Computationally attractive stability tests for the efficient method of moments

Pieter J. van der Sluis*
Department of Actuarial Science and Econometrics
University of Amsterdam
and
Tinbergen Institute

August 4, 1997

Abstract

Estimation using simulation techniques may be very time consuming. Specification tests for structural stability often require more than one of such computationally demanding estimators. Typically one for the sample, one for the post-sample and one for the combination of sample and post-sample is required. This paper describes structural stability tests for use with the Efficient Method of Moments technique. Computationally attractive post-sample estimators and test-statistics for structural stability are proposed. These computationally attractive test-statistics are modifications of the Lagrange Multiplier, Likelihood Ratio and Wald tests for structural stability and of the Hansen-type test statistics for structural stability. The modification ensures the same asymptotic optimality properties against certain local alternatives as those based on efficient computationally intensive estimators for the post-sample. However no time consuming estimators are needed for the post-sample and for the combination of sample and post-sample. Evaluation of these tests has been performed in the context of a stochastic volatility model for the S&P500.

Acknowledgements

This paper was presented at *TI Econometrics* at Erasmus University Rotterdam, December 4th 1996, at the seventh meeting of EC^2 on *Simulation Methods in Econometrics* at the European University Institute in Florence, Italy, 12-14 December 1996 and at the North American Summer Meeting of the Econometric Society at the California Institute of Technology, Pasadena (CA), USA, 25-29 June 1997. The author thanks Peter Boswijk, Ron Gallant, Jesper Lund and Neil Shephard for helpful feedback. The author thanks Jurgen Doornik for feedback on issues on the Ox programming language. Financial support from the Fondazione Marchi-Firenze, the NWO and the Tinbergen Institute is gratefully acknowledged.

Keywords: Efficient Method of Moments; Heteroskedasticity; Parameter updates; Stochastic volatility models; Structural stability tests.

JEL-classification: C15, C52

*Address for correspondence: Pieter J. van der Sluis, University of Amsterdam, Department of Actuarial Science and Econometrics, Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands, phone +31 20 525 4257, fax +31 20 525 4349, email sluis@fee.uva.nl.

1 Introduction

Indirect inference, as described in Gourieroux *et al*. (1993) or the *Efficient Method of Moments* (EMM) as described in Gallant and Tauchen (1996), may be of great help in case the likelihood of a model is intractable. These techniques require a repeated evaluation of very complex functions under simulation and are therefore in general very time consuming. Consequently, computationally attractive *structural stability tests* and *updates* of the estimates may be of interest with the arrival of post-sample data. Specification tests for structural stability often require more than one of such computationally demanding estimators. Typically, one for the sample, one for the post-sample and one for the combination of sample and post-sample is required. In this paper computationally attractive *structural stability tests* and *updates* of the estimates for EMM are proposed. As will be motivated in the next section the author judges the EMM technique superior to the original indirect inference technique of Gourieroux *et al*. (1993) from a viewpoint of *efficiency* and *stability*. Therefore the findings are presented for EMM only. The same rationale can easily be transcribed to an indirect inference setting as well. Because of its lack of efficiency, it is doubtful whether for substantial problems satisfying results will be obtained with the original indirect inference approach.

The rationale for indirect inference techniques is that there is some *auxiliary model* which is relatively easy to estimate and is rich in the sense that it mimics the properties of the structural model to a certain degree. Moreover, to obtain the full maximum likelihood efficiency of EMM, it is required that the auxiliary model in some sense *embeds* the structural model, see Gallant & Long (1997), Gallant and Tauchen (1996) and Tauchen (1996) for details. This auxiliary model will be used to evaluate the data at hand. Next, data will be simulated from the original model. Finally the parameters are calibrated in a such a way that the *dynamic properties* of the auxiliary model under the real data and under the simulated data match.

Since auxiliary models will be in general easy to estimate, it is relatively easy to make an update of the auxiliary parameters with the arrival of new data. From this shift in the estimates of the auxiliary parameter estimates we may deduce a first order approximation for the shift in the structural parameters, yielding one-step linearized EMM estimators. More important than these first order approximations are test-statistics for structural stability that can be based on such first order approximations.

Several tests for structural stability have been proposed in the GMM literature: LM/LR/Wald type test for structural stability (Andrews and Fair (1988)), the Hansen J-test for structural stability (Ghysels and Hall (1990b)) and the Prediction tests (Ghysels & Hall (1990a)). In a GMM context each of these test-statistics has optimal local power in a certain relevant direction of misspecification. These test statistics are easily transcribed to an EMM/Indirect inference context. Their optimality properties also carry over.

With the exception of the prediction tests, which have been analysed in an EMM context in Van der Sluis (1997a)[1], all these tests statistics require very demanding EMM estimates for the structural parameters, other than that of the sample: an EMM estimate of the post-sample and an EMM estimate of the combination of sample and post-sample is required. In this paper simple modifications of the LM/LR/Wald test-statistics and of the J-test for structural stability, are proposed that require only root-n consistent estimators and still have the same optimal power in the same direction as the originals. Of course their small sample properties will differ.

[1] A different class of stability tests viz. *post sample prediction test* (PSP), for EMM was proposed in Van der Sluis (1997a). This test was first introduced for GMM by Ghysels and Hall (1990a) and Hofman and Pagan (1989). The nature of this PSP test is already computationally attractive, since only EMM estimates for the sample are required. Therefore there is no need to consider modifications of the PSP test. In particular the PSP test arises as a special case here.

These modifications hinge on work of Ahn (1995) in the field of GMM specification testing. In Ahn (1995) test statistics are designed for any *root-n consistent* estimators that have the same asymptotic power properties as those computed with GMM estimators. These results are transcribed to the EMM case. As a root-n consistent estimator for example a simple one-step linearized EMM estimator may be used. Naturally, the first-order approximations to the structural model are root-n consistent. In the application in this paper also other estimators, like Bayesian estimators which are computationally more attractive than the EMM estimator are plugged in.

