UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 08-18
Economic Series (10)
April 2008

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

GENERAL TO SPECIFIC MODELLING OF EXCHANGE RATE VOLATILITY: A FORECAST EVALUATION*

Luc BAUWENS[1] and Genaro SUCARRAT[2]

29 April 2008

Abstract

The general-to-specific (GETS) methodology is widely employed in the modelling of economic series, but less so in financial volatility modelling due to computational complexity when many explanatory variables are involved. This study proposes a simple way of avoiding this problem when the conditional mean can appropriately be restricted to zero, and undertakes an out-of-sample forecast evaluation of the methodology applied to the modelling of weekly exchange rate volatility. Our findings suggest that GETS specifications perform comparatively well in both ex post and ex ante forecasting as long as sufficient care is taken with respect to functional form and with respect to how the conditioning information is used. Also, our forecast comparison provides an example of a discrete time explanatory model being more accurate than realised volatility ex post in 1 step forecasting.

JEL Classification: C53, F31
Keywords: Exchange Rate Volatility, General to Specific, Forecasting

1 CORE and Department of Economics, Université catolique de Louvain (Belgium). Email: bauwens@core.ucl.ac.be.

2 Corresponding author. Department of Economics, Universidad Carlos III de Madrid (Spain). Email: gsucarra@eco.uc3m.es. Homepage: Http://www.eco.uc3m.es/sucarrat/index.html.

*We are greatly indebted to Dagfinn Rime for providing us with some of the data, and for allowing us to draw on our joint research without wishing to hold him responsible for the result. We are also indebted to various people for useful questions, comments or suggestions at different stages, including Farooq Akram, Vincent Bodart, Eric Ghysels, Roger Hammersland, Sébastien Laurent, Ragnar Nymoen, Eric Renault, Fatemeh Shadman, two anonymous referees, participants at Forskermøtet in January 2007 (Tromsø), participants at the Zeuthen Workshop on Financial Econometrics in December 2006 (Copenhagen), participants at ESEM in August 2006 (Vienna), participants at ISF in June 2006 (Santander), seminar participants at Departamento de Economía de Empresa at UC3M in May 2006, participants at the International Conference on High Frequency Finance in May 2006 (Konstanz), seminar participants at CEMFI in Madrid February 2006, conference participants at the 16th (EC)2 in December 2005 (Istanbul), seminar participants at NHH in November 2005 (Bergen), seminar participants at SSB in September 2005 (Oslo), participants at the 3rd Oxmetrics Conference in August 2005 (London), and participants at the bi-annual doctoral workshop in economics at Université catolique de Louvain in May 2005 (Louvain la

1 Introduction

Exchange rate variability is an issue of great importance for both businesses and policymakers. Businesses use volatility models as tools in their risk management and as input in derivative pricing, whereas policymakers use them to acquire knowledge about what and how economic factors impact upon exchange rate variability for informed policymaking. Most volatility models are highly non-linear and thus require complex optimisation algorithms in empirical application. For models with few parameters and few explanatory variables this may not pose unsurmountable problems. But as the number of parameters and explanatory variables increases the resources needed for reliable estimation and model validation multiply. Indeed, this may even become an obstacle to the application of certain econometric modelling strategies, as for example argued by Granger and Timmermann (1999), and McAleer (2005) regarding automated general-to-specific (GETS) modelling of financial volatility.[1] GETS modelling is particularly suited for explanatory econometric modelling since it provides a systematic framework for statistical economic hypothesis-testing, model development and model (re-)evaluation, and the methodology is relatively popular among large scale econometric model developers and proprietors. However, since the initial model formulation typically entails many explanatory variables this poses challenges already at the outset for computationally complex models.

The recent developments in Hendry et al. (2007) and Doornik (2008) might be a step towards overcoming some of the computational challenges associated with maximum likelihood estimation of financial models when many variables are included in the variance specification. However, this remains to be investigated since their work is on the conditional mean using ordinary least squares estimation. Meanwhile, in this study we overcome the computational challenges traditionally associated with the application of the GETS methodology in the modelling of financial volatility by modelling volatility within an exponential model of variability (EMOV), where variability is defined as squared returns. The parameters of interests can thus be consistently estimated with ordinary least squares (OLS) under rather weak assumptions. This setup implies that the conditional mean is restricted to zero, but in return it enables us to apply GETS to a general specification with, in our case, a constant and twenty four regressors, including lags of log of

Neuve). Errors and interpretations being our own applies of course.

Genaro Sucarrat acknowledges financial support from The Finance Market Fund (Norway), from the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, MICFINMA, and from Norges Bank's fund for economic research.

[1] A distinction between general-to-specific specification search on the one hand and the GETS methodology on the other is useful at this point. General-to-specific specification search plays a central role in the GETS methodology, but the methodology additionally embodies a particular view about the relation between reality and empirical models, which gives rise to a certain set of model evaluation criteria and modelling objectives. The GETS methodology is also sometimes referred to as the "LSE methodology" after the institution in which the methodology to a large extent originated in, the "Hendry methodology" after the most influential and arguably the most important contributor to the development of the methodology, and sometimes even "British econometrics", see Gilbert (1989), Gilbert (1990), Mizon (1995), Hendry (2003) and Campos et al. (2005).

squared return, an asymmetry term, a skewness term, seasonality variables, and economic covariates. Compared with models of the autoregressive conditional heteroscedasticity (ARCH) and stochastic volatility (SV) classes we estimate and simplify our specification with little effort, and obtain a parsimonious encompassing specification with uncorrelated homoscedastic residuals and relatively stable parameters. Moreover, our out-of-sample forecast evaluation suggests that GETS specifications can be particularly valuable in conditional forecasting—as long as sufficient care is taken as to where and how the conditioning information enter, since the *ex post* EMOV specification performs particularly well.

Another contribution of this study consists of a qualificatory note on the evaluation of explanatory economic models of financial volatility against estimates based on continuous time theory. Highly simplified, the return volatility forecasting literature can be divided in two: Before and after the highly influential publication of Andersen and Bollerslev (1998). Although in-sample estimates suggest the widespread presence of ARCH, asymmetry effects, jumps, volume effects and so on in financial returns volatility, models that include these effects tend to explain a very small portion of return variability out-of-sample, see Poon and Granger (2003) for a review of the literature. Andersen and Bollerslev (1998) argued that this is because the standard estimates of volatility are very noisy and suggested instead that forecasts of volatility should be evaluated against high frequency *ex post* estimates, for example realised volatility (sums of intra-period squared returns). Andersen and Bollerslev (1998) were not the first to put forward this explanation and solution, but they nevertheless had the greatest impact. Subsequently the general view that has emerged is that discrete time models of financial volatility should be evaluated against estimates derived from continuous time theory, and not against return variability (for example squared returns), see *inter alia* Andersen et al. (1999), Andersen et al. (2003), Hansen and Lunde (2005), Andersen et al. (2005), Hansen and Lunde (2006), Andersen et al. (2006). The consequence of this is that little if any role is left for the residuals—directly or indirectly—to play in the forecast evaluation. This is counter to the GETS methodology where analysis of the residuals plays a key role in model evaluation and model comparison, since any empirical model is a highly simplified representation of the data generating process. Here we qualify the view that discrete time models of financial volatility should be evaluated against estimates derived from continuous time theory. Specifically, we argue that this is particularly inappropriate in the evaluation of explanatory economic models of financial volatility.

The rest of the paper is divided into four sections. The next section gives a brief exposition of the GETS methodology, explains why evaluation against high-frequency estimates based on continuous time theory in a sense is counter to the GETS methodology, and presents the EMOV and its relation with the more common ARCH and SV models. Then we present the data and empirical models in section 3, whereas section 4 contains the results of the *ex post* and *ex ante* out-of-sample forecast exercises. The *ex post* evaluation is of special interest in the current context. The GETS methodology is particularly suited for the development of explanatory models useful for conditional forecasting and scenario analysis more generally, and the accuracy of *ex post* forecasts is an indication of the usefulness for these purposes. In the final section we conclude and provide suggestions

for further research.

2 Modelling volatility

This section proceeds in two steps. In the first subsection we give a brief overview of the GETS methodology and its application in volatility modelling, our main objective being to explain why evaluation against high-frequency volatility estimates based on continuous time theory can be incompatible with the GETS methodology. In the second subsection we describe the EMOV and compare it with the more common ARCH and SV families of models.

2.1 The GETS methodology

A cornerstone of the GETS methodology is that empirical models are derived, simplified representations of the immensely complex and unknown joint density that generate the data (the DGP).[2] So instead of postulating a uniquely "true" model or class of models, the aim is to develop "congruent" encompassing models within the statistical framework of choice, where congruency refers to five properties that the empirical model should ideally exhibit: 1) That the error is an innovation, 2) that the conditioning variables are weakly exogenous with respect to the parameters of interest, 3) that the parameters are stable over the estimation sample, 4) that the model is economically founded, and 5) that the model is data-admissible. The GETS methodology mimics a reduction theory which originally was formulated in terms of discrete time variables, see Hendry and Richard (1990), and Hendry (1995, chapter 9). However, the non-restrictive modifications proposed in Sucarrat (2007) generalises reduction analysis to be applicable to continuous time models as well. In econometric practice GETS modelling proceeds in cycles of three steps. First formulate a general unrestricted model (GUM) which is congruent, second simplify the model sequentially in an attempt to derive a parsimonious congruent model while at each step checking that the model remains congruent, and finally test the resulting congruent model against the GUM. The test of the final model against the GUM serves as a parsimonious encompassing test, that is, a test of whether important information is lost or not in the simplification process. If the final model is not congruent or if it does not parsimoniously encompass the GUM, then the cycle starts all over again by re-specifying the GUM. As such the GETS methodology treats modelling as a process, where the aim is to derive a parsimonious congruent encompassing model while at the same time acknowledging that "the currently best available model" (Hendry and Richard 1990, p. 323) can always be improved.

[2] GETS modelling is an attempt to mimic statistical reduction theory in general and Hendry's reduction theory (1995, chapter 9) in particular. For more elaborate expositions of the GETS methodology and further discussion, see Hendry and Richard (1990), Hendry (1993), Gilbert (1990), Mizon (1995), Hendry (1995), Jansen (2002), Campos et al. (2005) and Sucarrat (2007).

To see the relation between discrete time models of volatility and the GETS methodology consider the discrete time model

$$r_t = \mu_t + e_t, \tag{1}$$

where r_t is the log-return of a financial asset, μ_t is equal to the conditional mean $E(r_t|\mathcal{I}_t)$, $\mathcal{I}_t$ is the conditioning information set in question at t, and $\{e_t\}$ is a sequence of errors. If (1) is congruent and $\{e_t\}$ the innovation errors with respect to $\{\mathcal{I}_t\}$ in the sense that $e_t|\mathcal{I}_t$ is distributed the same for each t, then $\{e_t\}$ is homoscedastic and there is no need for volatility modelling, since the conditional variance $Var(r_t|\mathcal{I}_t) = \sigma^2$, that is, volatility, is constant. Moreover, encompassing considerations are typically undertaken in terms of the $\{e_t\}$, either directly or indirectly. For example, in parsimonious encompassing evaluation in terms of the F-test then the $\{e_t\}$ are used directly, whereas in forecast encompassing evaluation of forecasts of, say, r_t or r_t^2 then the $\{e_t\}$ affect forecast precision indirectly. Now, consider the discrete time model

$$r_t = \mu_t + e_t, \qquad e_t = \sigma_t z_t, \tag{2}$$

where $E(z_t|\mathcal{I}_t) = 0$ for all t, and where the conditional variance $Var(r_t|\mathcal{I}_t) = \sigma_t^2$ is non-constant. Accordingly, $\{e_t\}$ is heteroscedastic and for (2) to be congruent $\{\sigma_t\}$ needs to be specified such that $\{z_t\}$ is an innovation process, that is, that $z_t|\mathcal{I}_t$ is distributed the same for each t. In analogy to the case where the $\{e_t\}$ are homoscedastic it is natural to consider encompassing properties in terms of the $\{z_t\}$, directly or indirectly.

