variance to the conditional mean of the latent regression for y_t^* (volatility feedback effect), as the in-mean parameters τ_1 are all positive in Table 3. This is also verified by the estimated APEs for the regressor σ_t^2 in Table 4, where we additionally observe that they are larger in the ordinal-response models with a moving average component. In the models that control for leverage effects, the posterior means of τ_2 parameters are positive and significant (Table 3), implying that a negative shock that takes place at time t will increase the conditional variance at time $t + 1$.

We conclude that both the in-mean and leverage effects affect the underlying continuous latent process of y^* that regulates the discretisation of the observed responses, and therefore, they are important factors in analyzing transaction stock prices. This is another empirical finding in the literature on discrete-response financial time series models.

The GARCH parameters $(\alpha_0, \alpha_1, \alpha_2)$ that govern the evolution of the conditional variance process are also significant across the models of Table 3. Furthermore, there is some persistence in the time-varying conditional variances; the posterior means for α_1 were found to be between 0.271 (for the AR-OR-GARCHL model) and 0.615 (for the ARMA-OR-GARCH). Also, most of the ARMA-OR-GARCH-type models yield larger volatility persistence than the non-ARMA-OR-GARCH-type models.

4.2.2 Forecasting results

In this section we examine how the models of Table 3 behave in terms of forecasting performance. In our forecast evaluation exercise, we computed point and density forecasts for 1-transaction, 5-transaction and 10-transaction ahead ($r = 1, 5$ and 10). To this end, we calculated log predictive scores (LSPs) and root mean squared forecast errors (RMSFEs). The results are presented in Tables 5 and 6, respectively.

Table 5 presents the results for the density forecasts. For short-horizon forecasts ($r = 1$), the AR-OR-GARCH model performs better than the rest of the models. The AR-OR-GARCHLM, which is the second best forecasting model, produces better density forecasts than its nested versions AR-OR-GARCHM (third best) and AR-OR-GARCHL (fourth best).

Notice also that the AR-OR-GARCH-type models outperform the ARMA-OR-GARCH-type models. Therefore, the addition of the moving average component to the AR-OR-GARCH variants does not contribute to their forecasting performance at short horizons.

For medium horizon forecasts ($r = 5$), the best forecasting performance is attained by the AR-OR-GARCHLM, followed by the AR-OR-GARCH model. Both the ARMA-OR-GARCHL and ARMA-OR-GARCHM models occupy the next two positions and they do better than the corresponding AR-OR-GARCHL and AR-OR-GARCHM models. That was not the case for $r = 1$. Hence, the role of the moving average term becomes more important in improving medium horizon density forecasts.

For longer horizons ($r = 10$), the AR-OR-GARCHLM model still remains the winner (as was also the case for $r = 5$), followed by its nested versions ARMA-OR-GARCHL (second best) and ARMA-OR-GARCHM (third best). We see that as the horizon of forecasts becomes bigger, some of the previously less dominant ARMA-OR-GARCH variants elevate their forecasting power, outperforming previously dominant AR-OR-GARCH variants. For instance, the second best and third best models at long horizons displaced the AR-OR-GARCH model, which was dominant at short and medium horizons.

The point forecast results are given in Table 6. For short horizons ($r = 1$), the most preferred model is the ARMA-OR-GARCH. Also, the ARMA-OR-GARCHM and ARMA-OR-GARCHL models occupy the second best and third best positions, respectively. In addition, we observe a superiority of almost all the ARMA-OR-GARCH variants (except for the main

model) over all AR-OR-GARCH variants. That was not the case for the short horizon density forecasts. Among the AR-OR-GARCH variants, the AR-OR-GARCH model produces the best point forecast. The same story is repeated for $r = 5$.

For $r = 10$, the AR-OR-GARCH model loses its dominance among the AR-OR-GARCH-type models. Also, all the models now with the moving average term (including the main specification) do better than the rest of the models, with the ARMA-OR-GARCHL model being the best.

Based on the LPS and the RMSFE values, the continuous ARMA-GARCHLM model has the worst forecasting performance across all horizons.

According to the MCS results (Table 6), the superior set of models contains the ARMA-OR-GARCH variants, indicating that their forecasting accuracies are not significantly different. These variants survive the elimination process for every value of r , except for the ARMA-OR-GARCH model, which is excluded from the MCS for $r = 5$ and $r = 10$. All the other models are excluded from the MCS at the 95% confidence level for all horizons; the forecasting ability of the ARMA-OR-GARCH variants is superior to that of the AR-OR-GARCH variants and of the traditional models (AR-OR-GARCH, AR-OR and ARMA-GARCHLM). These results are also verified by the produced RMSFE values.

We also considered weighted linear combinations of prediction models (linear pools), evaluated using the log predictive scoring rule (Geweke and Amisano, 2011). In particular, optimal weights can be computed in each transaction to form a prediction pool for the r -step ahead transaction ($r = 1, 5, 10$). Figure 5 shows the evolution of these weights.

The main specification, the ARMA-OR-GARCHLM, receives by far the largest weight over the entire out-of-sample forecasting period (Figure 5(a)). This weight increases over time at the expense of the ARMA-OR-GARCHL model (Figure 5(b)) and the ARMA-OR-GARCHM model (Figure 5(c)), both of which receive a negligible weight. Figure 5(d) displays the sum of weights “on all other models”, which is even smaller than the weights shown in Figures 5(b) and 5(c).