The plan of the paper is as follows. Section 2 will provide a short overview of EMM. This section is mainly included to introduce some notation. For a better understanding of the method the reader is strongly recommended to consult Gallant & Tauchen (1996) and Tauchen (1996). Section 3 will introduce the first order EMM approximations to the estimators of the structural model. Section 4 contains the derived computationally attractive specification tests. Besides that, formulae are given for the exact test statistics and noncentrality parameters. In section 5 data from the S&P500 in the context of stochastic volatility models are considered. The breakpoint is set at Black Monday 1987. Instability is associated with different volatility regimes. Besides the first order approximations, in this context the Bayesian method of Kim *et al.* (1996) provides very fast and possibly accurate estimators. The modes of these Bayesian posteriors are taken as root-n consistent estimators *in the classical sense* (see e.g. Barndorff-Nielsen & Cox (1994)) and plugged into the test-statistics for structural stability. The outcomes are interpreted. Easily available estimators and tests for structural stability are very much of interest in this context, because stochastic volatility models can be used for the pricing of options, see Hull & White (1987). The parameters of the stochastic volatility models are important determinants in option pricing formulae. Since the estimation of a stochastic volatility model using EMM is very time consuming and the next step of determining the option prices is very time consuming as well, this is an important issue. Section 6 concludes.

2 Efficient Method of Moments

In the original indirect inference setup of Gourieroux *et al.* (1993) the connection between the auxiliary model and the structural model is established through the parameters of the auxiliary model. Gallant and Tauchen (1996) solve the efficiency problems indirect inference has by proposing the *efficient method of moments* (EMM) technique. Here the connection between the auxiliary model and the dynamic model is achieved by means of the scores of the auxiliary model, where strict guidelines are given for the choice of the auxiliary model. This method does not only yield efficient estimates but also provides a more stable set-up: scores are better comparable over different models than parameters are. In short the EMM method goes as follows: the sequence of densities for the structural model will be denoted

$$\{p_1(x_1 \mid \theta), \{p_t(y_t \mid x_t, \theta)\}_{t:=1}^{\infty}\}$$

The sequence of densities for the auxiliary process will be denoted as

$$\{f_1(w_1 \mid \beta), \{f_t(y_t \mid w_t, \beta)\}_{t:=1}^{\infty}\}$$

where x_t and w_t are observable endogenous variables. We impose assumptions 1 and 2 from Gallant & Long (1997) on the structural model. These are technical assumptions that imply standard properties of quasi maximum likelihood estimators and properties of estimators based on *Hermite expansions* which will be explained below. Let us define

$$m(\theta, \beta) := \int \int \frac{\partial}{\partial \beta} \ln f(y \mid w, \beta) p(y \mid x, \theta) dy p(x \mid \theta) dx$$

the expected score of the auxiliary model under the dynamic model. The expectation is written in integral form to anticipate on the fact the we will approximate this integral by standard Monte Carlo techniques. The simulation approach solely consists of calculating this function as

$$m_N(\theta, \beta) := \frac{1}{N} \sum_{\tau:=1}^{N} \frac{\partial}{\partial \beta} \ln f(y_\tau(\theta) \mid w_\tau(\theta), \beta)$$

Let n denote the sample size, the EMM estimator is defined as

$$\widehat{\theta}_n(\mathcal{I}_n) := \arg \min_{\theta \in \Theta} m_N^{'}(\theta, \widehat{\beta}_n)(\mathcal{I}_n)^{-1} m_N(\theta, \widehat{\beta}_n)$$

where $\mathcal{I}_n$ is a weighting matrix and $\widehat{\beta}_n$ denotes an estimator for the parameter of the auxiliary model. The optimal weighting matrix here is obviously

$$\mathcal{I}_0 = \lim_{n \to \infty} V_0[\frac{1}{\sqrt{n}} \sum_{t:=1}^{n} \{\frac{\partial}{\partial \beta} \ln f_t(y_t \mid w_t, \beta^*)\}]$$

where β^* is a (pseudo) true value.

With the theory of misspecified models (White (1994)) one can prove consistency for the parameters of the auxiliary model under several assumptions posed in Gallant & Tauchen (1996),

$$\lim_{n \to \infty} (\widehat{\beta}_n - \beta^*) = 0 \text{ a.s.}$$

and asymptotic normality

$$\sqrt{n}(\widehat{\beta}_n - \beta^*) \xrightarrow{d} N(0, (\mathcal{J}_0)^{-1}(\mathcal{I}_0)(\mathcal{J}_0)^{-1})$$

Here

$$\begin{aligned} \mathcal{I}_n &= V_0[\frac{1}{\sqrt{n}} \sum_{t:=1}^{n} (\frac{\partial}{\partial \beta} \ln f_t(\widetilde{y}_t \mid \widetilde{w}_t, \widehat{\beta}_n)] \\ \mathcal{J}_n &= -\frac{\partial}{\partial \beta} m_N^{'}(\theta_0, \widehat{\beta}_n) \end{aligned}$$

where θ_0 denotes the (pseudo) true value. Obviously,

$$\begin{aligned} \lim_{n \to \infty} \mathcal{I}_n &= \mathcal{I}_0 \\ \lim_{n \to \infty} \mathcal{J}_n &= \mathcal{J}_0 \end{aligned}$$

One can also prove for the scores

$$\sqrt{n} m_N(\theta_0, \widehat{\beta}_n) \xrightarrow{d} N(0, \mathcal{I}_0)$$

Hence consistency and asymptotic normality of the estimator of the structural parameters $\widehat{\theta}_n$ follow:

$$\sqrt{n}(\widehat{\theta}_n(\mathcal{I}_0) - \theta_0) \xrightarrow{d} N(0, [\mathcal{M}_0^{'}(\mathcal{I}_0)^{-1}\mathcal{M}_0]^{-1})$$

where $\mathcal{M}_0 := \frac{\partial}{\partial \theta'} m(\theta_0, \beta^*)$.