Empirical models are data-based whereas continuous time models are theory constructs. To see the relation between continuous time models of volatility and the GETS methodology a useful distinction is that between *empirical* and *theoretical* congruency, that is, congruency of an empirical model on the one hand and congruency of a theory model on the other. An empirical model is said to be congruent if it is a congruent representation of the DGP, whereas a theory model is said to be congruent if it is a congruent representation of the theory mechanism, that is, the joint density that precedes the DGP in Hendry's reduction theory (see Hendry 1995, p. 345 and Sucarrat 2007). Now, consider the continuous time model

$$r(t) = A(t) + M(t), \tag{3}$$

where $r(t)$ is the log-return $p(t) - p(0)$ of the asset price under study from time 0 to t, $A(t)$ is a locally integrable and predictable process of finite variation, and $M(t)$ is a local martingale, see Andersen et al. (2001). Examples of continuous time models that are contained in this formulation are Itô, jump and jump-diffusion processes. For instance, by setting $A(t)$ equal to $\int_0^t \mu(s)ds$ and $M(t)$ equal to $\int_0^t \sigma(s)W(s)ds$, where μ and σ are continuous processes, and where W is a standard Wiener process, we obtain the Itô process

$$r(t) = \int_0^t \mu(s)ds + \int_0^t \sigma(s)W(s)ds. \tag{4}$$

In this particular case the quadratic variation $\int_0^t \sigma(s)^2 ds$ serves as the counterpart of volatility in the discrete time models (1) and (2), and a common estimator of quadratic variation is the sum of intra-period squared returns (realised volatility), see Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002), Andersen et al. (2003) and Aït-Sahalia (2006). Starting from (3) as if it were the theory mechanism can be of great use in many contexts, for example in derivative pricing. Nevertheless, from the point of view of the GETS methodology (3) is at best a congruent representation of the theory mechanism and not the theory mechanism itself. Since discrete time models like (1) and (2) are compatible with more continuous time models than those contained in (3), assuming that the latter is congruent in addition to (1) or (2) actually constitutes a probabilistic restriction (see Sucarrat 2007, section 4.2). At first sight the restrictions implied by (3) may seem innocuous, since the discrete time models contained in, say, (2) may be derived from (3). However, due to probabilistic, economic and practical reasons, starting from (3) as if it were the theoretical mechanism may actually constitute a very strong assumption—in particular in explanatory econometric modelling (Sucarrat 2008 contains a more complete discussion). First, for probabilistic reasons the set of possible worlds in which (3) is congruent is likely to be much smaller than the set of possible worlds in which (2) is congruent, since the former requires that price increments behave according to (3) at all increment-lengths, not only at certain increment-lengths as implied by (2).[3] Second, for economic reasons $A(t)$ and the explanatory component of $M(t)$ ($\int_0^t \sigma(s)ds$ in (4)) are likely to account for a decreasing portion of the total variation in $r(t)$ as the time-increment decreases, since time is needed for an event—or as is typically the case, a combination of events—to bring about another event. For example, the economic rationale behind Evans and Lyons' (2002) currency order flow measure is that information disseminates sequentially and aggregates temporally, so that time is needed for it to have an effect. Finally, for practical reasons neither $A(t)$ nor $M(t)$ in (3) are likely to account for a notable portion of the total variation in $r(t)$, since explanatory data is less often available at high frequencies.

2.2 The EMOV

If s_t denotes the log of an exchange rate and r_t its log-return, then the EMOV is given by

$$r_t^2 = \exp(\mathbf{b}'\mathbf{x}_t + \nu_t), \tag{5}$$

where $\mathbf{b}$ is a parameter vector, $\mathbf{x}_t$ is a vector of conditioning variables and $\{\nu_t\}$ are the errors. The exponential specification is motivated by several reasons. The most straightforward is that it results in simpler estimation compared with the more common ARCH and SV models, in particular when many explanatory variables are involved. Under the assumption that $\{r_t^2 = 0\}$ is an event with probability zero, then consistent and asymptotically normal estimates of $\mathbf{b}$ can be obtained almost surely with OLS under standard

[3] In terms of the concepts and terminology in Sucarrat (2007), the intersection of the set of possible worlds in which (3) is congruent and the set of possible worlds in which (2) is congruent, is either equal to or smaller than the set of possible worlds in which (2) is congruent.

assumptions, since

$$\log r_t^2 = \mathbf{b}'\mathbf{x}_t + \nu_t \qquad \text{with probability 1.} \tag{6}$$

Another motivation for the exponential specification is that large values of r_t^2 become less influential. A third motivation, pointed to by (amongst others) Engle (1982), Geweke (1986) and Pantula (1986), and which subsequently led Nelson (1991) to formulate the exponential general ARCH (EGARCH) model, is that it ensures positivity. This is particularly useful in empirical analysis because it ensures that fitted values of variability are not negative. Finally, another attractive feature of the exponential specification is that it produces residuals closer to the normal in (6) and thus presumably leads to faster convergence of the OLS estimator. In other words, the log-transformation is likely to result in sounder inference regarding $\mathbf{b}$ in (6) when an asymptotic approximation is used. Applying the conditional expectation operator in (5) gives

$$E(r_t^2|\mathcal{I}_t) = \exp(\mathbf{b}'\mathbf{x}_t) \cdot E[\exp(\nu_t)|\mathcal{I}_t], \tag{7}$$

where $\mathcal{I}_t$ denotes the information set in question. Estimates of $E(r_t^2|\mathcal{I}_t)$ are then readily obtained if either $\{\nu_t\}$ is IID or if $\{\exp(\nu_t)\}$ is a mean innovation, that is, if $E[\exp(\nu_t)|\mathcal{I}_t] = E[\exp(\nu_t)]$ for $t = 1, \ldots, T$, since the formula $\frac{1}{T}\sum_{t=1}^{T} \exp(\hat{\nu}_t)$ then provides a consistent estimate of the proportionality factor $E[\exp(\nu_t)|\mathcal{I}_t]$.

To see the relation between the EMOV and the ARCH and SV families of models, recall that the latter two decompose returns into a conditional mean μ_t and a remainder $e_t = \sigma_t z_t$ as in (2) above. If σ_t^2 follows a non-stochastic autoregressive process and if $Var(r_t^2|\mathcal{I}_t) = \sigma_t^2$, then (2) belongs to the ARCH family.[4] A common example is the GARCH(1,1) of Bollerslev (1986)

$$\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta\sigma_{t-1}^2, \tag{8}$$

with $z_t \sim IN(0,1)$. Explanatory terms, say, $\mathbf{c}'\mathbf{y}_t$, would typically enter additively in (8). If σ_t^2 on the other hand follows a stochastic autoregressive process, then (2) belongs to the SV family of models, and in the special case where σ_t and z_t are independent the conditional variance equals $E(\sigma_t^2|\mathcal{I}_t)$.

The EMOV can be seen both as a direct model of variability r_t^2, and as an approximation to the ARCH and SV families of models of volatility. To see this consider the specification

$$r_t = \sigma_t z_t, \tag{9}$$

where $\{z_t\}$ is a zero mean, unit variance innovation process, and where $\sigma_t = \exp(\mathbf{b}'\mathbf{x}_t)^{\frac{1}{2}} \cdot k$ with $k = E(u_t^2|\mathcal{I}_t)^{\frac{1}{2}}$ and $u_t = kz_t$. Squaring both sides gives (5) above with $\nu_t = \log u_t^2$, and then applying the log gives (6). Now, recall that expected variability within the ARCH family[5] is

$$E(r_t^2|\mathcal{I}_t) = \mu_t^2 + \sigma_t^2. \tag{10}$$

[4]Note that the conditioning information set $\mathcal{I}_t$ in the ARCH model may differ from the conditioning information set in the EMOV model.

[5]No generality is lost by only considering the ARCH family since the same type of argument applies with respect to the SV family under standard assumptions.

In words, the total expected exchange rate variation consists of two components, the squared conditional mean μ_t^2 and the conditional variance σ_t^2. As Jorion (1995, footnote 4 p. 510) has noted, σ_t^2 typically dwarfs μ_t^2 with a factor of several hundreds to one, so the "de-meaned" approximation

$$\mu_t^2 + \sigma_t^2 \approx \sigma_t^2 \tag{11}$$

is often reasonably good in practice. As a consequence, the expression $\exp(\mathbf{b}'\mathbf{x}_t)E[\exp(\nu_t)|\mathcal{I}_t]$ —or in its alternative form $\exp(\mathbf{b}'\mathbf{x}_t)E(u_t^2|\mathcal{I}_t)$—can be interpreted as both a model of variability r_t^2 and as a model of volatility σ_t^2.

3 Data and forecast models

This section presents the data of our study and our empirical forecast models. The economic motivation, justification and interpretation of the models have been dealt with in greater length elsewhere, see Bauwens et al. (2006) and Sucarrat (2006, chapters 1-3), so here we concentrate on their statistical properties. The first subsection describes our data in brief (the data appendix provides more details) and introduces notation. The second subsection contains specifications obtained through GETS modelling. The third and final subsection contains the benchmark or "simple" specifications that serve as a point of comparison. These models are relatively parsimonious and require little development and maintenance effort, thus they may be labelled "simple", and they have a documented forecasting record. Their motivation is that an important issue is whether GETS derived specifications improve upon the forecast accuracy provided by simple models.

3.1 Data and notation

Our weekly data span the period 8 January 1993 to 25 February 2005, a total of 634 observations, and the details of the data transformations and the data sources are given in the appendix. In order to undertake out-of-sample accuracy evaluation we split the sample in two. The estimation and model design sample is 8 January 1993 - 26 December 2003 (573 observations), and the reason we split the sample at this point is that the estimation sample then corresponds to that of Bauwens et al. (2006). In other words, our experiment becomes a true out-of-sample evaluation. No re-estimation is undertaken using data after the week of 26 December 2003. The remaining 61 observations are used for the out-of-sample evaluation. The exchange rate in question is the closing value of the BID NOK/EUR in the last trading day of the week and is denoted by S_t.[6] Note that before 1 January 1999 we use the BID NOK/DEM exchange rate converted to euro-equivalents with the official conversion rate 1.95583 DEM = 1 EURO. The weekly return

[6]During the sample period Norway experienced three different types of exchange rate regimes. Until 1998 Norges Bank (The Central Bank of Norway) actively sought to stabilise the Norwegian krone against its main trading partners. Then she shifted to partial inflation targeting before she was instructed by the Ministry of Finance to fully pursue inflation targeting in March 2001. For more details and a discussion, see Sucarrat (2006, pp. 6-9 in particular) and Bauwens et al. (2006).

is given by $r_t = \log S_t - \log S_{t-1}$, and the weekly variability by $V_t^w = r_t^2$. We will also make use of weekly realised volatility in the 1-week forecast comparison, and the series is computed using intra-weekly 30-minute log-returns.[7] We will make extensive use of the log-transformation applied on variabilities and generally we will follow the convention of denoting such variables in lower case. For example, the log of squared NOK/EUR returns is denoted v_t^w and defined as $v_t^w = \log V_t^w$. Graphs of S_t, r_t and v_t^w are contained in figure 1.

In addition to lags of log of squared returns we also include several other regressors in our specifications. To account for the possibility of skewness and asymmetries in r_t we use the lagged return r_{t-1} for the latter, and an impulse dummy ia_t equal to 1 when returns are positive and 0 otherwise for the former. We also include variables intended to account for the impact of holidays and seasonal variation (Christmas, Easter, etc.). These are denoted h_{lt} with $l = 1, 2, \ldots, 8$, see the appendix for further details. As a measure of week-to-week variation in market activity we use the relative change in the number of quotes. More precisely, if we denote the number of quotes in week t by Q_t and its log-counterpart by q_t, we use Δq_t as our measure of the relative change in market activity from one week to the next. As a measure of the general level of market activity due to (say) the number of traders active or other institutional characteristics we use a lagged smoothed variable, namely $\frac{1}{6}\sum_{i=1}^{6} q_{t-i}$, which is denoted $\bar{q}_{t-1}^{6}$.[8] As a measure of general currency market turbulence we use EUR/USD-variability. If $m_t = \log\,(\text{EUR/USD})_t$, then Δm_t denotes the weekly return of EUR/USD, M_t^w stands for weekly variability and m_t^w is its log-counterpart. The petroleum sector plays a major role in the Norwegian economy, so it makes sense to also include a measure of oil price variability. If the log of the oil price is denoted o_t, then the weekly return is Δo_t, weekly variability is O_t^w with o_t^w as its log-counterpart. We proceed similarly with Norwegian and US stock market variables. If x_t denotes the log of the main index of the Oslo stock exchange, then the associated variables are Δx_t, X_t^w and x_t^w. In the US case u_t is the log of the New York stock exchange (NYSE) composite index and the associated variables are Δu_t, U_t^w and u_t^w. The foreign interest-rate variables that we include are constructed using an index made up of the short term market interest-rates of the EMU countries. Specifically, if IR_t^{emu} denotes this interest-rate index then we include a variable that is denoted ir_t^{emu} and which is defined as $(\Delta IR_t^{emu})^2$. The Norwegian interest-rate variables that we include are constructed using the main policy interest rate variable of the Norwegian central bank. Let F_t denote the main policy interest rate in percentages and let ΔF_t denote the change from the end of one week to the end of the next. Furthermore, let I_a denote an indicator function equal to 1 in the period 1 January 1999 - 30 March 2001 and 0 otherwise, and let I_b denote an indicator function equal to 1 after 30 March 2001 and 0 before. In the first period the Bank pursued a "partial" inflation targeting policy, whereas in the second it pursued a "full" inflation targeting policy. We then have $\Delta F_t^a = \Delta F_t \times I_a$ and $\Delta F_t^b = \Delta F_t \times I_b$, respectively, and f_t^a

[7] This series only spans the period 14 February 2003 - 25 February 2005, because we do not have access to intraday data before this period.