The pool of models optimized over all the out-of-sample predictive densities for all the 500 data points (see expression (19)) is given in the second row of Table 10. For example, for $r = 1$, the optimal prediction weight for the ARMA-OR-GARCHLM model is 0.799, for the

ARMA-OR-GARCHL model is 0.140, and for the ARMA-OR-GARCHM model is 0.061. The sum of these weights is equal to one, with the ARMA-OR-GARCHLM receiving most of the weight. All the other models are excluded from the optimal pool as they were assigned zero weight. For $r = 5$ and 10, the optimal pools also include ARMA-OR-GARCH-type models only, where the ARMA-OR-GARCHLM dominates again.

As a next step in our analysis we examined graphically how well the models in question can forecast the price changes in consecutive trades ($r = 1$) for four different time intervals of 5000 observations each. To this end, we produced Figure 6 that displays the evolution of the real cumulative price changes (blue line) against that of the predicted cumulative price changes obtained from various forecasting models.

Among the ordinal-response models, the ARMA-OR-GARCHLM model (red line), had the best fit to the real path of cumulative price changes. We, then, compared the predicted path generated by this model against that generated by the traditional models ARMA-GARCHLM and AR-OR-GARCH. As can be seen from the first three plots (Figures 6(a)-6(c)), the traditional models do worse than the ARMA-OR-GARCHLM model, failing to follow closely the true path of values for the three time intervals in question and there are cases where the divergence is significant.

In Figure 6(d) we plotted the time series of the predicted cumulative price changes obtained from the optimal pool of models (ARMA-OR-GARCHLM, ARMA-OR-GARCHL, ARMA-OR-GARCHM) for the last 5000 observations of our data set. We observe that when these three models are combined together, the resulting series is able to replicate/capture very well the movement of the true one.

As a last issue we evaluated the proposed models, in terms of profitability. It is known that traders want to buy low and sell high so as to achieve high returns with low risk. Hence, it is worth examining if our proposed models are realistic models for trading purposes. As such, we simulated a simple trading rule according to which we sell when we predict a decrease in price and buy otherwise. The decrease or increase has to be “significant” in the sense that the price difference changes category. The predictions are extracted from the ARMA-GARCHLM model and the optimal combination of models. In either case, we produce a forecast for $r = 1, 5, 10$ based on the relevant predictive density.

Then, the performance of the trading strategy for each case (ARMA-GARCHLM or optimal pool) is evaluated by the average excess return it can generate. We found that the out-of-sample average excess returns earned by the trading rule are greater for the optimal pool of models than for the ARMA-GARCHLM model by 5.30%, 4.17% and 3.22% for $r = 1, 5$ and 10, respectively; see second row of Table 11.

4.3 General findings for the rest of the data sets

The same analysis was repeated for the rest of the data sets of Table 1. Due to the large number of models and data sets in question, we provide in Tables 7-11 an overall summary of the forecasting results that were obtained from the data sets I-X. In the Online Appendix we report the full forecasting results (relevant tables and figures) for data set II only (as did for data set I). For the data sets III-X, these results are available upon request.

Tables 7 and 8 report the best forecasting model for each data set and for each horizon ($r = 1, 5, 10$), based on the LPS and RMSFE values, respectively. In terms of density forecasts, the ARMA-OR-GARCH variants occupy the first best position, with the ARMA-OR-GARCHLM being the dominant model in most of the cases. The traditional benchmark AR-OR-GARCH produced the best density forecast result only for data set I and only for the short-term horizon. Also, the AR-OR-GARCH-type models produced worse density forecasts than the ARMA-OR-GARCH-type models.

Similar analysis holds for the point forecast results, as far as the dominance of the ARMA-OR-GARCH specifications is concerned. The continuous-response model ARMA-GARCHLM beats the rest of the models but only for data set X and only for $r = 1$. As was also the case for the density forecasts, the AR-OR-GARCH variants fail to defeat their counterparts that control for an MA term.

Based on the MSC results (Table 9), the set of models with equal predictability contains, in any case, the ARMA-OR-GARCH specifications, which again dominate. Only for data set II, the AR-OR-GARCH variants survive the elimination process, where the ARMA-GARCHLM model is also included in the superior set of models for every r . The continuous model belongs to the superior set of models and for data X as well.

In Table 10, we report the optimal prediction pool optimized over all the last 500 obser-

variations. In doing so for each horizon, we obtain the optimal weights for the models that are included in the pool. These weights add up to one. As can be seen from Table 10, most of the weight is allocated to the ARMA-OR-GARCHLM model. The rest of the ARMA-OR-GARCH variants receive the remaining weight. The various AR-OR-GARCH models are excluded from the optimal pool, as is also the case for the continuous ARMA-GARCHLM model, except for the last data set.

Using also the trading rule that we developed in the previous section, we were able to calculate the profitability obtained from the optimal pool of models against that obtained from the ARMA-GARCHLM model. From Table 11, we observe that the average excess returns generated from the optimal pool of models are always greater than those generated from the ARMA-GARCHLM model by 2.16% up to 9.32%.

5 Conclusions

In this paper we proposed a novel autoregressive (AR) ordinal-response (OR) time series model with generalized autoregressive conditional heteroscedasticity (GARCH). In an attempt to account for potential microstructure effects that may be present in intraday transaction stock prices, the proposed model specification was equipped with leverage effects, a moving average (MA) component and conditional heteroscedasticity in mean. For model estimation, a Markov Chain Monte Carlo algorithm was developed, while a substantive forecasting exercise was conducted. We found that the proposed ordinal-response models produce forecast gains over the traditional benchmarks, with the ARMA-OR-GARCH-type models being dominant.