In order to obtain *full maximum likelihood efficiency* it is required that the auxiliary model in some sense embeds the structural model. The semi-nonparametric (SNP) density of Gallant & Nychka (1987)[2] may be a

[2] Building on earlier work of Phillips (1983). See also Fenton & Gallant (1996a+b) for recent results on SNP densities.

good choice, see Gallant & Tauchen (1996) and Gallant & Long (1997). The auxiliary model is built as follows. Let $y_t(\theta_0)$ denote the process under investigation, $\mu_t(\theta_0) := \mathsf{E}_{t-1}[y_t(\theta_0)]$, the conditional mean, $\sigma_t^2(\theta_0) := \mathsf{Var}_{t-1}[y_t(\theta_0) - \mu_t(\theta_0)]$ the conditional variance and $z_t(\theta_0) := [y_t(\theta_0) - \mu_t(\theta_0)] \cdot [\sigma_t^2(\theta_0)]^{-1/2}$ the standardized process. The SNP density now takes the following form

$$f(z_t;\theta) = \frac{1}{\sqrt{\sigma_t^2}} \frac{[P_K(z_t, x_t)]^2 \phi(z_t)}{\int [P_K(u, x_t)]^2 \phi(u) du}$$

where ϕ denotes the standard normal density, $x := (y_{t-1}, ..., y_{t-L})$ and the polynomials

$$P_K(z, x_t) := \sum_{i:=0}^{K_z} a_i(x_t) z^i := \sum_{i:=0}^{K_z} [\sum_{j:=0}^{K_x} a_{ij} x_t^j] z^i$$

A specific form for the polynomials is taken, namely orthogonal Hermite polynomials (see Gallant *et al*. (1991) and Andersen & Lund (1997)). Relevant formulae for the derivatives can be found in Abramowitz & Stegun (1972) and Fenton & Gallant (1996a). The model $\sigma_t^2(\theta)$ and $\mu_t(\theta)$ is chosen as a *leading term* in the Hermite expansion to relieve the expansion of some of its task, improving its small sample properties. We note that the SNP density offers a very convenient way of modelling. It omits the curse of dimensionality often encountered in the area of non parametric models while sharing many of the optimality properties of the Kernel estimators of Silverman (1986).

In this paper *structural stability* is of interest, therefore the following nomenclature is employed: $n_1$will denote the number of observations in the sample and n_2 will denote the number of observations in the post-sample. The auxiliary estimator that employs sample data only, will be denoted $\widehat{\beta}_{n_1}$. The auxiliary estimator that employs post-sample data only will be denoted $\widehat{\beta}_{n_2}$. Likewise the EMM estimators $\widehat{\theta}_{n_1}$ and $\widehat{\theta}_{n_2}$ are denoted. Furthermore estimators that are based on both sample and post-sample data are denoted $\widehat{\beta}_{n_1+n_2}$ and $\widehat{\theta}_{n_1+n_2}$.

In specification testing the following estimator for θ_0 will often be of (theoretical) interest, see e.g. Rothenberg (1973)

$$\widetilde{\theta}_{n_1+n_2} := \arg\min_{\theta \in \theta} [m_N'(\theta, \widehat{\beta}_{n_1}) \widehat{\mathcal{I}}_{n_1}^{-1} m_N(\theta, \widehat{\beta}_{n_1}) + m_N'(\theta, \widehat{\beta}_{n_2}) \widehat{\mathcal{I}}_{n_2}^{-1} m_N(\theta, \widehat{\beta}_{n_2})]$$

where $\widehat{\mathcal{I}}_{n_i}$ denotes a consistent estimator for the asymptotic covariance matrix of $m_N(\theta_0, \widehat{\beta}_{n_i})$, $\mathcal{I}_i$ for $i = 1, 2$.

In this paper we will take $p := \dim(\theta)$, the dimension of the parameters of the structural model, $q_1 := \dim(\beta_1)$, the dimension of the parameters of the auxiliary model for the sample and $q_2 := \dim(\beta_2)$, the dimension of the parameters of the auxiliary model for the post-sample. The number of moment conditions for the sample q_1 and for the post-sample q_2 will be determined using several criteria. For EMM, it is necessary that q_1 increases with n_1 and q_2 with n_2. Note in this respect the conceptual difference with GMM. It will automatically happen that q_1 and q_2 will increase with n_1 and n_2, respectively, using any of the model specification criteria such as the Akaike Information Criterion (AIC, Akaike (1973)), the Schwarz Criterion (BIC, Schwarz (1978)) or the Hannan-Quinn Criterion (HQC, Hannan & Quinn (1979) and Quinn (1980))[3]. The theory of model selection in the context of SNP models is not very well developed yet. Results in Eastwood (1991) may lead to

[3] AIC$= T \ln L - |\beta|$

BIC$= T \ln L - 0.5|\beta| \ln T$

HQC$= T \ln L - |\beta| \ln \ln T$

In this context, L denotes the loglikelihood of the auxiliary model in the optimum, T denotes the number of observations and $|\beta|$ the number of freely estimated auxiliary parameters. For some properties in the context of multivariate time-series see Lütkepohl (1990).

believe AIC is optimal in this case. However, as for multivariate ARMA models, the AIC may overfit the model to noise in the data so we may be better off by following the BIC or the HQC. The same findings were reported in Andersen & Lund (1997). In their seminal paper Gallant & Tauchen (1996) rely on the BIC in their applications. If testing by a battery of specification tests indicates that the BIC-optimal model is inadequate, see Bansal *et al.* (1995) and Tauchen *et al.* (1996), the model is further expanded.