[8] See Sucarrat (2006, pp. 29-31) for its motivation.

and f_t^b stand for $|\Delta F_t^a|$ and $|\Delta F_t^b|$, respectively. Finally, we also include a step dummy sd_t equal to 0 before 1997 and 1 after to account for what appears to be a structural increase in variability.

3.2 GETS forecasting models

This subsection presents our four forecasting models obtained through GETS modelling. Three of the models will be used to generate both *ex post* and *ex ante* forecasts, whereas one of the models will be used to generate *ex ante* forecasts only. The motivation for evaluating the accuracy of *ex post* forecasts is that they mimic situations of conditional forecasting (stress testing, scenario analysis, etc.) and counterfactual analysis. The GETS methodology is particularly suited for the development of models intended for these types of analyses, so the distinction is of great importance. The first model from which both *ex post* and *ex ante* forecasts are generated is obtained through simplification of a general unrestricted model given by

$$\begin{aligned}\text{GUM EMOV1:}\quad v_t^w &= b_0 + b_1 v_{t-1}^w + b_2 v_{t-2}^w + b_3 v_{t-3}^w + b_4 v_{t-4}^w + b_5 \bar{q}_{t-1}^6 \\ &+ b_6 \Delta q_t + b_7 m_t^w + b_8 o_t^w + b_9 x_t^w + b_{10} u_t^w + b_{11} f_t^a + b_{12} f_t^b \\ &+ b_{13} ir_t^{emu} + b_{14} sd_t + b_{15} ia_t + b_{16} r_{t-1} + \sum_{l=1}^{8} b_{16+l} h_{lt} + \nu_t, \quad (12)\end{aligned}$$

and specifically the simplified model is

$$\begin{aligned}\text{GETS EMOV1:}\quad v_t^w &= b_0 + b_2(v_{t-2}^w + v_{t-3}^w) + b_6 \Delta q_t + b_7 m_t^w \\ &+ b_9(x_t^w + u_t^w) + b_{12} f_t^b + b_{13} ir_t^{emu} + b_{14} sd_t + \nu_t. \quad (13)\end{aligned}$$

The Autometrics feature of PcGive 12 (Doornik and Hendry 2007, Doornik 2008), a software that automates GETS specification search, was used in the derivation of GETS EMOV1 with a chosen 10% level both for regressor significance and diagnostic testing.[9] The software proposes a specification with a constant and the nine regressors contained in

[9] Autometrics is similar to PcGets (Hendry and Krolzig 2001), and both Autometrics and PcGets have been tested and calibrated for situations where the residuals of the simulation DGP are serially uncorrelated, homoscedastic and normal. In particular, both use t-tests and F-tests in most of the significance tests. Our residuals are uncorrelated and homoscedastic according to our diagnostic tests, but they are not normal. However, through the "advanced settings" Autometrics can be configured in such a way that inference is valid in large samples even though residuals are not normal (essentially by unchecking normality in the diagnostic options), since t and F-tests are valid in large samples when residuals are non-normal as long as residuals are uncorrelated and homoscedastic. Admittedly, it remains an open question how well Autometrics actually performs in situations where the residuals are not normal, since the calibration of Autometrics is based on simulations with normal residuals. An important characteristic of Autometrics is that it implements multiple path specification search as opposed to a single path specification search. As pointed out by Hoover and Perez (1999), a single specification search might result in "path dependence", in the sense that a relevant variable being removed early on in the search whereas irrelevant variables that proxy its role are retained. In our case a manual/non-automated single path specification search with GUM

GETS EMOV1, and then further coefficient restrictions are imposed and tested which ultimately leads to GETS EMOV1. Finally a parsimonious encompassing test is undertaken of GETS EMOV1 against GUM EMOV1 in terms of a joint restriction test. The two other GETS models from which both *ex post* and *ex ante* forecasts are generated belong to the ARCH framework, and specifically they are

$$\text{QGETS GARCH1:} \quad r_t = b_1\Delta m_t + b_2\Delta o_t + b_4\Delta x_t + b_5\Delta IR_t^{emu} + b_6\Delta F_t^b + e_t,$$
$$e_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma_1\Delta q_t \quad (14)$$

$$\text{QGETS EGARCH1:} \quad r_t = b_1\Delta m_t + b_2\Delta o_t + b_3\Delta u_t + b_4\Delta x_t + b_5\Delta IR_t^{emu} + b_6\Delta F_t^b + e_t,$$
$$e_t = \sigma_t z_t, \qquad \log\sigma_t^2 = \omega + \alpha|\frac{e_{t-1}}{\sigma_{t-1}}| + \beta\log\sigma_{t-1}^2 + \gamma_1\Delta q_t + \gamma_2 sd_t, \quad (15)$$

where σ_t is the conditional standard deviation of r_t, and where $\{z_t\}$ is an IID sequence of zero mean, unit variance variables. QGETS GARCH1 and QGETS EGARCH1 are obtained through a modelling strategy which we call "Quasi" GETS. Since we are unsuccessful in estimating ARCH models due to numerical issues when including all the information used in GUM EMOV1, we use only the significant information in GETS EMOV1 in formulating GUM versions of a GARCH(1,1) and an EGARCH(1,1) model, respectively. Also, in order to avoid numerical issues the information is put either in the conditional mean or in the conditional variance specifications, not in both. Then we simplify each model (through non-automated, single path specification search) by removing insignificant regressors, and the results are QGETS GARCH1 and QGETS EGARCH1. In addition to the fact that the conditional variance σ_t^2 is modelled exponentially, the EGARCH differs from the GARCH in how persistence is measured and in the condition for covariance stationarity. In GARCH(1,1) models the higher $\alpha + \beta$ the higher persistence, and a necessary condition for covariance stationarity is $\alpha + \beta < 1$. In EGARCH(1,1) models the higher $|\beta|$ the higher persistence, and a necessary condition for covariance stationarity is $|\beta| < 1$, see Nelson (1991).

Recursive parameter stability analysis of the first specification (GUM EMOV1) is contained in figure 2, and estimation results and recursive parameter stability analysis of the second specification (GETS EMOV1) is contained in table 1 and in figure 3. Both GUM EMOV1 and GETS EMOV1 exhibit innovation errors in the sense that the nulls of no serial correlation, no autoregressive conditional heteroscedasticity and no heteroscedasticity are not rejected at the 10% significance level, and the recursive parameter stability analysis

EMOV1 as initial model produces a specification identical to that of Autometrics. So path-dependence does not appear to be an important issue in our case. This is in line with White's (1990) theorem which holds that the path dependence problem reduces as the size of the sample increases. Our sample of 573 observations is considerably larger than those investigated by Lovell (1983), Hoover and Perez (1999) and Hendry and Krolzig (1999), the sequence of studies that resulted in PcGets and then Autometrics. Whereas Lovell (1983) used only 23 observations, the other two studies employed a maximum of 140 observations.

suggests parameters are relatively stable.[10] For both GUM EMOV1 and GETS EMOV1 the Chow forecast and breakpoint tests do not signify at the 1% level, but the 1-step forecast tests on the other hand show some signs of instability.[11] The number of spikes that exceeds the 1% critical value in the break-point tests are 12 and 14, respectively. This suggests the presence of some structural instability since on average we would expect only 5 spikes to exceed the 1% critical value (1% of 473 is just below 5).[12] Estimation results of the last two specifications QGETS GARCH1 and QGETS EGARCH1 are contained in table 2, and both exhibit uncorrelated standardised residuals and squared standardised residuals according to the diagnostic tests. The persistence estimate $\alpha + \beta$ (0.150 + 0.852 = 1.002) of the QGETS GARCH1 is slightly greater than one, which usually is interpreted as an indication of a strong persistence of shocks on the conditional variance. However, in our case it is probably due to the structural break around the beginning of 1997. In QGETS GARCH1 the structural break variable sd_t is not retained due to insignificance. In the QGETS EGARCH1 model on the other hand the variable is retained, with the consequence that the persistence estimate of β (0.925) is notably lower than 1 though still relatively high.[13] Removing the structural break variable sd_t increases the estimate of β to 0.982 and has little impact on the other estimates and significance results. Compared with GETS EMOV1 the influence of the US stock market is insignificant in both QGETS GARCH1 and QGETS EGARCH1, whereas comparing QGETS GARCH1 with QGETS EGARCH1 the influence of the Norwegian stock market and the structural break at the beginning of 1997 are only significant in the latter. Finally, the usual asymmetry term of EGARCH models is not included in QGETS EGARCH1 due to insignificance at the 10% level.

The motivation behind a forecast evaluation of *ex post* forecast is to evaluate the usefulness of GETS models in conditional forecasting. Examples of situations where it is useful to generate forecasts under the assumption that the conditioning information is correct are when studying the effect of a policy intervention (say, an interest rate change or a currency market intervention), in stress testing, in event analysis, in conditional asset pricing, in conditional Value-at-Risk analysis, and so on. In the generation of *ex post* forecasts one week ahead we therefore use observed values on the conditioning information on all the variables on the right-hand side. In the generation of forecasts two weeks ahead and onwards we use forecasted values of v_t^w and r_t, and observed values of the other right-hand variables. In other words, forecasts are generated as if the conditioning information—apart from v_t^w and r_t—is known. The *ex post* forecast accuracy of GETS EMOV1, QGETS GARCH1

[10] We do not report the parameter estimates of GUM EMOV1 for expository reasons. The p-values of the tests AR_{1-10}, $ARCH_{1-10}$, $Het.$, $Hetero.$ and JB of the GUM EMOV1 are 0.86, 0.70, 0.47, 0.94 and 0.00, respectively.

[11] If t denotes the sample size, k the number of parameters in $\mathbf{b}$ and M the observation at which recursive estimation starts, then for $t = M, \ldots, T$ the 1-step, breakpoint and forecast tests are computed in PcGive as $F(1, t-k-1)$, $F(T-t+1, t-k-1)$ and $F(t-M+1, M-k-1)$, respectively.

[12] The number 473 is due to the fact that the recursive estimation was initialised at observation number 100.

[13] The restriction $\beta = 1$ is rejected at 2.7% in a one-sided test against the alternative $\beta < 1$, using a $N(0,1)$ distribution on the test statistic.

and QGETS EGARCH1 hence constitute an indication of their potential for out-of-sample conditional forecasting.[14] In the case of QGETS GARCH1 we generate K-weeks ahead volatility forecasts under the assumption that $z_{t+k} \sim iid(0,1)$ for $k = 1, \ldots, K$, using the formula

$$\hat{\sigma}^2_{t+K} = \hat{\omega}\left(\sum_{j=0}^{K-1}(\hat{\alpha}+\hat{\beta})^j\right) + (\hat{\alpha}+\hat{\beta})^{K-1}(\hat{\alpha}\hat{z}_t^2+\hat{\beta})\hat{\sigma}_t^2 + \sum_{j=0}^{K-1}(\hat{\alpha}+\hat{\beta})^j\hat{\gamma}_1\Delta q_{t+K-j}. \tag{16}$$

In the case of EGARCH(1,1) we also generate K-weeks ahead volatility forecasts under the assumption that $z_{t+k} \sim iid(0,1)$ for $k = 1, \ldots, K$, but use the formula

$$\hat{\sigma}^2_{t+K} = \hat{\sigma}_t^{2\hat{\beta}^K}\exp(\hat{\omega}\sum_{j=0}^{K-1}\hat{\beta}^j)\prod_{j=0}^{K-1}E\left[\exp(\hat{\alpha}\hat{\beta}^j|z_{t+K-1-j}|)\right] \tag{17}$$

instead, where each $E\left[\exp(\hat{\alpha}\hat{\beta}^j|z_{t+K-1-j}|)\right]$ is estimated by means of the expression $\frac{1}{T}\sum_{t=1}^{T}\exp(|\hat{z}_t|)^{\hat{\alpha}\hat{\beta}^j}$.