References

- Ahmed, J. (2016). A conditionally heteroscedastic binary choice model for macro-financial time series. *Journal of Statistical Computation and Simulation*, 86, 2007-2035.
- Awartani, B.M.A., & Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries. *International Journal of Forecasting*, 21, 167–183.
- Barra, I., & Koopman, S.J. (2017). Bayesian dynamic modeling of high-frequency integer price changes. *Working paper*
- Benjamin, M.A., Rigby, R.A., & Stasinopoulos, D.M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98, 214–223.
- Bianco, S., & Renò, R. (2006). Dynamics of intraday serial correlation in the Italian futures market. *Journal of Futures Markets*, 26, 61–84.
- Bianco, S., & Renò, R. (2009). Unexpected volatility and intraday serial correlation. *Quantitative Finance*, 9, 465-475.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- Calzolari, G., & Fiorentini, G. (1998). A tobit model with GARCH errors. *Econometric Reviews*, 17, 85 - 104.
- Campbell, J.Y., Lo, A.W., & MacKinlay, A.C. (1997). *The Econometrics of Financial Markets*. Princeton (NJ): Princeton University Press.
- Chakravarty, S., Wood, R.A., & Ness, R.A.V. (2004). Decimals and liquidity: A study of the NYSE. *Journal of Financial Research*, 27, 75-94.
- Chan, J., & Grant, A. (2016). Modeling energy price dynamics: GARCH versus Stochastic Volatility. *Energy Economics*, 54, 182–189.
- Chordia, T., & Subrahmanyam, A. (1995). Market making, the tick size and payment for-order-flow: Theory and evidence. *Journal of Business*, 68, 543–576.
- Cordella, T., & Foucault, T. (1999). Minimum price variations, time priority and quote

dynamics. *Journal of Financial Intermediation*, 8, 141–173.

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, 30, 368–380.

Christensen, B.J., Nielsen, M., & Zhu, J. (2010). Long memory in stock market volatility and the volatility-in-mean effect: The FIEGARCH-M Model. *Journal of Empirical Finance*, 17, 460–470.

Diebold, F., & Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–263.

Dimitrakopoulos, S., & Dey, K.D. (2017). Discrete-response state space models with conditional heteroscedasticity: An application to forecasting the federal funds rate target. *Economics Letters*, 154, 20–23.

Dueker, M. (1999). Conditional heteroscedasticity in qualitative response models of time series: A Gibbs-sampling approach to the bank prime rate. *Journal of Business and Economic Statistics*, 17, 466–472.

Engle, R.F., & Russell, J.R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66, 1127 – 1162.

Engle, R.F., & Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48, 1749–1778.

Geweke, J., & Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164, 130–141.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74, 1545–1578.

Girolami, M., & Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 123–214.

Glosten, L.R., Jagannathan, R., & Runkle, D.E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of*

Finance, 48, 1779-1801.

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

Hausman, J.A., Lo, A.W., & MacKinlay, A.C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 31, 319–379.

Koopman, S.J., Lit, R., & Lucas, A. (2017). Intraday stochastic volatility in discrete price changes: the dynamic Skellam model. *Journal of the American Statistical Association*, 112, 1490-1503.

Müller, G., & Czado, C. (2005). An autoregressive ordered probit model with application to high-frequency financial data. *Journal of Computational and Graphical Statistics*, 14, 320-338.

Müller, G.J., & Czado, C. (2009). Stochastic volatility models for ordinal valued time series with application to finance. *Statistical Modelling*, 9, 69–95.

Münnix, M.C., Schäfer, R., & Guhr, T. (2010). Impact of the tick-size on financial returns and correlations. *Physica A*, 389, 4828-4843.

Nakatsuma, T. (2000). Bayesian analysis of ARMA–GARCH models: A Markov chain sampling approach. *Journal of Econometrics*, 95, 57–69.

Nash, S.G. (1984). Newton–type minimization via the Lanczos method. *SIAM Journal on Numerical Analysis*, 21, 770–788.

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59, 347–370.

O’Hara, M., Saar, G., & Zhong, Z. (2014). Relative tick size and the trading environment. *Working paper*.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, 4, 1127– 1139.

Ronen, T., & Weaver, D.G. (2001). “teenies” anyone?. *Journal of Financial Markets*, 4, 231–260.

Startz, R. (2008). Binomial autoregressive moving average models with an application to

US recessions. *Journal of Business and Economics Statistics* 26, 1-8.

Yang, J.W., & Parwada, J. (2012). Predicting stock price movements: An ordered probit analysis on the Australian securities exchange. *Quantitative Finance*, 12, 791-804.

Ye, M., & Yao, C. (2014). Tick size Constrains, market structure, and liquidity. *Working paper*.

Zheng, T., Xiao, H., & Chen, R. (2015). Generalized ARMA models with martingale difference errors. *Journal of Econometrics*, 189, 492–506.