3 Updating based on different auxiliary models for sample and post-sample

Say we have a sample of size n_1 and estimators $\widehat{\theta}_{n_1}$ and $\widehat{\beta}_{n_1}$ based on this sample at hand. The arrival of new data of size n_2 or the desire to apply tests for structural stability, may force us to redo the whole EMM estimation in order to obtain a $\widehat{\theta}_{n_2}$ for the post-sample or to obtain a $\widehat{\theta}_{n_1+n_2}$ or $\widetilde{\theta}_{n_1+n_2}$ for the combined data set. This may be very time consuming. However in general estimation of $\widehat{\beta}_{n_2}$ will not cost much too much computing time. Note that the auxiliary model need not be the same for sample and post-sample. At time n_1 we already have $m_N(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})$. Transcription of the one-step linearized GMM estimator yields a one-step linearized EMM estimator

$$\widehat{\theta}^*_{n_2} := \widehat{\theta}_{n_1} - \widehat{\mathcal{V}}_2^{-1}(\widehat{\theta}_{n_1})\mathcal{M}_2(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_2})^{'}(\mathcal{I}_{n_2})^{-1} m_N(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_2})$$

where $\widehat{\mathcal{V}}_i(\theta) := \mathcal{M}_i^{'}(\theta, \widehat{\beta}_{n_i})(\widehat{\mathcal{I}}_{n_i})^{-1}\mathcal{M}_i(\theta, \widehat{\beta}_{n_i})$.

Property 1 *Application of lemma 4 of Newey (1985) yields*

$$\sqrt{n_2}(\widehat{\theta}^*_{n_2} - \theta_2) = o_p(1)$$

Pooling of the estimators in an optimal way may be done in several different ways. The poolings in theorems 1 till 4 are all based on Taylor expansions around consistent estimators under the null. Theorem 1 is based on a Taylor expansion around $\widehat{\theta}_{n_1}$ and $\widehat{\theta}_{n_2}$, theorem 2 around $\widehat{\theta}_{n_1}$, theorem 3 around $\widehat{\theta}_{n_1}$ and $\widehat{\theta}^*_{n_2}$ and theorem 4 around $\widehat{\theta}^*_{n_2}$.

Theorem 1 *The pooled estimator given by*

$$\begin{aligned}\widehat{\theta}^{I}_{n_1+n_2} &= [\widehat{\mathcal{V}}_1(\widehat{\theta}_{n_1}) + \widehat{\mathcal{V}}_2(\widehat{\theta}_{n_2})]^{-1}[\widehat{\mathcal{V}}_1(\widehat{\theta}_{n_1})\widehat{\theta}_{n_1} + \widehat{\mathcal{V}}_2(\widehat{\theta}_{n_2})\widehat{\theta}_{n_2} \\ &-\mathcal{M}_1^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})(\mathcal{I}_1)^{-1} m(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1}) - \mathcal{M}_2^{'}(\widehat{\theta}_{n_2}, \widehat{\beta}_{n_2})(\mathcal{I}_2)^{-1} m(\widehat{\theta}_{n_2}, \widehat{\beta}_{n_2})]\end{aligned}$$

is asymptotically equivalent with $\widetilde{\theta}_{n_1+n_2}$.

Theorem 2 *The estimator*

$$\begin{aligned}\widehat{\theta}^{II}_{n_1+n_2} &= [\widehat{\mathcal{V}}_1(\widehat{\theta}_{n_1}) + \widehat{\mathcal{V}}_2(\widehat{\theta}_{n_1})]^{-1}[\widehat{\mathcal{V}}_1(\widehat{\theta}_{n_1})\widehat{\theta}_{n_1} + \widehat{\mathcal{V}}_2(\widehat{\theta}_{n_1})\widehat{\theta}_{n_1} \\ &-\mathcal{M}_1^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})(\mathcal{I}_1)^{-1} m(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1}) - \mathcal{M}_2^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_2})(\mathcal{I}_2)^{-1} m(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_2})]\end{aligned}$$

is asymptotically equivalent with $\widetilde{\theta}_{n_1+n_2}$.

Theorem 3 *The estimator*

$$\begin{aligned}\widehat{\theta}^{III}_{n_1+n_2} &= [\widehat{\mathcal{V}}_1(\widehat{\theta}_{n_1}) + \widehat{\mathcal{V}}_2(\widehat{\theta}^*_{n_2})]^{-1}[\widehat{\mathcal{V}}_1(\widehat{\theta}_{n_1})\widehat{\theta}_{n_1} + \widehat{\mathcal{V}}_2(\widehat{\theta}^*_{n_2})\widehat{\theta}^*_{n_2} \\ &-\mathcal{M}_1^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})(\mathcal{I}_1)^{-1} m(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1}) - \mathcal{M}_2^{'}(\widehat{\theta}^*_{n_2}, \widehat{\beta}_{n_2})(\mathcal{I}_2)^{-1} m(\widehat{\theta}^*_{n_2}, \widehat{\beta}_{n_2})]\end{aligned}$$

is asymptotically equivalent with $\widetilde{\theta}_{n_1+n_2}$

Theorem 4 *The estimator*

$$\widehat{\theta}^{IV}_{n_1+n_2} := \widehat{\theta}^*_{n_2} - [\widehat{\mathcal{V}}_1(\widehat{\theta}^*_{n_2}) + \widehat{\mathcal{V}}_2(\widehat{\theta}^*_{n_2})]^{-1}[\mathcal{M}_1'(\widehat{\theta}^*_{n_2}, \widehat{\beta}_{n_1})(\mathcal{I}_1)^{-1}m(\widehat{\theta}^*_{n_2}, \widehat{\beta}_{n_1}) + \mathcal{M}_2'(\widehat{\theta}^*_{n_2}, \widehat{\beta}_{n_2})(\mathcal{I}_2)^{-1}m(\widehat{\theta}^*_{n_2}, \widehat{\beta}_{n_2})]$$

is asymptotically equivalent with $\widetilde{\theta}_{n_1+n_2}$

A conceptually different update strategy is based on the specification (via AIC, HQC or BIC) of an overall auxiliary model for the sample and post-sample. We are thus interested in the estimator

$$\widehat{\theta}_{n_1+n_2} = \arg\min_{\theta \in \Theta} m'(\theta, \widehat{\beta}_{n_1+n_2})\mathcal{I}_{1+2}^{-1} m(\theta, \widehat{\beta}_{n_1+n_2})$$