If conditional forecasting is not the objective, or if the conditioning information is uncertain, then it is inappropriate to evaluate a model according to its *ex post* forecast accuracy. This leads to *ex ante* forecasting. In order to mimic such a setting we specify models that use the parameter estimates of GETS EMOV1, QGETS GARCH1 and QGETS EGARCH1, and that use simple rules in forecasting the conditioning information. We refer to these specifications as *ex ante* models and specifically they are

$$\text{GETS EMOV2:}\quad v_t^w = b_0 + b_2(v_{t-2}^w + v_{t-3}^w) + b_7\bar{m}^w + b_9(\bar{x}^w + \bar{u}^w) + b_{13}\bar{ir}^{emu} + b_{14}sd_t + \nu_t, \tag{18}$$

$$\text{QGETS GARCH2:}\quad r_t = e_t, \qquad e_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{19}$$

$$\text{QGETS EGARCH2:}\quad r_t = e_t, \qquad e_t = \sigma_t z_t, \qquad \log\sigma_t^2 = \omega + \alpha\left|\frac{e_{t-1}}{\sigma_{t-1}}\right| + \beta\log\sigma_{t-1}^2 + \gamma_2 sd_t. \tag{20}$$

The variables Δm_t, Δo_t, Δu_t, Δx_t, ΔIR_t^{emu}, ΔF_t^b, Δq_t and f_t^b are all set to zero because this is approximately equal to their in-sample average. The variables m_t^w, x_t^w, u_t^w and ir_t^{emu} are set equal to their sample averages $\bar{m}^w$, $\bar{x}^w$, $\bar{u}^w$ and $\bar{ir}_t^{emu}$ over the period 1 January 1999 - 26 December 2003.[15] In forecasting volatility K-weeks ahead we use similar formulas to

[14] Of course, if a structural break in the parameters occurs prior to the forecast sample, then lack of accuracy is not necessarily indicative of lack of usefulness, see Clements and Hendry (2005) for a discussion.

[15] This sample was chosen because the volatility of r_t looks relatively stable over this period. Specifically, the values of $\bar{m}^w$, $\bar{x}^w$, $\bar{u}^w$ and $\bar{ir}^{emu}$ are 0.280, 0.633, 0.412 and 0.006.

those of the *ex post* models above, that is, similar to equations (16) and (17). Finally, we also include a second "*ex ante*" version of the EMOV GETS1 obtained through automated GETS specification search (using Autometrics) of a general unrestricted specification that does not include contemporaneous information. The motivation for this model is that it is of interest whether modelling of a GUM with *ex ante* information only changes the results. The specific model is given by

$$\text{GETS EMOV3:} \quad v_t^w = b_0 + b_2(v_{t-2}^w + v_{t-3}^w) + b_{14} sd_t + \nu_t, \tag{21}$$

and the estimation results are contained in table 1. The estimates of b_0 and b_2 are similar to those for GETS EMOV1, but the estimate of b_{14} is somewhat higher. Presumably this is due to the absence of the contemporaneous conditioning variables.

3.3 Benchmark models

We include five benchmark models in our forecast evaluation, four GARCH specifications and one using realised volatility (RV) as an estimate of volatility. The four GARCH specifications all have the conditional mean set to zero, and specifically they are

$$\text{Historical:} \quad r_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega \tag{22}$$

$$\text{RiskMetrics:} \quad r_t = \sigma_t z_t, \qquad \sigma_t^2 = 0.06 e_{t-1}^2 + 0.94 \sigma_{t-1}^2 \tag{23}$$

$$\text{GARCH3:} \quad r_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{24}$$

$$\text{EGARCH3:} \quad r_t = \sigma_t z_t, \qquad \log \sigma_t^2 = \omega + \alpha |\tfrac{e_{t-1}}{\sigma_{t-1}}| + \beta \log \sigma_{t-1}^2, \tag{25}$$

where $z_t \sim IID(0,1)$. The first specification labelled Historical is a GARCH(0,0) estimated on the sample 1/1/1999 - 26/12/2003 (261 observations). In other words, it is the ARCH-counterpart of the sample variance because it specifies volatility as non-varying, and the estimation sample is due to the fact that return variability appears comparatively stable over this period (see figure 1). The second specification is an exponentially weighted moving average (EWMA), that is, a GARCH(1,1) with ω restricted to zero, with parameter values suggested by RiskMetrics (Hull 2000, p. 372): $\alpha = 0.06$ and $\beta = 0.94$.[16] The third specification is an unrestricted GARCH(1,1) model, whereas the fourth an final specification is an unrestricted EGARCH(1,1) model. The GARCH3 and EGARCH3 models are estimated over the sample 8/1/1993 - 26/12/2003 (573 observations).

[16] To be more precise, the parameter values are those suggested by the 1995 version of RiskMetrics. At least two versions of RiskMetrics have superseded the 1995 May edition (see `http://www.riskmetrics.com/`), but $\alpha = 0.06$ and $\beta = 0.94$ have nevertheless come to be known as "the" RiskMetrics specification.

Estimates and residual diagnostics of the benchmark models are contained in table 3. All models yield standardised residuals that are uncorrelated according to the AR_{1-10} test, but serially uncorrelated squared standardised residuals according to the $ARCH_{1-10}$ test is not achieved by the RiskMetrics specification. This suggests that the RiskMetrics specification is too restrictive. The estimates of α, β are similar to those above, which suggests that GARCH3 and EGARCH3 are likely to produce forecast similar to those of QGETS GARCH2 and QGETS EGARCH2. Again we have not included an asymmetry term in the EGARCH(1,1) due to insignificance at 10%, and forecasts are generated with formulas similar to equations (16) and (17) above.

Our last benchmark model is weekly RV made up of intradaily 30 minute squared returns using end-of-interval mid-point quotes from Olsen Financial Technologies (OFT). Specifically, our RV model is given by

$$\text{RV:} \qquad r_t = \sigma_t z_t, \qquad \sigma_t^2 = \sum_{n(t)=1}^{N(t)} r_{n(t)}^2 \tag{26}$$

$$R^2 \; 0.00 \quad AR_1 \; \underset{[0.77]}{0.088} \quad ARCH_1 \; \underset{[0.19]}{1.758} \quad JB \; \underset{[0.70]}{0.710} \quad T = 46$$

where $r_{n(t)}^2$ is squared NOK/EUR log-return in percent over 30 minute interval number $n(t)$ in week t (each week contains 336 intervals). The term σ_t^2 is RV at t and the diagnostic tests AR_1, $ARCH_1$ and JB are of the standardised residual z_t. The 1-week ahead forecast is given by σ_t^2, that is, RV at t, which effectively is an *ex post* forecast. Forecasts 2-weeks ahead and further, by contrast, are generated by means of an AR(1) model of RV, and are thus *ex ante* forecasts.[17]

4 Out-of-sample forecast evaluation

A view that has gained widespread acceptance lately is that discrete time models of financial volatility should be evaluated against estimates based on continuous time theory, see for example Andersen and Bollerslev (1998), Andersen et al. (1999) and Andersen et al. (2001), Andersen et al. (2003), Hansen and Lunde (2005), Andersen et al. (2005), Hansen and Lunde (2006), and Andersen et al. (2006). Typically these estimators use high frequency intra period data and a common example of such an estimator is realised volatility, that is, the sum of squared intra period returns. An important motivation for using high frequency estimators derived from continuous time theory is that they are more efficient or less "noisy". Although this is possibly the case in situations where one already at the outset chooses to employ a continuous time model for the purpose of, say, derivative pricing,

[17] The NLS estimated AR(1) specification is given by $\sigma_t^2 = 1.372 + 0.424\sigma_{t-1}^2 + u_t$, and only one lag is included because further lags are insignificant at 10%. The in-sample R^2 of the fitted model is 18%, and diagnostic tests suggest the standardised residuals are serially uncorrelated and normal, although somewhat heteroscedastic.

comparing an empirical model with estimates obtained from another model is suboptimal in progressive explanatory empirical modelling. Analysis and comparison of residuals is integral to model evaluation and encompassing considerations. So restricting model evaluation to a comparison of volatility estimates from a low frequency model to that of a high frequency model, effectively implies that the residuals play little or no part. As argued in subsection 2.1, when evaluating explanatory volatility models the natural approach is to undertake encompassing and model comparison considerations in terms of the standardised residuals $\{\hat{z}_t\}$, for example in terms of the sample kurtosis.[18] However, this might unfairly benefit the ARCH models, since most of their conditioning information appears in the conditional mean rather than in the conditional variance. Also, the simulations in Sucarrat (2008) suggest that the sample kurtosis is not reliable in reproducing the correct ranking among explanatory models. For these reasons we compare models in terms of their predictability of r_t^2. In the ARCH case $E(r_t^2|\mathcal{I}_t)$ is equal to $\mu_t^2 + \sigma_t^2$ and serves as the prediction of r_t^2, whereas in the EMOV case $E(r_t^2|\mathcal{I}_t)$ is equal to $\exp(\mathbf{b}'\mathbf{x}_t)E[\exp(u_t)|\mathcal{I}_t)$ and serves as the prediction of r_t^2.

We evaluate both *ex post* and *ex ante* forecasts. The objective of an *ex post* evaluation is to study the accuracy of GETS models in conditional forecasting situations. In other words, how well the GETS models forecast given that the values of the conditioning variables are correct. Examples of such situations are when studying the effect of an interest rate change or a currency market intervention (or, more generally, a policy intervention), in stress testing, in event analysis, in conditional asset pricing, in conditional Value-at-Risk analysis, and so on. Since the GETS methodology is especially suited for the modelling of explanatory models that possibly include many variables, *ex post* accuracy is arguably more important than *ex ante* accuracy. If correctly predicting the values of the conditioning variables does not improve upon forecast accuracy beyond that of non-explanatory models—in our case the benchmark models, then this suggests the GETS models do not constitute an improvement in conditional forecasting compared with the non-explanatory models. The objective of an *ex ante* evaluation by contrast is to shed light on the accuracy of explanatory models when the values of the conditioning variables are uncertain. This is typically the case when conditional forecasting is not the objective, and then an evaluation of *ex ante* forecasts is more appropriate. In *ex ante* forecasting the conditioning information has to be forecasted, so one cannot expect the GETS models to forecast better than the benchmark models. But ideally the GETS models should forecast at least as well. If this is the case, then the GETS models serve both purposes, conditional forecasting and *ex ante* forecasting.

The out-of-sample evaluation is undertaken on the period 2 January 2004 to 25 February 2005 (61 observations), and the section proceeds in three steps. The first subsection contains a comparison of so-called Mincer-Zarnowitz (1969) regressions of squared returns on a constant and 1-step forecasts, whereas the second contains an out-of-sample fore-

[18] By construction the second moment of z_t is equal to 1, so moments of higher order are needed. The third moment is equal to zero if the standardised residual is distributed symmetrically, so the estimate of the fourth moment suggests itself naturally.

cast accuracy comparison in terms of the mean of the squared forecast errors. The third and final subsection sheds additional light on the results by examining some of the 1-step forecast trajectories more closely.

4.1 1-step Mincer-Zarnowitz regressions

A simple way of evaluating forecast models is by regressing the variable to be forecasted on a constant and on the forecasts, so-called Mincer-Zarnowitz (1969) regressions, see Andersen and Bollerslev (1998), Patton (2007) and Sucarrat (2008). In our case this proceeds by estimating the specification

$$r_t^2 = a + b\hat{V}_t + e_t, \tag{27}$$

where $\hat{V}_t$ is the 1-step forecast and e_t is the error term. Ideally, a should equal zero and b should equal one (these characteristics constitute conditions for "unbiasedness"), and the fit should be high. Among these properties the simulation results in Sucarrat (2008) suggest that the fit in terms of R^2 and the joint test $a = 0, b = 1$ are the most informative.

Table 4 contains the regression output organised according to forecast model categories, GETS *ex post*, GETS *ex ante* and benchmark. One specification stands out according to the majority of the criteria, namely the *ex post* model GETS EMOV1. Its estimate of a is not significantly different from zero, the estimate of b is positive and significant at the 5% level, the joint restriction $a = 0, b = 1$ is not rejected at conventional significance levels, and its R^2 is 0.25. This is higher than the 14% attained by the *ex post* benchmark model RV, and it is substantially higher than any of the R^2s cited in Andersen and Bollerslev (1998, pp. 890-891)—the typical R^2 they cite is around 0.03 and the highest is 0.11, and must be very close to—if not exceeding—the population upper bound of R^2 for a GARCH(1,1) model, see Andersen and Bollerslev (1998, p. 892). The relatively high R^2 of GETS EMOV1 suggests therefore that the poor *ex post* forecasting performance of r_t^2 by ARCH-models can be improved upon substantially through explanatory modelling. That the source of the high R^2 is explanatory variables is suggested by the fact that the *ex ante* version of GETS EMOV1, that is, GETS EMOV2, only achieves an R^2 of 3%. However, there are no clear signs that the other *ex post* models perform as well as GETS EMOV1. In particular, QGETS GARCH1 and QGETS EGARCH1 achieve an R^2 of only 0% and 4%, respectively, and the joint restriction of $a = 0$ and $b = 1$ is rejected at 11% and 3% in their cases. GETS models are comparatively useful in conditional forecasting and scenario analysis more generally if they perform better than the benchmark models. The R^2 of the first four benchmark models is zero and the joint restriction $a = 0, b = 1$ is rejected at 0%, 3% and 2%, respectively, for the three models in which the joint test can be undertaken. RV is the only benchmark model that does relatively well with an R^2 of 14%, and the tests $a = 0$ and $a = 0, b = 1$ are not rejected at the 10% level. However, RV by itself cannot provide conditional forecasting and scenario analysis more generally. So all in all the Mincer-Zarnowitz regressions provide some support for the usefulness of GETS models in 1-week conditional forecasting of exchange rate volatility. The usefulness of GETS models seems to be dependent on the model framework (EMOV *vs.* ARCH)

and/or where the conditioning information is included (mean *vs.* variance equation).