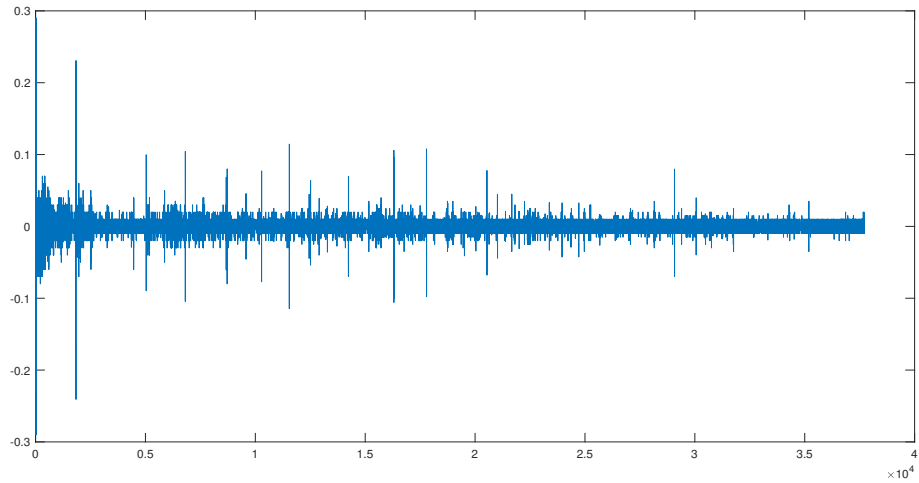


Figure 1: Empirical results. Time plot of price changes (measured in cents) in consecutive transactions for Caterpillar stock traded on January 4, 2010.

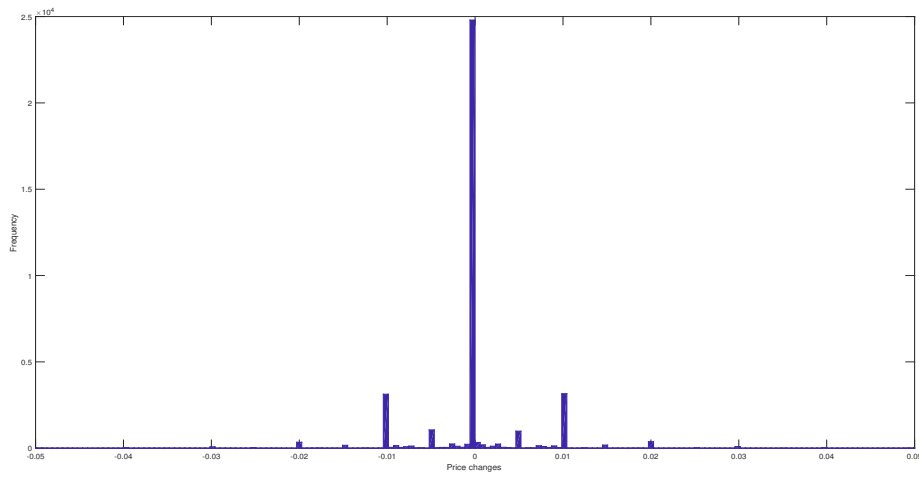


Figure 2: Empirical results. Histogram of price changes for Caterpillar stock traded on January 4, 2010.

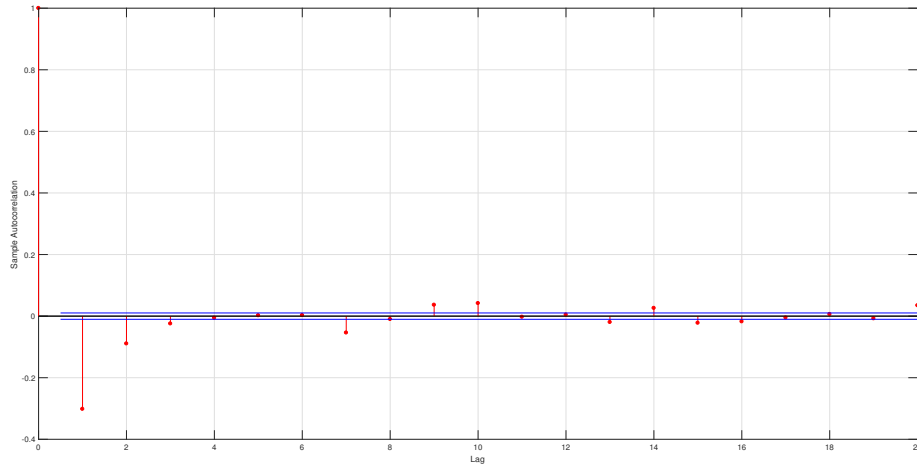


Figure 3: Empirical results. Sample autocorrelation function of the price changes for Caterpillar stock traded on January 4, 2010. The two blue horizontal lines represent the 95% confidence bounds.

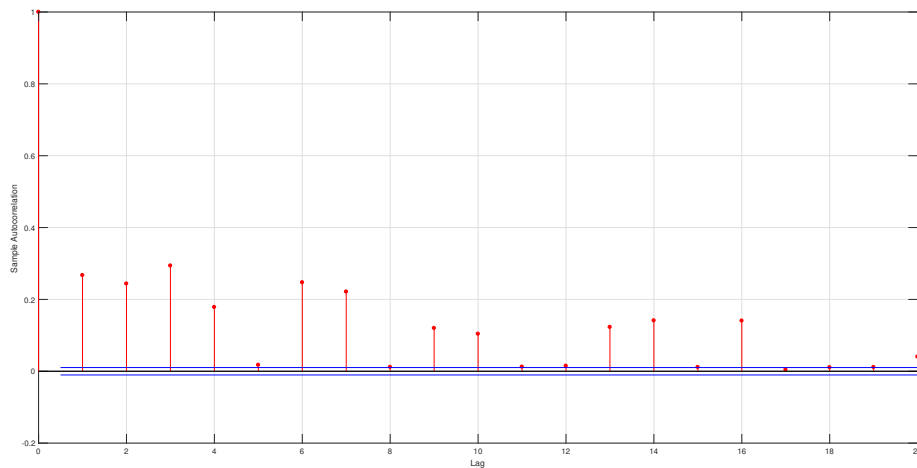
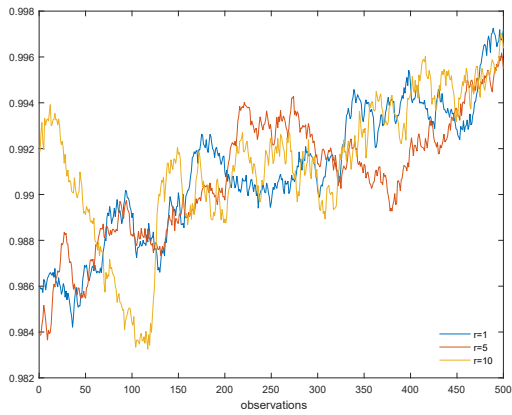
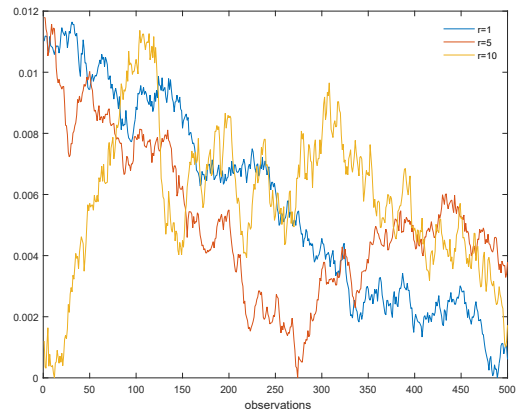