Taylor expansion around $\widehat{\theta}_{n_1}$ gives

$$\widehat{\theta}^{\mathrm{V}}_{n_1+n_2} = \arg\min_{\theta \in \Theta} [m'(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1+n_2}) + \mathcal{M}_{1+2}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1+n_2})(\theta - \widehat{\theta}_{n_1})]' \mathcal{I}_{1+2}^{-1} [m'(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1+n_2}) + \mathcal{M}_{1+2}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1+n_2})(\theta - \widehat{\theta}_{n_1})]$$

This comes down to ordinary GLS and therefore an explicit expression for $\widehat{\theta}^{\mathrm{V}}_{n_1+n_2}$ is found

$$\widehat{\theta}^{\mathrm{V}}_{n_1+n_2} = \widehat{\theta}_{n_1} - [\mathcal{M}_{1+2}'(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1+n_2})\mathcal{I}_{1+2}^{-1}\mathcal{M}_{1+2}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1+n_2})]^{-1}\mathcal{M}_{1+2}'(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1+n_2})\mathcal{I}_{1+2}^{-1} m(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1+n_2})$$

The usual arguments provide asymptotic equivalence of $\widehat{\theta}^{\mathrm{V}}_{n_1+n_2}$ and $\widehat{\theta}_{n_1+n_2}$.

Intuitively, an estimator based on an estimate of the auxiliary parameter of the combined sample and the post-sample, such as $\widehat{\theta}^{\mathrm{V}}_{n_1+n_2}$, will give a better approximation to $\widehat{\theta}_{n_1+n_2}$ than an approximation that is based on auxiliary estimators of the separate sample and post-sample such as $\widehat{\theta}^{\mathrm{III}}_{n_1+n_2}$. Reversely, tests for structural stability based on estimators of the auxiliary model for the separate sample and post-sample will have higher power than tests that are solely based on auxiliary estimators for the combined sample and post-sample. Future research is needed in order to substantiate this intuition.

4 Stability testing

Following, among others, Newey (1985), Ghysels and Hall (1990b) and Ahn (1995) for GMM, we identify the following individual null hypotheses for the EMM case.

$$\begin{aligned} \mathsf{H}_0^1 &: \quad m(\theta_0, \beta_1) = 0, \ \forall\, t \in T_1 \\ \mathsf{H}_0^2 &: \quad m(\overline{\theta}_0, \beta_2) = 0, \ \forall\, t \in T_2 \\ \mathsf{H}_0^3 &: \quad \theta_0 = \overline{\theta}_0 \end{aligned}$$

where the function m is the expected score of the auxiliary model under the dynamic model as defined in the previous section. The *stability* hypothesis is defined as

$$\mathsf{H}_0^s : \mathsf{H}_0^1, \mathsf{H}_0^2 \text{ and } \mathsf{H}_0^3 \text{ hold}$$

The asymptotic power of the test depends on which of the individual null-hypotheses is violated. Define the following alternative hypotheses

$$\begin{aligned} \mathsf{H}_A^A &: \quad \text{At least one of } \mathsf{H}_0^1, \mathsf{H}_0^2 \text{ and } \mathsf{H}_0^3 \text{ does not hold} \\ \mathsf{H}_A^B &: \quad \mathsf{H}_0^1 \text{ holds, at least one of } \mathsf{H}_0^2 \text{ and } \mathsf{H}_0^3 \text{ does not hold} \\ \mathsf{H}_A^C &: \quad \mathsf{H}_0^1 \text{ and } \mathsf{H}_0^2 \text{ hold, } \mathsf{H}_0^3 \text{ does not hold} \end{aligned}$$

This leads to the following local alternatives

$$\begin{aligned}
\mathsf{H}_n^1 &: \sqrt{n_1} m(\theta_0, \beta_1) = \delta_1 + o(1) \\
\mathsf{H}_n^2 &: \sqrt{n_2} m(\theta_0, \beta_2) = \delta_2 + o(1) \\
\mathsf{H}_n^3 &: m(\overline{\theta}_{0n}, \beta_2) = 0, \ \sqrt{n_2}(\overline{\theta}_{0n} - \theta_0) = \delta
\end{aligned}$$

Let $\mathsf{H}_l^A := \{(\mathsf{H}_n^1, \mathsf{H}_n^2)\}_{n:=1}^{\infty}$, $\mathsf{H}_l^B := \{(\mathsf{H}_0^1, \mathsf{H}_n^2)\}_{n:=1}^{\infty}$ and $\mathsf{H}_l^C := \{(\mathsf{H}_0^1, \mathsf{H}_n^3)\}_{n:=1}^{\infty}$. The usual Hansen $J-$test may be used to test H_0^s. However, Ghysels and Hall (1990b) show that in the context of structural stability this test has no power against local alternatives that are of the form H_l^C. In Ghysels and Hall (1990a) a modification of the Hansen test is proposed, the JSS test. This test has optimal power against H_l^A.

Several interesting asymptotic equivalencies for the Wald/LR/LM type tests for structural stability proposed by Andrews and Fair (1988) can be given. The upshot is that any of their statistics has maximum local power against alternatives of the form H_l^C.

Ahn (1995) shows, building on earlier work of Newey (1985) that the PSP test for GMM is an optimal GMM test that has maximum power toward H_l^B. In the context of EMM these PSP statistics are investigated in an accompanying paper, Van der Sluis (1997a).

As said, in a GMM context a modification of the Hansen test is proposed by Ghysels and Hall (1990a). For EMM we define the following analogy

$$\mathsf{JSS} := n_1 m_N^{'}(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_1})(\widehat{\mathcal{I}}_{n_1})^{-1} m_N(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_1}) + n_2 m_N^{'}(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_2})(\widehat{\mathcal{I}}_{n_2})^{-1} m_N(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_2})$$

this statistic is already computationally attractive since only $\widetilde{\theta}_{n_1+n_2}$ must be determined. However, for use in an update setting a modified version will be proposed.