In situations where the conditioning information has to be forecasted, then it is inappropriate to evaluate the GETS models in terms of their *ex post* forecast accuracy. The *ex ante* forecasts provide some support for the usefulness of GETS models when the conditioning information has to be forecasted, but the support is weaker than for *ex post* forecasting. Both GETS EMOV2 and GETS EMOV3 models attain an R^2 of only 3%, which is higher than any of the *ex ante* benchmark models (RV is an *ex post* benchmark model). Also, contrary to the benchmark models, for neither GETS EMOV2 nor GETS EMOV3 is the joint restriction $a = 0, b = 1$ rejected. The fit for QGETS GARCH2 and QGETS EGARCH2 by contrast is zero, and for both is the restriction $a = 0, b = 1$ rejected at 2% and 0%, respectively. So again does the usefulness of GETS models seem to be dependent on model framework and/or where the conditioning information is included.

4.2 Out-of-sample MSE comparison

Given a sequence of squared returns $\{V_k\}$ over the forecast periods $k = 1, \ldots, K$ and a corresponding sequence of forecasts $\{\hat{V}_k\}$, we compare the out-of-sample forecast accuracy in terms of mean squared error (MSE)

$$MSE = \frac{1}{K}\sum_{k=1}^{K}(V_k - \hat{V}_k)^2. \tag{28}$$

Error-based measures like MSE are "pure" precision measures in the sense that evaluation is based solely on the discrepancy between the forecast and the actual value. There is a case to be made for the view that precision-based measures are the most appropriate when evaluating the forecast properties of a certain modelling strategy, since this leaves open what the ultimate use of the model is. On the other hand, this is also a weakness since considerations regarding the final use of the model do not enter the evaluation.[19] The values of the MSEs together with the p-values of White's (2000) Reality Check (RC) and Hansen's (2005) test for Superior Predictive Ability (SPA) against Historical are contained in table 5, whereas the p-values of the Modified Diebold-Mariano (MDM) test (Harvey et al. 1997) of each model against Historical are contained in table 6. None of the p-values suggest any model is more accurate than Historical at any horizon (the highest p-value of 19% is produced by GETS EMOV1 in 2 week forecasting by means of the MDM test). But the simulations in Sucarrat (2008) show at any rate that the RC, SPA and MDM tests have very little power in situations similar to the current one. So a low p-value should not necessarily be interpreted as insignificant superior accuracy than Historical.

[19] Several alternative approaches to out-of-sample forecast comparison have been proposed. In West et al. (1993), for example, the expected utility of a risk averse investor serves as the ranking criterion. Engle et al. (1993) provide a methodology in which the profitability of a certain trading strategy ranks the forecasts. Yet another approach takes densities as the object of interest, see Diebold et al. (1998), whereas Lopez (2001) has proposed a framework that provides probability forecasts of the event of interest. The study by González-Rivera et al. (2004) is eclectic in that it includes several loss functions, including an option price based, a utility based, a Value-at-Risk (VaR) based and a likelihood based.

Explanatory models are particularly useful for conditional forecasting and scenario analysis more generally, and a measure of their usefulness for these purposes is out-of-sample *ex post* accuracy beyond that of benchmark models. If explanatory models with *ex post* information do not forecast better than non-explanatory models, then they do not provide insights beyond that of non-explanatory models. Among the GETS *ex post* models only one specification does well compared with the benchmark models, namely GETS EMOV1. Not only does it perform better than the two other GETS *ex post* at all horizons up to 6 week ahead, but it also comes first compared with the other models as well. In 12 weeks forecasting GETS EMOV1 comes 6th, beaten by (amongst others) Historical and the *ex ante* counterpart of GETS EMOV1. This suggests indeed that the GETS methodology can be useful in developing models for conditional forecasting and scenario analysis purposes. However, a caveat is that the two other *ex post* models obtained by (quasi) GETS modelling, that is, QGETS GARCH1 and QGETS EGARCH1, do not perform better than Historical at any forecast horizon. In other words, the GETS methodology by itself does not guarantee a satisfactory model. Care is needed with respect to where and how the information enter in the mean and variance specifications, and with respect to the rigour with which the GETS methodology is implemented. Of some notable interest is the fact that GETS EMOV1 beats the RV model in 1 week (*ex post*) forecasting. RV fares comparatively well in 1 week forecasting, since it comes second overall. However, thereafter it performs miserably since it comes last overall at each horizon. Closer inspection of the forecasts of RV reveals that the reason for its inaccuracy is that RV is biased and tends to overpredict, see figure 4. Indeed, part of the reason why RV does relatively well according to the MSE measure 1 week ahead (second overall) is because of its bias. This underlines the importance of accounting for market microstructure effects when using high-frequency data to estimate continuous time entities, see Sucarrat (2008) for a discussion.

In forecast situations where impact analysis or conditional forecasting is not the objective, then the conditioning information is uncertain and has to be forecasted. This leads to *ex ante* forecasting, and the GETS *ex ante* specifications try to mimic such a situation. One cannot necessarily expect explanatory models to perform better than simple models like, say, Historical, in such situations. However, according to one criterion GETS models should perform at least as good. A first characteristic that emerges from the results in table 5 is that, again, the EMOV specifications GETS EMOV2 and GETS EMOV3 perform better than the ARCH specifications QGETS GARCH2 and QGETS EGARCH2 at all horizons. A second characteristic is that the performance of the *ex ante* versions of the EMOV models is very similar. For the first two horizons GETS EMOV3 comes first, whereas for the other horizons GETS EMOV2 comes first. A third characteristic is that GETS EMOV2 and GETS EMOV3 perform better than the benchmark models up to and including 2 weeks ahead. (In 1 week forecasting RV performs better than GETS EMOV2 and GETS EMOV3, but 1 week ahead the RV forecasts are *ex post* and therefore not comparable.) However, from 3-weeks and onwards they do not forecast better than Historical. Summarised, then, the results suggest that GETS models are more or less at par with—but not better than—simple benchmark models like Historical. This suggests that GETS mod-

els are useful also in *ex ante* forecasting, or at least in short term *ex ante* forecasting. A caveat is that the accuracy of GETS models again depends on functional form and how the information is used (and possibly on the rigour with which the GETS methodology is implemented), since the *ex ante* version of the QGETS models of the ARCH class do not perform very well.

Among the benchmark models Historical stands out since it performs better than the other benchmark models at all horizons except 1 week ahead, where it unsurprisingly is beaten by the *ex post* forecasts of RV. The other benchmark models, that is, RiskMetrics GARCH3, EGARCH3 and RV, are all particularly suited in the presence of autoregressive heteroscedasticity. The fact that these benchmark models do not perform better than Historical suggests there is none or very little of it in weekly exchange rate returns.

4.3 Explaining the forecast results

An important part of an out-of-sample study consists of explaining the results, and to this end figure 5 provides a large part of the answer. The figure contains the out-of-sample trajectories of squared NOK/EUR log-returns in percent r_t^2, the 1-step forecasts of GETS EMOV1 and the 1-step forecasts of Historical, and the figure provides insights on the forecast accuracy results. First, the series of r_t^2 seems to be characterised by some occasional large values but little persistence in the sense that large values do not tend to follow each other. Indeed, only at two instances is a large value followed by another, and for a relatively large portion of the sample r_t^2 stays rather low. This explains to some extent the forecast accuracy of Historical. Second, in the 5th and in the 11th weeks of the forecast sample Norges Bank changed its main policy interest rate. This is reflected in the large values of r_t^2 in the 5th and 11th weeks, and explains the forecast accuracy of GETS EMOV1 (it contains a variable for policy interest rate changes) and its relatively high R-squared in the 1-step forecast Mincer-Zarnowitz regressions. However, the QGETS GARCH1 and QGETS EGARCH1 models contain the same interest rate information, but do not seem to be able to make equally efficient use of it. Finally, the other explanatory variables included in GETS EMOV1 are probably the reason why it also follows r_t^2 relatively well at other instances when r_t^2 moves substantially.

5 Conclusions

This study has evaluated the out-of-sample forecast accuracy of models of weekly NOK/EUR volatility derived by means of the GETS methodology. The results suggest that such models produce unbiased *ex post* and *ex ante* forecasts, and that they perform comparatively well at all horizons. In particular, the explanatory GETS EMOV specification comes first at all horizons up to 6 weeks ahead in *ex post* forecasting, which is indicative of usefulness for conditional forecasting and scenario analysis more generally, whereas in *ex ante* forecasting the GETS EMOV models come first up to 2 weeks ahead and fare comparatively well thereafter. But our results also suggest that the application of GETS specification

search by itself does not guarantee good forecasting models. Care is needed with respect to functional form and as to how and where the conditioning information enter in the mean and variance specifications. The rigour with which the GETS methodology is implemented might also be a factor. Another result of interest in our comparison is that explanatory *ex post* models are capable of providing better predictions of squared returns 1 week ahead than the *ex post* forecasts of realised volatility.

Our findings suggest several lines for further research. First, the generality of our results must be established. Is GETS-modelling of financial volatility useful on higher frequencies—which typically exhibit more volatility persistence—than the weekly? On other exchange rates and for other financial assets? Second, contrary to Granger and Timmermann's (1999) and McAleer's (2005) assertions, automated GETS-modelling of financial volatility can be readily implemented and should be investigated more fully. Finally, a drawback with our approach (the EMOV framework) is that the conditional mean is restricted to zero, which means that predictability in the direction of exchange rate changes cannot be exploited. One interesting line of research is thus to make use of multi-step least squares estimators of conditional heteroscedasticity models so that the numerical issues and problems associated with GETS modelling of volatility are avoided, possibly combined with the procedures in Hendry and Krolzig (2005) and in Hendry et al. (2007), in order to efficiently handle many variables in the initial GUM.

References

Aït-Sahalia, Y. (2006). Estimating Continuous-Time Models with Discretely Sampled Data. Invited lecture at the 2005 World Congress of the Econometric Society. To appear as chapter 9 in volume 3 of the conference proceedings, and currently available via `http://cemmap.ifs.org.uk/`.

Andersen, T., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006). Volatility and correlation forecasting. In G. Elliott, C. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting, Volume 1*. Amsterdam: North Holland.

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*, 885–905.

Andersen, T. G., T. Bollerslev, F. S. Diebold, and P. Labys (2001). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association 96*, 42–55. Correction published in 2003, volume 98, page 501.

Andersen, T. G., T. Bollerslev, F. S. Diebold, and P. Labys (2003). Modeling and Forecasting Realized Volatility. *Econometrica 72*, 579–625.

Andersen, T. G., T. Bollerslev, and S. Lange (1999). Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon. *Journal of Empirical Finance 6*, 457–477.

Andersen, T. G., T. Bollerslev, and N. Meddahi (2005). Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities. *Econometrica 73*, 279–296.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253–280.

Bauwens, L., D. Rime, and G. Sucarrat (2006). Exchange Rate Volatility and the Mixture of Distribution Hypothesis. *Empirical Economics 30*, 889–911.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 31*, 307–327.

Bollerslev, T. and J. Wooldridge (1992). Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances. *Econometric Reviews 11*, 143–172.

Campos, J., N. R. Ericsson, and D. F. Hendry (2005). General-to-Specific Modeling: An Overview and Selected Bibliography. In J. Campos, D. F. Hendry, and N. R. Ericsson (Eds.), *General-to-Specific Modeling, Volume 1*. Cheltenham: Edward Elgar Publishing.

Clements, M. P. and D. F. Hendry (2005). Evaluating a Model by Forecast Performance. *Oxford Bulletin of Economics and Statistics 67*, 931–956.

Diebold, F., T. Gunther, and A. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review 39*, 863–883.

Doornik, J. (2006). *Ox: An Object Oriented Matrix Programming Language*. London: Timberlake Consultants Ltd. 5th. Edition.

Doornik, J. (2008). Autometrics. Unpublished manuscript, Economics Department, University of Oxford.