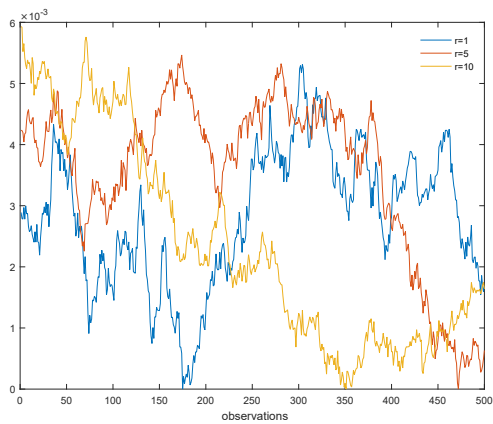
Figure 4: Empirical results. Sample autocorrelation function of the squared price changes for Caterpillar stock traded on January 4, 2010. The two blue horizontal lines represent the 95% confidence bounds.



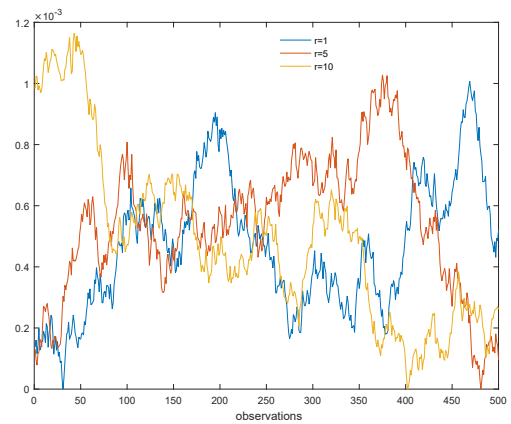
(a) Weight on model ARMA-OR-GARCHLM.



(b) Weight on model ARMA-OR-GARCHL.

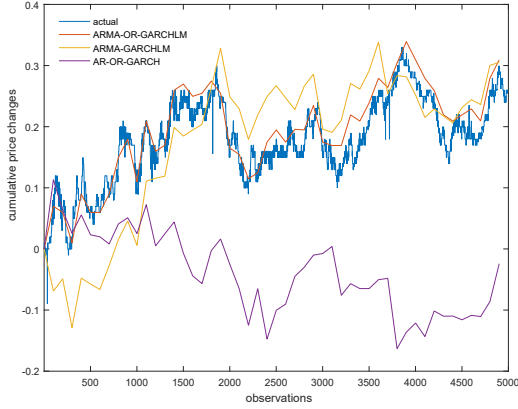


(c) Weight on model ARMA-OR-GARCHM.

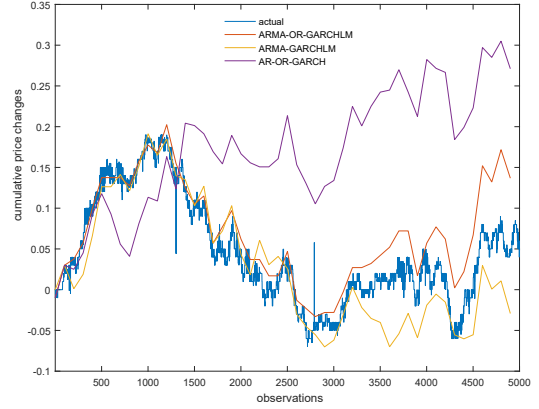


(d) Sum of weights on all other models.

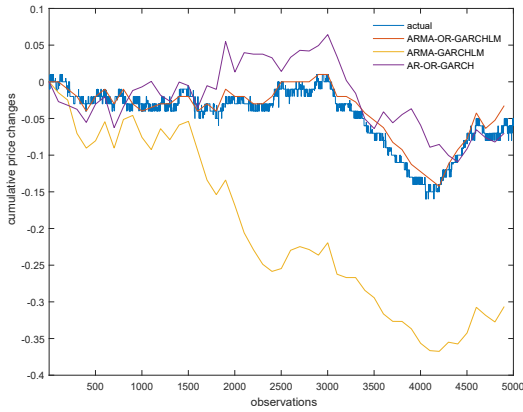
Figure 5: Empirical results. Evolution of model weights in the nine-model pool of Caterpillar stock $r = 1, 5$ and 10-step ahead predictive densities.