For EMM the following Wald, LR and LM type tests for structural stability are proposed. These are analogies with the test statistics for structural stability proposed in Andrews and Fair (1988). Note the presence of the computationally unattractive estimators $\widehat{\theta}_{n_2}$ and $\widetilde{\theta}_{n_1+n_2}$ in the expressions for the test-statistics below.

$$\begin{aligned}
\mathsf{W} &: = (\widehat{\theta}_{n_2} - \widehat{\theta}_{n_1})^{'}[n_1 \widehat{\mathcal{V}}_1^{-1}(\widehat{\theta}_{n_1}) + n_2[\widehat{\mathcal{V}}_2(\widehat{\theta}_{n_2})]^{-1}(\widehat{\theta}_{n_2} - \widehat{\theta}_{n_1}) \\
\mathsf{LR} &: = [n_1 m_N^{'}(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_1})(\widehat{\mathcal{I}}_{n_1})^{-1} m_N(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_1}) + \\
& \quad n_2 m_N^{'}(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_2})(\widehat{\mathcal{I}}_{n_2})^{-1} m_N(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_2})] - \\
& \quad [n_1 m_N^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})(\widehat{\mathcal{I}}_{n_1})^{-1} m_N(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1}) + \\
& \quad n_2 m_N^{'}(\widehat{\theta}_{n_2}, \widehat{\beta}_{n_2})(\widehat{\mathcal{I}}_{n_2})^{-1} m_N(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})] \\
\mathsf{LM} &: = n_1 m_N^{'}(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_1})(\widehat{\mathcal{I}}_{n_1})^{-1} \mathcal{M}_1(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1}) \widehat{\mathcal{V}}_1^{-1}(\widehat{\theta}_{n_1}) \mathcal{M}_1^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})(\widehat{\mathcal{I}}_{n_1})^{-1} m_N(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_1}) + \\
& \quad n_2 m_N^{'}(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_2})(\widehat{\mathcal{I}}_{n_2})^{-1} \mathcal{M}_2(\widehat{\theta}_{n_2}, \widehat{\beta}_{n_2}) \widehat{\mathcal{V}}_2^{-1}(\widehat{\theta}_{n_2}) \mathcal{M}_2^{'}(\widehat{\theta}_{n_2}, \widehat{\beta}_{n_2})(\widehat{\mathcal{I}}_{n_2})^{-1} m_N(\widetilde{\theta}_{n_1+n_2}, \widehat{\beta}_{n_2})
\end{aligned}$$

For GMM, Newey (1985) shows that the W,LR and LM tests are asymptotically equivalent. Trivially, these results carry over to an EMM setting. Furthermore these tests have optimal local asymptotic power against alternatives of different form than the PSP test for EMM[4] see Van der Sluis (1997a). The PSP test has optimal local asymptotic power against alternatives of the form H_l^B. As said above the JSS test has optimal local asymptotic power against yet another class of alternatives, H_l^A.

A class of modified J-tests is proposed here.

[4] For completeness' sake: the PSP test for EMM reads

$$\mathsf{PSP} := n_2 m_N^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_2})\{\widehat{\mathcal{I}}_{n_2} + k\mathcal{M}_2(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_2})[\mathcal{M}_1^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})\widehat{\mathcal{I}}_{n_1}^{-1}\mathcal{M}_1(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_1})]^{-1}\mathcal{M}_2^{'}(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_2})\}^{-1} m_N(\widehat{\theta}_{n_1}, \widehat{\beta}_{n_2})$$

Definition 1 *For any root-n consistent estimator $\widehat{\theta}$:*

$$\mathsf{JSS}^*(\widehat{\theta}) := n_1 m_N^{'}(\widehat{\theta},\widehat{\beta}_{n_1})(\widehat{\mathcal{I}}_{n_1})^{-1} m_N(\widehat{\theta},\widehat{\beta}_{n_1}) + n_2 m_N^{'}(\widehat{\theta},\widehat{\beta}_{n_2})(\widehat{\mathcal{I}}_{n_2})^{-1} m_N(\widehat{\theta},\widehat{\beta}_{n_2}) - \Psi(\widehat{\theta})$$

where

$$\Psi(\widehat{\theta}) := [\sum_{i:=1}^{2} n_i m_N^{'}(\widehat{\theta},\widehat{\beta}_{n_i})(\widehat{\mathcal{I}}_{n_i})^{-1}\mathcal{M}_i(\widehat{\theta},\widehat{\beta}_{n_i})][\sum_{i:=1}^{2} n_i \widehat{\mathcal{V}}_i(\widehat{\theta})]^{-1}[\sum_{i:=1}^{2} n_i \mathcal{M}_i^{'}(\widehat{\theta},\widehat{\beta}_{n_i})(\widehat{\mathcal{I}}_{n_i})^{-1} m_N(\widehat{\theta},\widehat{\beta}_{n_i})]$$

This modification is based on the following property

Property 2 *Under H_l^A this $\mathsf{JSS}^*(\widehat{\theta})$ is asymptotically identical to JSS. (Proof: trivial from Ahn (1995) for the GMM case).*

A class of modified LM/LR/Wald test is also proposed.

Definition 2 *For any root-n consistent estimator $\widehat{\theta}$, the statistic $\mathsf{P}(\widehat{\theta})$ is defined as*

$$\mathsf{P}(\widehat{\theta}) = \sum_{i=1}^{2} n_i m_N^{'}(\widehat{\theta},\widehat{\beta}_{n_i})(\widehat{\mathcal{I}}_{n_i})^{-1}\mathcal{M}_i(\widehat{\theta},\widehat{\beta}_{n_i})\widehat{\mathcal{V}}_i^{-1}(\widehat{\theta})\mathcal{M}_i^{'}(\widehat{\theta},\widehat{\beta}_{n_i})(\widehat{\mathcal{I}}_{n_i})^{-1} m_N(\widehat{\theta},\widehat{\beta}_{n_i}) - \Psi(\widehat{\theta})$$

here $\Psi(\widehat{\theta})$ is taken as above.