Doornik, J. A. and D. F. Hendry (2007). *Empirical Econometric Modelling - PcGive 12: Volume I*. London: Timberlake Consultants Ltd.

Engle, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations. *Econometrica 50*, 987–1008.

Engle, R., C.-H. Hong, A. Kane, and J. Noh (1993). Arbitrage Valuation of Variance Forecasts with Simulated Options. In D. Chance and R. Trippi (Eds.), *Advances in Futures and Options Research*. Greenwich, Connecticut: JAI Press.

Evans, M. D. and R. K. Lyons (2002). Order flow and exchange rate dynamics. *Journal of Political Economy 110*, 170–180.

Geweke, J. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews 5*, 57–61.

Gilbert, C. L. (1989). LSE and the British Approach to Time Series Econometrics. *Oxford Economic Papers 41*, 108–128.

Gilbert, C. L. (1990). Professor Hendry's Econometric Methodology. In C. W. Granger (Ed.), *Modelling Economic Series*. Oxford: Oxford University Press. Earlier publised in Oxford Bulletin of Economics and Statistics 48 (1986), pp. 283-307.

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*, 629–645.

Granger, C. W. and A. Timmermann (1999). Data mining with local model specification uncertainty: a discussion of Hoover and Perez. *Econometrics Journal 2*, 220–225.

Hansen, P. R. (2005). A Test for Superior Predictive Ability. *Journal of Business and Economic Statistics 23*, 365–380.

Hansen, P. R. and A. Lunde (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics 20*, 873–889.

Hansen, P. R. and A. Lunde (2006). Consistent ranking of volatility models. *Journal of Econometrics 131*, 97–121.

Hansen, P. R. and A. Lunde (2007). MULCOM 1.00. Econometric Toolkit for Multiple Comparisons. `http://www.asb.dk/~alunde/mulcom/mulcom.htm`.

Harvey, D., S. Leybourne, and P. Newbold (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting 23*, 801–824.

Hendry, D. F. (1993). *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (2003). J. Denis Sargan and the Origins of LSE Econometric Methodology. *Econometric Theory 19*, 457–480.

Hendry, D. F., S. Johansen, and C. Santos (2007). Automatic selection of indicators in a fully saturated regression. *Computational Statistics*. DOI 10.1007/s00180-007-0054-z.

Hendry, D. F. and H.-M. Krolzig (1999). Improving on 'Data Mining Reconsidered' by K.D. Hoover and S.J. Perez. *Econometrics Journal 2*, 202–219.

Hendry, D. F. and H.-M. Krolzig (2001). *Automatic Econometric Model Selection using PcGets*. London: Timberlake Consultants Press.

Hendry, D. F. and H.-M. Krolzig (2005). The Properties of Automatic Gets Modelling. *Economic Journal 115*, C32–C61.

Hendry, D. F. and J.-F. Richard (1990). On the Formulation of Empirical Models in Dynamic Econometrics. In C. W. Granger (Ed.), *Modelling Economic Series*. Oxford: Oxford University Press. Earlier published in Journal of Econometrics 20 (1982), pp. 3-33.

Hoover, K. D. and S. J. Perez (1999). Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search. *Econometrics Journal 2*, 167–191.

Hull, J. C. (2000). *Options, Futures and Other Derivatives*. London: Prentice Hall International. Fourth Edition.

Jansen, E. (2002). Statistical Issues in Macroeconomic Modelling. *Scandinavian Journal of Statistics 29*, pp. 193–217.

Jarque, C. and A. Bera (1980). Efficient Tests for Normality, Homoskedasticity, and Serial Independence of Regression Residuals. *Economics Letters 6*, 255–259.

Jorion, P. (1995). Predicting Volatility in the Foreign Exchange Markets. *The Journal of Finance 50*, 507–528.

Ljung, G. and G. Box (1979). On a Measure of Lack of Fit in Time Series Models. *Biometrika 66*, 265–270.

Lopez, J. A. (2001). Evaluating the Predictive Accuracy of Volatility Models. *Journal of Forecasting 20*, 87–109.

Lovell, M. C. (1983). Data Mining. *The Review of Economics and Statistics 65*, 1–12.

McAleer, M. (2005). Automated inference and learning in modeling financial volatility. *Econometric Theory 21*, 232–261.

Mincer, J. and V. Zarnowitz (1969). The Evaluation of Economic Forecasts. In J. Zarnowitz (Ed.), *Economic Forecasts and Expectations*. New York: National Bureau of Economic Research.

Mizon, G. (1995). Progressive Modeling of Macroeconomic Time Series: The LSE Methodology. In K. D. Hoover (Ed.), *Macroeconometrics. Developments, Tensions and Prospects*. Kluwer Academic Publishers.

Nelson, D. B. (1991). Conditional Heteroscedasticity in Asset Returns: A New Approach. *Econometrica 51*, 485–505.

Pantula, S. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews 5*, 71–73.

Patton, A. J. (2007). Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies. Available as `http://www.economics.ox.ac.uk/members/andrew.patton/patton_robust_aug07.pdf`.

Poon, S.-H. and C. W. J. Granger (2003). Forecasting in Financial Markets: A Review. *Journal of Economic Literature*, 478–539.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Sucarrat, G. (2006). *Essays in the Study and Modelling of Exchange Rate Volatility*. Ph. D. thesis, Department of Economics, Université catolique de Louvain, Louvain la Neuve, Belgium. Thesis Number 508/2006. Available as `http://edoc.bib.ucl.ac.be:81/ETD-db/collection/available/BelnUcetd-09152006-141220/unrestricted/these_sucarrat.pdf`.

Sucarrat, G. (2007). Econometric Reduction Theory and Philosophy: The First Stage in Hendry's Reduction Theory Revisited. Available as `http://www.eco.uc3m.es/sucarrat/research/hendry.pdf`.

Sucarrat, G. (2008). Forecast Evaluation of Explanatory Models of Financial Variability. Available as `http://www.eco.uc3m.es/sucarrat/research/volamodeval.pdf`.

West, K., H. Edison, and D. Cho (1993). A Utility Based Comparison of Some Models for Exchange Rate Volatility. *Journal of International Economics 35*, 23–45.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. *Econometrica 48*, 817–838.

White, H. (1990). A consistent model selection procedure based on m-testing. In C. W. Granger (Ed.), *Modelling Economic Series: Readings in Econometric Methodology*. Oxford: Clarendo Press.

White, H. (2000). A Reality Check for Data Snooping. *Econometrica 68*, 1097–1126.

Appendix: Data transformations and sources

The data transformations are undertaken in Ox (Doornik 2006), R (The R Development Core Team 2007) and EViews.

S_t BID NOK/1EUR closing value of the last trading day of week t. Before 1.1.1999 the BID NOK/1EUR rate is obtained by the formula BID NOK/100DEM $\times$ 0.0195583, where 0.0195583 is the official DEM/1EUR conversion rate 1.95583 DEM = 1 EUR divided by 100. The source of the BID NOK/100DEM series is Olsen Financial Technologies and the source of the BID NOK/1EUR series is Reuters.

r_t $(\log S_t - \log S_{t-1}) \times 100$

V_t^w $\{\{\log[S_t + I(S_t = S_{t-1}) \times 0.0009] - \log(S_{t-1})\} \times 100\}^2$. $I(S_t = S_{t-1})$ is an indicator function equal to 1 if $S_t = S_{t-1}$ and 0 otherwise, and $S_t = S_{t-1}$ occurs for $t = 10/6/1994$, $t = 19/8/1994$ and $t = 17/2/2000$.

v_t^w $\log V_t^w$

V_t^r $\sum_n [\log(S_n^m / S_{n-1}^m) \times 100]^2$, where $n = 1(t), 2(t), ..., N(t)$ and $1(t) - 1 = N(t-1)$ (realised weekly volatility). The number $N(t)$ is the number of 30-minute intervals during the week, and $S_{n(t)}$ is the mid-quote at the end of interval $n(t)$ in week t. The source of the raw-data is Olsen Financial Technologies.

M_t BID USD/EUR closing value of the last trading day of week t. Before 1.1.1999 the BID USD/EUR rate is obtained with the formula 1.95583/(BID DEM/USD). The source of the BID DEM/USD and BID USD/EUR series is Reuters.

m_t $\log M_t$

M_t^w $\{\{\log[M_t + I(M_t = M_{t-1}) \times 0.0009] - \log(M_{t-1})\} \times 100\}^2$. $I(M_t = M_{t-1})$ is an indicator function equal to 1 if $M_t = M_{t-1}$ and 0 otherwise.

m_t^w $\log M_t^w$

Q_t Weekly number of NOK/EUR quotes (NOK/100DEM before 1.1.1999). The underlying data is a daily series from Olsen Financial Technologies, and the weekly values are obtained by summing the values of the week.

q_t $\log Q_t$. This series is "synthetic" in that it has been adjusted for changes in the underlying quote collection methodology at Olsen Financial Technologies. More precisely q_t has been generated under the assumption that Δq_t is equal to zero in the weeks containing Friday 17 August 2001 and Friday 5 September 2003, respectively. In the first week the underlying feed was changed from Reuters to Tenfore, and on the second a feed from Oanda was added.

Δq_t $q_t - q_{t-1}$. The values of this series has been set to zero in the weeks containing Friday 24 August 2001 and Friday 5 September 2003, respectively.

O_t Closing value of the Brent Blend spot oil price in USD per barrel in the last trading day of week t. The untransformed series is Bank of Norway database series D2001712.

o_t $\log O_t$

O_t^w $\{\{\log[O_t + I(O_t = O_{t-1}) \times 0.009] - \log(O_{t-1})\} \times 100\}^2$. $I(O_t = O_{t-1})$ is an indicator function equal to 1 if $O_t = O_{t-1}$ and 0 otherwise, and $O_t = O_{t-1}$ occurs three times, for $t = 1/7/1994$, $t = 13/10/1995$ and $t = 25/7/1997$.

o_t^w $\log O_t^w$

X_t Closing value of the main index of the Norwegian Stock Exchange (TOTX) in the last trading day of week t. The source of the daily untransformed series is EcoWin series ew:nor15565.

x_t $\log X_t$

X_t^w $\{[\log(X_t/X_{t-1})] \times 100\}^2$. $X_t = X_{t-1}$ does not occur for this series.

x_t^w $\log X_t^w$

U_t Closing value of the composite index of the New York Stock Exchange (the NYSE index) in the last trading day of week t. The source of the daily untransformed series is EcoWin series ew:usa15540.

U_t^w $\{[\log(U_t/U_{t-1})] \times 100\}^2$. $U_t = U_{t-1}$ does not occur for this series.

u_t^w $\log U_t^w$

IR_t^{emu} Average of closing values of the 3-month market interest rates of the European Monetary Union (EMU) countries in the last trading day of week t. The source of the daily untransformed series is EcoWin series ew:emu36103.

ir_t^{emu} $(\Delta IR_t^{emu})^2$.

F_t The Norwegian central bank's main policy interest-rate, the so-called "folio", at the end of the last trading day of week t. The source of the untransformed daily series is Norges Bank's webpages.

f_t^a $|\Delta F_t| \times I_a$, where I_a is an indicator function equal to 1 in the period 1 January 1999 - Friday 30 March 2001 and 0 elsewhere

f_t^b $|\Delta F_t| \times I_b$, where I_b is an indicator function equal to 1 after Friday 30 March 2001 and 0 before

id_t Russian moratorium impulse dummy, equal to 1 in the week containing Friday 28 August 1998 and 0 elsewhere.

sd_t Step dummy, equal to 0 before 1997 and 1 thereafter.

ia_t Skewness term, equal to 1 when $r_t > 0$ and 0 otherwise.

h_{lt} $l = 1, 2, \ldots, 8$. Holiday variables with values equal to the number of official Norwegian holidays that fall on weekdays. For example, if 1 January falls on a Saturday then h_{1t} is equal to 0, whereas if 1 January falls on a Monday, then h_{1t} is equal to 1. h_{2t} is associated with Maundy Thursday and Good Friday and thus always equal to 2, h_{3t} with Easter Monday and thus always equal to 1, h_{4t} with Labour Day (1 May), h_{5t} with the Norwegian national day (17 May), h_{6t} with Ascension Day, h_{7t} with Whit Monday and h_{8t} with Christmas (Christmas Day and Boxing Day). Source: Http://www.timeanddate.com.