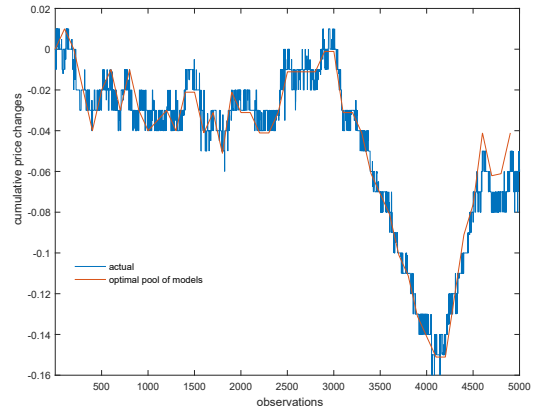
(a) Actual and predicted time series of cumulative price changes for particular models ($t = 5000, \dots, 10000$).



(b) Actual and predicted time series of cumulative price changes for particular models ($t = 15000, \dots, 20000$).



(c) Actual and predicted time series of cumulative price changes for particular models ($t = 20000, \dots, 25000$).



(d) Actual and predicted time series of cumulative price changes for the optimal pool of models ($t = 32715, \dots, 37715$). The optimal pool of models consists of the models reported in Figures 5(a)-5(c); see also Table 10.

Figure 6: Empirical results: Actual and predicted cumulative price changes for $r = 1$ -step ahead transactions (Caterpillar stock).

Table 1: Data sets.

Data set	Description	Total price changes
I	Caterpillar stock traded on January 4, 2010.	37715
II	Jonson and Jonson (JNJ) stock traded on October 5, 2010.	47805
III	General Electric (GE) stock traded on December 2, 2003.	13048
IV	Starbucks stock traded on July 29, 2011.	74396
V	Boeing stock traded on December 1, 2008.	43893
VI	Alcoa stock traded on June 8, 2010.	83016
VII	IBM stock traded on June 9, 1999.	4149
VIII	IBM stock traded on January 13, 2000.	5552
IX	Microsoft stock traded on December 16, 1997.	6422
X	Microsoft stock traded on May 16, 2000.	23775

Notes: The transaction data sets were obtained from the Trades and Quotes (TAQ) database of the New York Stock Exchange (NYSE). We focus on normal trading hours from 9:30 AM to 4:00 PM Eastern time. The tick size on NYSE was $\frac{1}{8}$ before June 24, 1997 and $\frac{1}{16}$ before January 29, 2001. Since January 29, 2001 all NYSE stocks are traded in decimals and the price changes in consecutive transactions tend to occur in multiples of 1 cent.

Table 2: Empirical application. Response categories, intervals of price changes and observed frequencies for Caterpillar stock data on January 4, 2010.

Price change intervals (cents)	< -2	$[-2, -1)$	$[-1, 0]$	$(0, 1]$	$(1, 2]$	> 2
Responses (y_t)	1	2	3	4	5	6
Frequencies	228	638	30239	5672	691	247

Table 3: Empirical results. Competing autoregressive ordinal-response models (Caterpillar stock).

Models/Parameters	ρ	γ	τ_1	τ_2	α_0	α_1	α_2
ARMA-OR-GARCHLM	0.410* (0.032)	0.182* (0.016)	0.017* (0.0032)	0.032* (0.013)	0.053* (0.012)	0.303* (0.025)	0.572* (0.044)
ARMA-OR-GARCHL	0.372* (0.025)	0.091* (0.029)		0.044* (0.021)	0.049* (0.017)	0.382* (0.017)	0.488* (0.032)
ARMA-OR-GARCHM	0.313* (0.019)	0.177* (0.023)	0.025* (0.0093)		0.061* (0.014)	0.399* (0.021)	0.588* (0.037)
ARMA-OR-GARCH	0.332* (0.017)	0.022* (0.0045)			0.059* (0.0083)	0.615* (0.022)	0.287* (0.015)
AR-OR-GARCHLM	0.414* (0.032)		0.012* (0.0038)	0.035* (0.007)	0.061* (0.0044)	0.313* (0.042)	0.544* (0.017)
AR-OR-GARCHL	0.372* (0.041)			0.027* (0.0013)	0.057* (0.0032)	0.271* (0.036)	0.655* (0.015)
AR-OR-GARCHM	0.288* (0.018)		0.017* (0.0044)		0.044* (0.0041)	0.301* (0.028)	0.598* (0.013)
AR-OR-GARCH	0.217* (0.011)				0.058* (0.012)	0.288* (0.032)	0.582* (0.019)
AR-OR	0.188* (0.015)						

Reported posterior means. Standard errors in parenthesis.* denotes significant based on the 95% highest posterior density intervals.

Table 4: Empirical results. Marginal effects for competing autoregressive ordinal-response models (Caterpillar stock).

Models/Regressors	y_{t-1}^*	σ_t^2
ARMA-OR-GARCHLM	0.388* (0.071)	0.103* (0.044)
ARMA-OR-GARCHL	0.414* (0.082)	
ARMA-OR-GARCHM	0.055 (0.077)	0.138* (0.054)
ARMA-OR-GARCH	0.455* (0.071)	
AR-OR-GARCHLM	0.388* (0.081)	0.027* (0.0092)
AR-OR-GARCHL	0.431* (0.061)	
AR-OR-GARCHM	0.444* (0.075)	0.035* (0.012)
AR-OR-GARCH	0.303* (0.061)	
AR-OR	0.289* (0.055)	

Reported posterior means. Standard errors in parenthesis.* denotes significant based on the 95% highest posterior density intervals.