This modification is based on the following properties which have been first proposed in the GMM case by Ahn (1995):

Property 3 *For any root-n consistent estimator $\widehat{\theta}$, $\mathsf{P}(\widehat{\theta})$ is asymptotically identical to the LR statistic.*

Property 4 *$\mathsf{P}(\widetilde{\theta}_{n_1+n_2})$ equals(!) the LM statistic.*

Property 5 *$\mathsf{P}(\widehat{\theta}_{n_1})$ is asymptotically identical to the W statistic.*

Application of Newey (1985) proposition 3 tells us that $\mathsf{P}(\widetilde{\theta}_{n_1+n_2})$ has maximum power against H_l^C. Because of the asymptotic equivalence between any $\mathsf{P}(\widehat{\theta})$ and the Wald, LR and the LM statistic, we define for EMM the following computationally attractive Wald, LR and LM type tests for structural stability which have maximum power against H_l^C

$$\begin{array}{lcl} \mathsf{W}^* & := & \mathsf{P}(\widehat{\theta}_{n_1}) \\ \mathsf{LR}^*(\widehat{\theta}) & := & \mathsf{P}(\widehat{\theta}) \text{ for any root-}n\text{ consistent } \widehat{\theta} \\ \mathsf{LM} & = & \mathsf{P}(\widetilde{\theta}_{n_1+n_2}) \end{array}$$

The W^* test requires only the computation of $\widehat{\theta}_{n_1}$ and therefore it is computationally attractive. The LM^* test also requires only one computationally intensive EMM-optimization round for the sample in order to find

under H_l^B we have $\mathsf{PSP}_{\mathsf{EMM}}$ has a noncentral $\chi^2_{q_2}$ distribution with noncentrality parameter $\lambda_{\mathsf{PSP}_{\mathsf{EMM}}}$, given by

$$\lambda_{\mathsf{PSP}} := \delta_2^{'}\mathcal{I}_2^{-1}[\mathcal{I}_2 - c_2\mathcal{M}_2(\theta_0,\beta_2)[\sum_{i:=1}^{2}\mathcal{M}_i^{'}(\theta_0,\beta_i)\mathcal{I}_i^{-1}\mathcal{M}_i(\theta_0,\beta_2)]^{-1}\mathcal{M}_2^{'}(\theta_0,\beta_i)]\mathcal{I}_2^{-1}\delta_2$$

where $c_2 := \lim_{n_1,n_2\to\infty}\frac{n_2}{n_1+n_2}$.

See Van der Sluis (1997a) for details and for an application in the context of exchange rates.

$\widetilde{\theta}_{n_1+n_2}$. Compared to the original form of the LM test which requires three optimization rounds namely for $\widehat{\theta}_{n_1}, \widehat{\theta}_{n_2}$ and $\widetilde{\theta}_{n_1+n_2}$, this is an enormous gain. However using $\widetilde{\theta}_{n_1+n_2}$ does not fit into the updating framework, presented here. Let $\beta_i := \text{plim}_{n_i \to \infty} \widehat{\beta}_{n_i}$ and the factor $c_i := \lim_{n_1, n_2 \to \infty} \frac{n_i}{n_1+n_2}$

All $\mathsf{P}(\theta)$ statistics have a noncentral χ^2_p with noncentrality parameter λ_{P} given by

$$
\begin{aligned}
\lambda_{\mathsf{P}} \quad : \quad &= \delta^{'} \mathcal{M}_2^{'}(\theta_0, \beta_2) \mathcal{I}_2^{-1} [\mathcal{I}_2 - c_2 \mathcal{M}_2(\theta_0, \beta_2) \cdot \\
&[c_1 \mathcal{M}_1^{'}(\theta_0, \beta_1) \mathcal{I}_1^{-1} \mathcal{M}_1(\theta_0, \beta_1) + c_2 \mathcal{M}_2^{'}(\theta_0, \beta_2) \mathcal{I}_2^{-1} \mathcal{M}_2(\theta_0, \beta_2)]^{-1} \cdot \\
&\mathcal{M}_2^{'}(\theta_0, \beta_2)] \mathcal{I}_2^{-1} \mathcal{M}_2(\theta_0, \beta_2) \delta
\end{aligned}
$$

If $p = q_2$ the $\mathsf{P}(\theta)$ and PSP tests in Van der Sluis (1997a) are asymptotically identical. For $q_2 > p$ the PSP tests have more degrees of freedom than the $\mathsf{P}(\theta)$ based tests where the noncentrality parameters are equal under H_l^C. Consequently the $\mathsf{P}(\theta)$ statistics have higher power against H_l^C. On the other hand the PSP test has maximum power towards H_l^B.

The JSS^* or JSS statistics have a noncentral $\chi^2_{q_1+q_2-p}$ distribution with noncentrality parameter λ_{JSS^*} given by

$$
\lambda_{\mathsf{JSS}^*} := \delta_{1,2}^{'} [\mathcal{I}_{1,2}^{-1} - \mathcal{I}_{1,2}^{-1} \mathcal{M}_{1,2} [\mathcal{M}_{1,2}^{'} \mathcal{I}_{1,2}^{-1} \mathcal{M}_{1,2}]^{-1} \mathcal{M}_{1,2}^{'} \mathcal{I}_{1,2}^{-1}] \delta_{1,2}
$$

where

$$
\begin{aligned}
\mathcal{M}_{1,2} \quad : \quad &= [c_1 \mathcal{M}_1^{'}(\theta_0, \beta_1), c_2 \mathcal{M}_2^{'}(\theta_0, \beta_2)]^{'} \\
\mathcal{I}_{1,2} \quad : \quad &= \text{diag}(\mathcal{I}_1, \mathcal{I}_2) \\
\delta_{1,2} \quad : \quad &= [\sqrt{c_1} \delta_1^{'}, \sqrt{c_2} \delta_2^{'}]^{'}
\end{aligned}
$$