Table 1: Estimation results of GETS EMOV1 and GETS EMOV3

Regressor	(13)		(21)	
	Est.	*Pval.*	*Est.*	*Pval.*
const.	-2.922	0.00	-2.966	0.00
$v^w_{t-2} + v^w_{t-3}$	0.080	0.00	0.080	0.00
Δq_t	1.034	0.00		
m^w_t	0.072	0.09		
$x^w_t + u^w_t$	0.117	0.00		
f^b_t	3.657	0.00		
ir^{emu}_t	4.395	0.05		
sd_t	1.170	0.00	1.429	0.00
R^2	0.20		0.13	
AR_{1-10}	3.08	0.98	3.74	0.96
$ARCH_{1-10}$	7.01	0.72	9.88	0.36
$Het.$	13.28	0.43	0.61	0.74
$Hetero.$	18.99	0.98	6.74	0.15
JB	127.81	0.00	117.04	0.00
$Obs.$	569		569	

Note: The estimation sample is 8 January 1993 to 26 December 2003 and computations are in EViews 5.1 with OLS estimation. Standard errors are non-robust, *Pval* stands for p-value and corresponds to a two-sided test with zero as null, AR_{1-10} is the χ^2 version of the Lagrange-multiplier test for serially correlated residuals up to lag 10, $ARCH_{1-10}$ is the χ^2 version of the Lagrange-multiplier test for serially correlated squared residuals up to lag 10, $Het.$ and $Hetero.$ are the χ^2 versions of White's (1980) heteroscedasticity tests without and with cross products, respectively, JB is the Jarque and Bera (1980) test for non-normality in the residuals, and *Obs.* is the number of non-missing observations.

Table 2: Estimation results of the QGETS GARCH models

Regressor	(14)		(15)	
	Est.	*Pval.*	*Est.*	*Pval.*
Δm_t	0.158	0.00	0.139	0.00
Δo_t	-1.464	0.00	-1.130	0.01
Δx_t			-0.020	0.06
ΔIR_t^{emu}	0.550	0.00	0.531	0.00
ΔF_t^b	-1.953	0.02	-2.073	0.02
var.const.	0.006	0.07	-0.387	0.00
e_{t-1}^2	0.150	0.00		
$\lvert\frac{e_{t-1}}{\sigma_{t-1}}\rvert$			0.269	0.00
σ_{t-1}^2	0.852	0.00		
$\log \sigma_{t-1}^2$			0.925	0.00
Δq_t	0.083	0.00	1.249	0.00
sd_t			0.146	0.01
R^2	0.14		0.14	
LogL.	-505.28		-477.38	
AR_{1-10}	11.34	0.33	11.32	0.33
$ARCH_{1-10}$	4.78	0.91	9.78	0.46
JB	206.87	0.00	70.90	0.00
Obs.	572		572	

Note: The estimation sample is 8 January 1993 to 26 December 2003 and computations are in EViews 5.1 with robust standard errors of the Bollerslev and Wooldridge (1992) type. *Pval* stands for p-value and corresponds to a two-sided test with zero as null, *LogL* stands for Gaussian log-likelihood, AR_{1-10} is the Ljung and Box (1979) test for serial correlation in the standardised residuals up to lag 10, $ARCH_{1-10}$ is the Ljung and Box (1979) test for serial correlation in the squared standardised residuals up to lag 10, JB is the Jarque and Bera (1980) test for non-normality in the standardised residuals, and *Obs.* is the number of non-missing observations.

Table 3: Estimation results and diagnostics of the benchmark models

Regressor	(22)		(23)		(24)		(25)	
	Est.	*Pval.*	*Est.*	*Pval.*	*Est.*	*Pval.*	*Est.*	*Pval.*
var.const.	0.681	0.00			0.006	0.14	-0.223	0.00
e_{t-1}^2			0.060	-	0.146	0.01		
$\lvert\frac{e_{t-1}}{\sigma_{t-1}}\rvert$							0.293	0.00
σ_{t-1}^2			0.940	-	0.867	0.00		
$\log \sigma_{t-1}^2$							0.982	0.00
LogL.	-370.34		-741.60		-580.96		-577.07	
AR_{1-10}	8.94	0.54	12.38	0.26	11.43	0.32	12.48	0.25
$ARCH_{1-10}$	9.49	0.49	18.95	0.04	4.97	0.89	6.26	0.79
JB	87.85	0.00	868.00	0.00	635.57	0.00	394.53	0.00
Obs.	261		572		572		572	

Note: The estimation sample is 1 January 1999 to 26 December 2003 for (22), and 8 January 1993 to 26 December 2003 for (23), (24) and (25). Computations are in G@RCH 4.0 and EViews 6.0 with robust standard errors of the Bollerslev and Wooldridge (1992) type when applicable. Otherwise see table 2.

Table 4: Mincer-Zarnowitz regressions of r_t^2 on a constant and 1-step out-of-sample forecasts ($K = 61$)

	Model	a	b	R^2	AR_1	$Wald$
GETS ex post:	GETS EMOV1	-0.10	1.26	0.25	0.23	0.38
		[0.78]	[0.03]		[0.63]	[0.83]
	QGETS GARCH1	0.86	-0.02	0.00	0.39	4.42
		[0.05]	[0.97]		[0.53]	[0.11]
	QGETS EGARCH1	0.33	0.40	0.04	0.00	7.07
		[0.24]	[0.14]		[0.97]	[0.03]
GETS ex ante:	GETS EMOV2	-0.31	1.64	0.03	0.65	0.61
		[0.79]	[0.36]		[0.42]	[0.74]
	GETS EMOV3	-0.21	1.20	0.03	0.65	0.17
		[0.84]	[0.36]		[0.42]	[0.92]
	QGETS GARCH2	0.99	-0.17	0.00	0.39	7.46
		[0.02]	[0.69]		[0.53]	[0.02]
	QGETS EGARCH2	0.96	-0.13	0.00	0.47	15.24
		[0.03]	[0.71]		[0.49]	[0.00]
Benchmark:	Historical	-	1.24	0.00	0.38	0.73
			[0.00]		[0.54]	[0.39]
	RiskMetrics	0.92	-0.06	0.00	0.39	13.07
		[0.20]	[0.91]		[0.53]	[0.00]
	GARCH3	0.80	0.05	0.00	0.39	6.98
		[0.04]	[0.90]		[0.53]	[0.03]
	EGARCH3	0.60	0.19	0.00	0.34	8.29
		[0.21]	[0.60]		[0.56]	[0.02]
	RV	-0.11	0.83	0.14	0.02	4.13
		[0.79]	[0.07]		[0.90]	[0.13]

Note: The Mincer-Zarnowitz regression is given by $r_t^2 = a + b\hat{V} + u_t$, where $\hat{V}$ is the forecast of r_t^2. The columns $\hat{a}$ and $\hat{b}$ contain OLS estimated coefficient estimates of a and b, and the associated p-values in square brackets of a two-sided test with zero as the null hypothesis. The column R^2 contains the squared multiple correlation coefficient of the regression, the AR_1 column contains a $\chi^2(1)$ distributed LM test statistic for first order serial correlation in $\{\hat{u}_t\}$ with associated p-value, and the $Wald$ column contains a $\chi^2(2)$ distributed test statistic of the joint restriction $a = 0, b = 1$ with associated p-value. (In the case of Historical only the restriction $b = 1$ is tested.) The estimate of the variance-covariance matrix is of the White (1980) type.

Table 5: MSE forecast statistics and tests for greater accuracy than Historical

	Model	1-week	2-week	3-week	6-week	12-week
GETS ex post:	GETS EMOV1	1.605	1.631	1.106	1.085	1.136
	QGETS GARCH1	2.220	2.165	1.356	1.316	1.260
	QGETS EGARCH1	2.425	2.617	1.798	1.486	1.065
GETS ex ante:	GETS EMOV2	2.075	2.102	1.209	1.197	1.023
	GETS EMOV3	2.049	2.067	1.230	1.201	1.038
	QGETS GARCH2	2.239	2.162	1.346	1.293	1.231
	QGETS EGARCH2	2.472	2.528	1.503	1.235	1.094
Benchmark:	Historical	2.132	2.165	1.198	1.192	1.022
	RiskMetrics	2.517	2.504	1.762	1.772	1.758
	GARCH3	2.285	2.266	1.479	1.465	1.648
	EGARCH3	2.379	2.329	1.374	1.310	1.304
	RV	1.901	3.897	3.643	3.958	3.951
Forecast tests:	RC p-value	0.22	0.23	0.77	0.63	1.00
	SPA p-value	0.29	0.34	0.77	0.83	1.00

Note: The p-values of White's (2000) Reality Check (RC) and Hansen's (2005) Superior Predictive Ability (SPA) test are calculated using the Ox package SPA 2.02 (see Hansen and Lunde 2007) with 10 000 bootstraps and dependence parameter q equal to 0.5.

Table 6: One-sided p-values of the Modified Diebold-Mariano (MDM) test statistic of greater forecast accuracy than Historical in terms of MSE

	Model	1-week	2-week	3-week	6-week	12-week
GETS ex post:	GETS EMOV1	0.32	0.19	0.35	0.34	0.92
	QGETS GARCH1	0.60	0.50	0.85	0.83	0.88
	QGETS EGARCH1	0.60	0.85	0.85	0.89	0.58
GETS ex ante:	GETS EMOV2	0.38	0.27	0.65	0.67	0.53
	GETS EMOV3	0.41	0.31	0.69	0.62	0.66
	QGETS GARCH2	0.65	0.49	0.87	0.84	0.91
	QGETS EGARCH2	0.75	0.86	0.88	0.68	0.88
Benchmark:	Historical	—	—	—	—	—
	RiskMetrics	0.72	0.83	0.99	0.99	0.99
	GARCH3	0.64	0.69	0.91	0.90	0.96
	EGARCH3	0.66	0.85	0.94	0.91	0.98
	RV	0.42	0.96	1.00	1.00	1.00

Note: The p-values are computed using a Student's t distribution with 1 degree of freedom (DF) in the 1-week test, 2 DFs in the 2-week test, and so on.

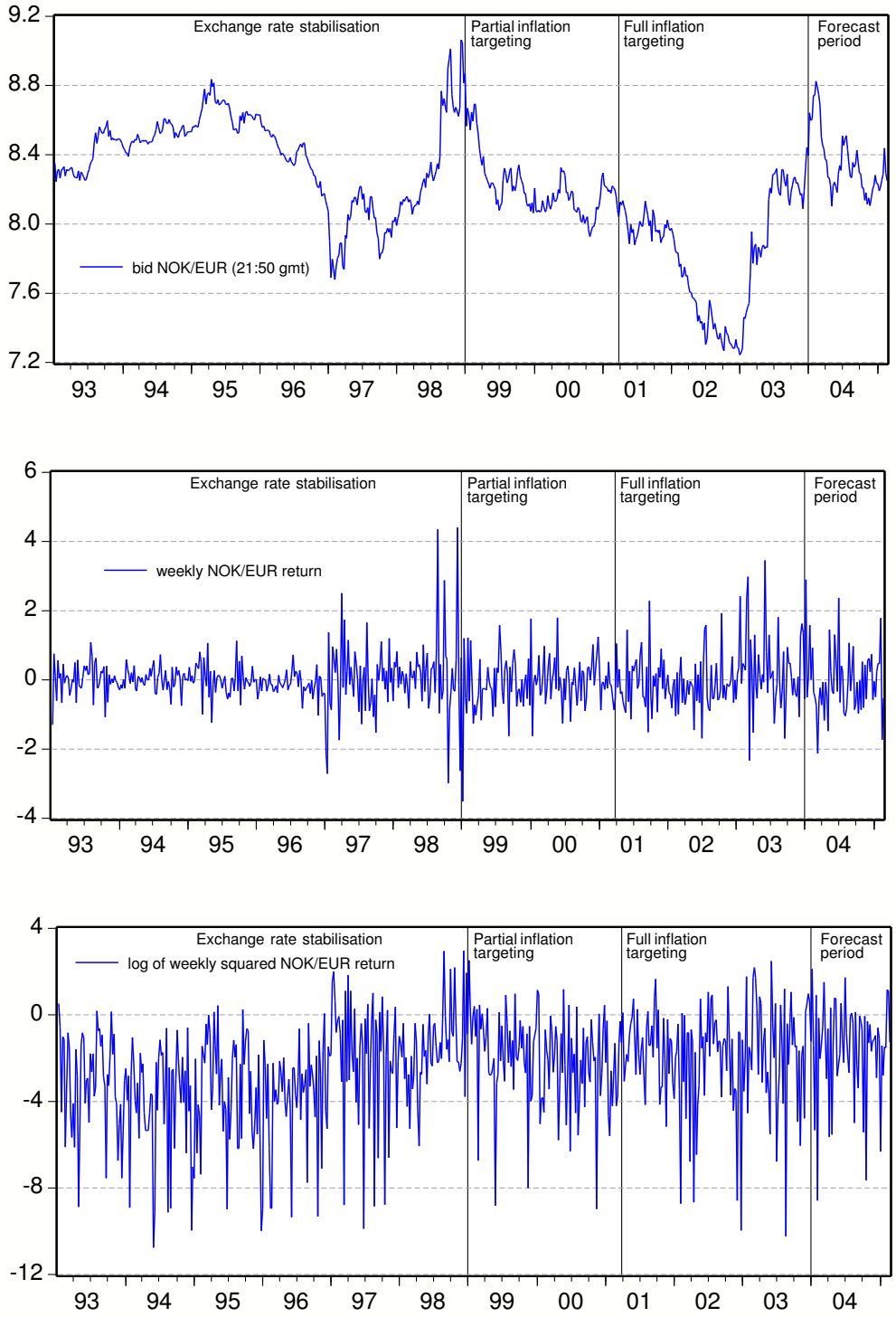

Figure 1: Bid NOK/EUR at 21:50 GMT in the last trading day of the week (denoted S_t in the text) in the upper graph, log-return r_t in the middle graph and log of r_t^2 in the bottom graph from 8 January 1993 to 25 February 2005.