Table 5: Empirical results. Relative log predictive scores (LPS) for competing autoregressive ordinal-response models (Caterpillar stock).

Models	$r = 1$	$r = 5$	$r = 10$
ARMA-OR-GARCHLM	0.233	0.245	0.239
ARMA-OR-GARCHL	0.144	0.132	0.127
ARMA-OR-GARCHM	0.132	0.138	0.128
ARMA-OR-GARCH	0.166	0.212	0.27
AR-OR-GARCHLM	0.083	0.092	0.114
AR-OR-GARCHL	0.105	0.177	0.185
AR-OR-GARCHM	0.097	0.182	0.130
AR-OR-GARCH	0.081	0.103	0.155
AR-OR	1	1	1
ARMA-GARCHLM	1.115	1.225	1.334

Notes: Values in bold indicate that the corresponding models have the best forecasting performance.

Table 6: Empirical results. Relative root mean squared forecast errors (RMSFEs) for competing autoregressive ordinal-response models (Caterpillar stock).

Models	$r = 1$	$r = 5$	$r = 10$
ARMA-OR-GARCHLM	<u>1.022*</u>	<u>1.027*</u>	<u>1.029*</u>
ARMA-OR-GARCHL	<u>1.037*</u>	<u>1.043*</u>	1.051*
ARMA-OR-GARCHM	<u>1.038*</u>	<u>1.040*</u>	<u>1.044*</u>
ARMA-OR-GARCH	1.045*	1.044*	1.040*
AR-OR-GARCHLM	1.017*	1.013*	1.010*
AR-OR-GARCHL	1.012*	1.011*	1.010
AR-OR-GARCHM	1.033*	1.022*	1.017*
AR-OR-GARCH	1.035*	1.032*	1.015
AR-OR	1	1	1
ARMA-GARCHLM	0.972*	0.835	0.770

Notes:

1) Values in bold indicate that the corresponding models have the best forecasting performance.

2) * indicates that the respective model performs significantly different from the baseline model (AR-OR model) at the 5% level, based on the p -values of the Giacomini-White (2006) test. The table of these values is given in the Online Appendix.

3) Underlined numbers indicate that the corresponding models are included in the Model Confidence Set (Hansen et al., 2011). The confidence level for MCS is 95%. The table of the MCS p -values is given in the Online Appendix.

Table 7: Summary table for the LPS results.

Data	$r = 1$	$r = 5$	$r = 10$
I	AR-OR-GARCH	AR-OR-GARCHLM	AR-OR-GARCHLM
II	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
III	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
IV	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
V	ARMA-OR-GARCHLM	ARMA-OR-GARCHL	ARMA-OR-GARCHLM
VI	ARMA-OR-GARCHM	ARMA-OR-GARCHM	ARMA-OR-GARCHLM
VII	ARMA-OR-GARCHM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
VIII	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
IX	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
X	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM

Notes: This table reports the best forecasting model for each data set and for each horizon r , based on the LPS values. In the Online Appendix we report the full forecasting results for the data set II only. The forecasting results for the data sets III-X are available upon request.

Table 8: Summary table for the RMSFE results.

Data	$r = 1$	$r = 5$	$r = 10$
I	ARMA-OR-GARCH	ARMA-OR-GARCH	ARMA-OR-GARCHL
II	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
III	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
IV	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
V	ARMA-OR-GARCHLM	ARMA-OR-GARCH	ARMA-OR-GARCHL
VI	ARMA-OR-GARCHLM	ARMA-OR-GARCH	ARMA-OR-GARCHM
VII	ARMA-OR-GARCHLM	ARMA-OR-GARCH	ARMA-OR-GARCHLM
VIII	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHL
IX	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM
X	ARMA-GARCHLM	ARMA-OR-GARCHLM	ARMA-OR-GARCHLM

Notes: This table reports the best forecasting model for each data set and for each horizon r , based on the RMSFE values. In the Online Appendix we report the full forecasting results for the data set II only. The forecasting results for the data sets III-X are available upon request.

Table 9: Summary table for the MCS results.

Data	$r = 1$	$r = 5$	$r = 10$
I	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCHM ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCHM	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCHM
II	ARMA-OR-GARCH AR-OR-GARCHLM AR-OR-GARCHL AR-OR-GARCHM AR-OR-GARCH ARMA-GARCHLM	ARMA-OR-GARCH AR-OR-GARCHLM AR-OR-GARCHL AR-OR-GARCHM AR-OR-GARCH ARMA-GARCHLM	ARMA-OR-GARCH AR-OR-GARCHLM AR-OR-GARCHL AR-OR-GARCHM AR-OR-GARCH ARMA-GARCHLM
III	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCHM	ARMA-OR-GARCHL ARMA-OR-GARCHM	ARMA-OR-GARCHLM ARMA-OR-GARCHM
IV	ARMA-OR-GARCHLM ARMA-OR-GARCHL	ARMA-OR-GARCHLM ARMA-OR-GARCHL	ARMA-OR-GARCHLM ARMA-OR-GARCHL
V	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCHM	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCHM	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCHM
VI	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCHM ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCHL ARMA-OR-GARCH
VII	ARMA-OR-GARCHM ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCHM ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCH
VIII	ARMA-OR-GARCHLM ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCHM ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCH
IX	ARMA-OR-GARCHLM ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCHM ARMA-OR-GARCH	ARMA-OR-GARCHLM ARMA-OR-GARCH
X	ARMA-OR-GARCHLM ARMA-GARCHLM	ARMA-OR-GARCHLM ARMA-OR-GARCHM ARMA-GARCHLM	ARMA-OR-GARCHLM ARMA-OR-GARCH ARMA-GARCHLM

Notes: This table reports the Model Confidence Set for each data set and for each horizon r , based on the MCS p -values. In the Online Appendix we report the full forecasting results for the data set II only. The forecasting results for the data sets III-X are available upon request.