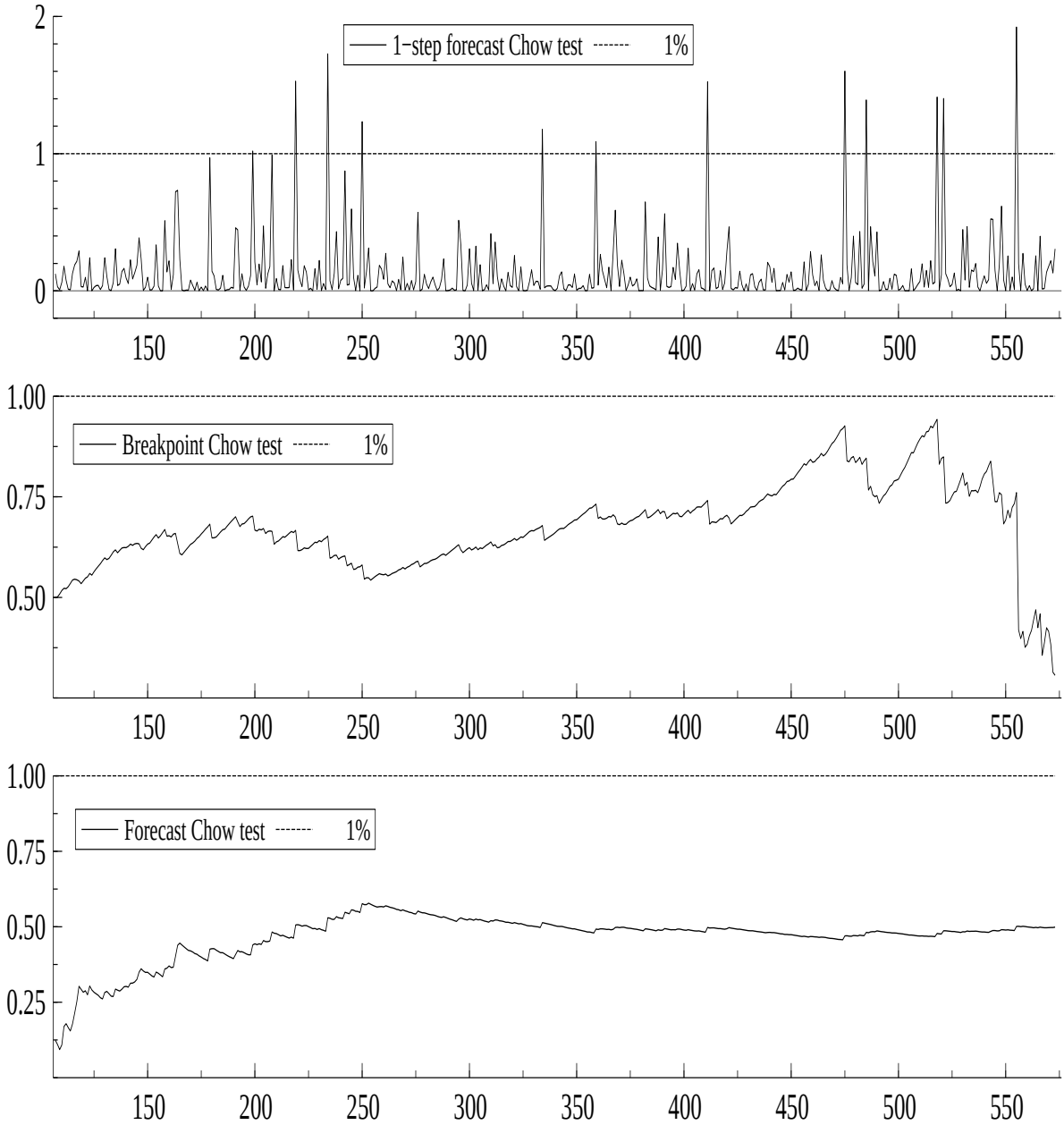

Figure 2: Recursive analysis of GUM EMOV1. Computations are in PcGive with OLS and initialisation at observation number 100.

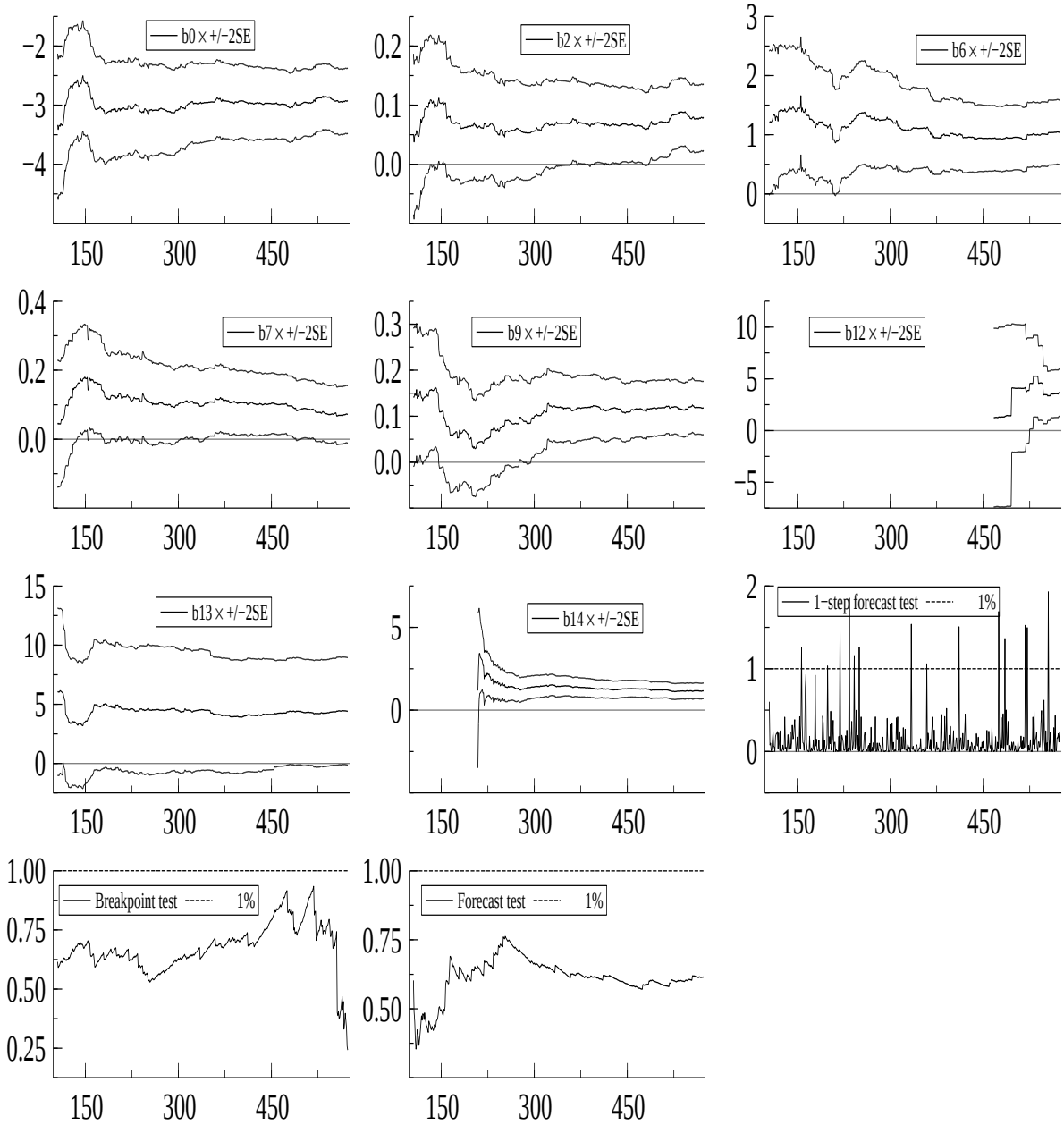

Figure 3: Recursive analysis of GETS EMOV1. Computations are in PcGive with OLS and initialisation at observation number 100.

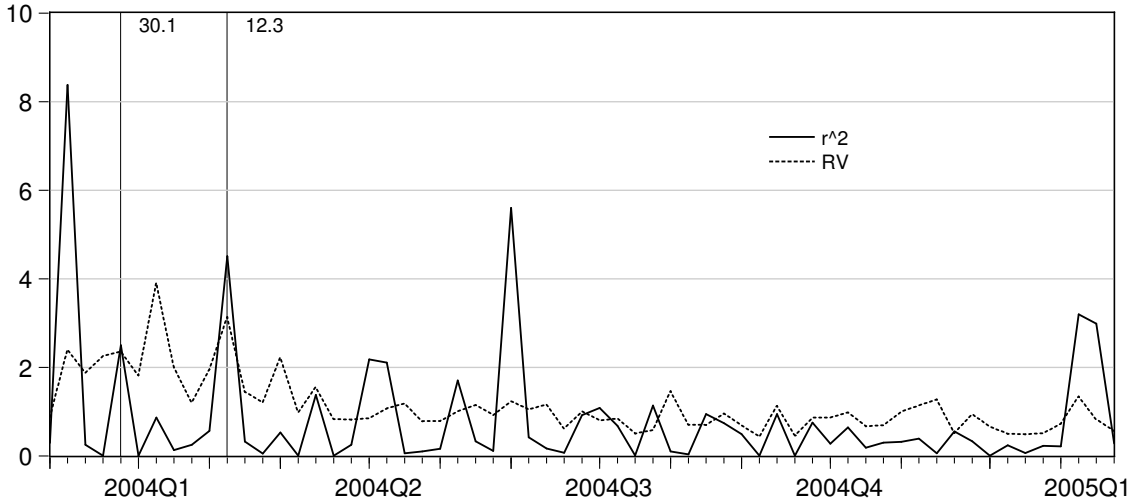

Figure 4: Out-of-sample trajectories of r_t^2 and RV. Vertical lines indicate weeks in which Norges Bank changed their main policy interest rate.

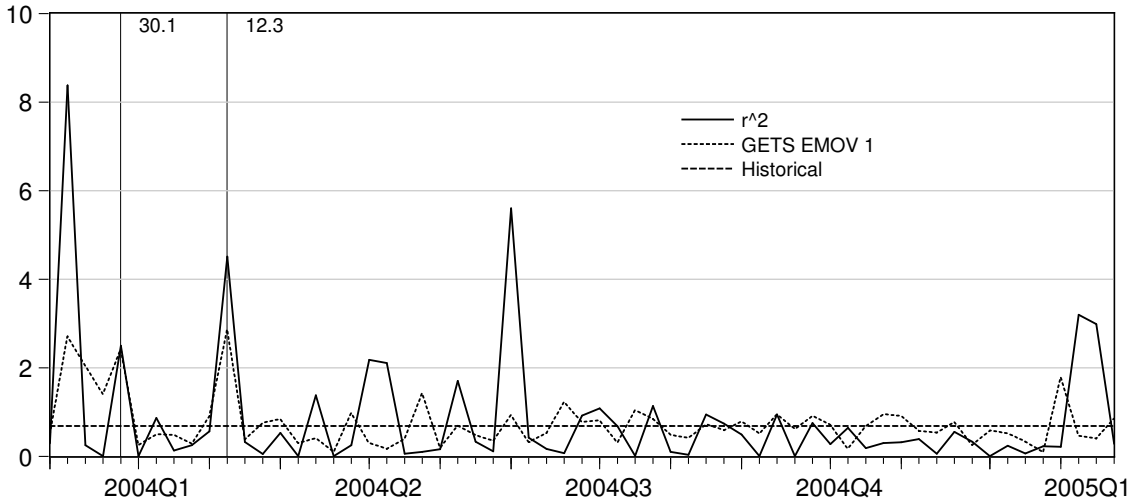

Figure 5: Out-of-sample trajectories of r_t^2, GETS EMOV1 and Historical. Vertical lines indicate weeks in which Norges Bank changed their main policy interest rate.