Table 10: Summary table for the Geweke-Amisano results.

Data	$r = 1$	$r = 5$	$r = 10$
I	ARMA-OR-GARCHLM (0.799) ARMA-OR-GARCHL (0.140) ARMA-OR-GARCHM (0.061)	ARMA-OR-GARCHLM (0.855) ARMA-OR-GARCHL (0.120) ARMA-OR-GARCH (0.025)	ARMA-OR-GARCHLM (0.620) ARMA-OR-GARCHL (0.214) ARMA-OR-GARCHM (0.152) ARMA-OR-GARCH (0.014)
II	ARMA-OR-GARCHLM (0.614) ARMA-OR-GARCHL (0.221) ARMA-OR-GARCHM (0.032) ARMA-OR-GARCH (0.133)	ARMA-OR-GARCHLM (0.789) ARMA-OR-GARCHL (0.211)	ARMA-OR-GARCHLM (0.892) ARMA-OR-GARCHL (0.053) ARMA-OR-GARCHM (0.022) ARMA-OR-GARCH (0.033)
III	ARMA-OR-GARCHLM (0.732) ARMA-OR-GARCHL (0.128) ARMA-OR-GARCHM (0.14)	ARMA-OR-GARCHLM (0.812) ARMA-OR-GARCHL (0.052) ARMA-OR-GARCHM (0.071) ARMA-OR-GARCH (0.065)	ARMA-OR-GARCHLM (0.935) ARMA-OR-GARCHL (0.065)
IV	ARMA-OR-GARCHLM (0.810) ARMA-OR-GARCHL (0.190)	ARMA-OR-GARCHLM (0.834) ARMA-OR-GARCHL (0.166)	ARMA-OR-GARCHLM (0.944) ARMA-OR-GARCHL (0.056)
V	ARMA-OR-GARCHLM (0.940) ARMA-OR-GARCHL (0.060)	ARMA-OR-GARCHLM (0.717) ARMA-OR-GARCHL (0.133) ARMA-OR-GARCHM (0.120) ARMA-OR-GARCH (0.030)	ARMA-OR-GARCHLM (0.832) ARMA-OR-GARCHL (0.056) ARMA-OR-GARCHM (0.071) ARMA-OR-GARCH (0.041)
VI	ARMA-OR-GARCHLM (0.540) ARMA-OR-GARCHL (0.230) ARMA-OR-GARCHM (0.120) ARMA-OR-GARCH (0.11)	ARMA-OR-GARCHLM (0.550) ARMA-OR-GARCHM (0.450)	ARMA-OR-GARCHLM (0.815) ARMA-OR-GARCHM (0.117) ARMA-OR-GARCH (0.068)
VII	ARMA-OR-GARCHLM (0.730) ARMA-OR-GARCH (0.270)	ARMA-OR-GARCHLM (0.710) ARMA-OR-GARCHM (0.220) ARMA-OR-GARCH (0.070)	ARMA-OR-GARCHLM (0.810) ARMA-OR-GARCH (0.190)
VIII	ARMA-OR-GARCHLM (0.85) ARMA-OR-GARCH (0.15)	ARMA-OR-GARCHLM (0.713) ARMA-OR-GARCHM (0.133) ARMA-OR-GARCH (0.154)	ARMA-OR-GARCHLM (0.832) ARMA-OR-GARCH (0.168)
IX	ARMA-OR-GARCHLM (0.600) ARMA-OR-GARCH (0.400)	ARMA-OR-GARCHLM (0.710) ARMA-OR-GARCHM (0.23) ARMA-OR-GARCH (0.06)	ARMA-OR-GARCHLM (0.640) ARMA-OR-GARCH (0.360)
X	ARMA-OR-GARCHLM (0.550) ARMA-GARCHLM (0.450)	ARMA-OR-GARCHLM (0.623) ARMA-OR-GARCHM (0.167) ARMA-GARCHLM (0.210)	ARMA-OR-GARCHLM (0.644) ARMA-OR-GARCH (0.140) ARMA-GARCHLM (0.216)

Notes: This table reports the optimal pool of forecasting models for each data set and for each horizon r . The numbers in parentheses are the weights for the pool optimized over all the last 500 observations. The sum of these numbers is equal to 1. The model in bold received most of the weight. In the Online Appendix we report the full forecasting results for the data set II only. The forecasting results for the data sets III-X are available upon request.

Table 11: Summary table for the trading rule.

Data	$r = 1$	$r = 5$	$r = 10$
I	5.30%	4.17%	3.22%
II	2.17%	4.72%	5.51%
III	7.33%	6.88%	6.32%
IV	8.25%	8.71%	9.32%
V	3.71%	4.15%	5.08%
VI	4.88%	5.14%	5.33%
VII	4.77%	4.90%	5.12%
VIII	2.44%	4.85%	4.85%
IX	3.15%	4.03%	3.01%
X	2.16%	2.22%	2.28%

Notes: This table reports the percentage by which the average excess return from the trading rule for the optimal pool of models exceeds that for the ARMA-GARCHLM model.