As usual there is a connection between the efficiency of the estimator and the power of a test. The non-centrality parameter often reveals this connection. Following Tauchen (1996) in his exposition of the full efficiency of estimators, let $\mathcal{V}_0^{-1}$ be the asymptotic variance-covariance matrix of the ML estimator of θ_0. Let $\mathcal{V}_{f_{K_i,L_i}}^{-1}$ be the asymptotic variance-covariance matrix of the EMM estimator based on a score f with lags K_i and L_i and observations n_i

Property 6 *Theorem 2 in Gallant & Long (1997), we have that, under certain conditions, for any* i :

$$
\lim_{K_i \to \infty} \lim_{L_i \to \infty} \mathcal{V}_{f_{K_i,L_i}}^{-1} = \mathcal{V}_0^{-1} \textit{as } n_i \to \infty
$$

In our case this result extends to the noncentrality parameter of the P statistics

Property 7

$$
\lim_{K_1, K_2, L_1, L_2 \to \infty} \lambda_{\mathsf{P}} = c_1 \delta^{'} \mathcal{V}_0^{-1} \delta \textit{ as } n_1, n_2 \to \infty
$$

(Proof: simple algebraic manipulations)

By the Cramer-Rao inequality we have that conditionally on c_i, the highest asymptotic power is achieved in the limiting case, yielding in a *uniformly most powerful test*. Consequently, we should increase q_1 with n_1 and q_2 with n_2.

It is not very likely that for the λ_{JSS^*} such property exists. Maybe further research will be more conclusive on this. The main thing to note here is that the JSS test becomes more stringent with an increasing number of moment conditions. In many applications the test will probably become too stringent and it may be wise to test for structural stability using only a subset of the moment conditions. The same is true for the PSP test in Van der Sluis (1997a). A full Monte Carlo study will be needed to assess the finite sample properties of these tests and it may turn out that it is wise to adjust the critical values of these tests. Another possibility is to apply the same techniques as in Liu and Zhang (1996) to these stability tests. In this paper a new specification tests is proposed that minimizes the inference bias of the Hansen overidentifying J test caused by the approximation error in the auxiliary model.

It should be noted that *any* root-n consistent estimator may be plugged into the modifications of the test-statistics. This makes it possible to plug in estimates from different estimation techniques. Note that in case we choose for the LM/LR/Wald tests our moments in an optimal way, that is following the EMM methodology, asymptotically the highest power possible is obtained, namely the power associated with the Cramer-Rao lower bound from maximum likelihood, while leaving the hypothesis H_A^C unchanged. A different choice of the moment conditions for the JSS test and the PSP test induces different overidentifying restrictions, so different alternative hypotheses.

5 Application to stochastic volatility models

The following classes of stochastic volatility models are considered. The reader is advised to consult Ghysels *et al*. (1995) in case of nonfamiliarity with these type of models. Stochastic volatility models are mainly used to model the conditional heteroskedasticity in financial time-series, such as stock prices, exchange rates and interest rates. The main difference with ARCH type of models is the fact that some elements in the specification are *latent*. The first class of stochastic volatility models we consider is the following

$$\begin{aligned}
&y_t = \sigma_t \epsilon_t \\
&\ln \sigma_t^2 = \omega + \textstyle\sum_{i=1}^p \rho_i \ln \sigma_{t-i}^2 + \eta_t + \sum_{i=1}^q \zeta_i \eta_{t-i} \\
&\epsilon_t \sim \mathsf{NID}(0,1), \eta_t \sim \mathsf{NID}(0,\sigma_\eta^2) \\
&t = 1, \ldots, T
\end{aligned}$$

For obvious reasons this class of models will be referred to as SARMAV(p,q) models[5]. In the recent literature mainly the SARMAV$(1,0)$ model is considered. In the current application this class of model will also mainly be considered. Several modifications of this model are considered. One modification is the SARMAV$(p,q)-\mathsf{t}(\nu)$ model. This class of models may capture more of the excess *kurtosis* observed in financial time-series. This model reads

$$\begin{aligned}
&y_t = \sigma_t \epsilon_t \\
&\ln \sigma_t^2 = \omega + \textstyle\sum_{i=1}^p \rho_i L^i \ln \sigma_t^2 + \sigma_\eta (1 + \sum_{j=1}^q \zeta_j L^j) \eta_t \\
&\epsilon_t \sim \mathsf{t}_\nu / \sqrt{\tfrac{\nu}{\nu-2}}, \eta_t \sim \mathsf{NID}(0,\sigma_\eta^2) \\
&t = 1, \ldots, T
\end{aligned}$$

where ν is treated as a parameter to be estimated. Since financial time-series often show asymmetric behaviour an asymmetric modification of the SARMAV models may be considered: the ASARMAV[6] model. This model

[5] Stochastic AutoRegressive Moving Average Volatility models

[6] Asymmetric Stochastic AutoRegressive Moving Average Volatility. A different form of the specification may be of more interest, namely where ϵ_{t-1} and η_t are correlated.

is defined as follows

$$
\begin{aligned}
& y_t = \sigma_t \epsilon_t \\
& \ln \sigma_t^2 = \omega + \sum_{i=1}^{p} \rho_i L^i \ln \sigma_t^2 + \sigma_\eta (1 + \sum_{j=1}^{q} \zeta_j L^j) \eta_t \\
& \epsilon_t \sim \mathsf{NID}(0,1), \eta_t \sim \mathsf{NID}(0,1), \mathsf{Corr}(\epsilon_{t-1}, \eta_t) = \lambda, \ \ -1 \leq \lambda \leq 1 \\
& t = 1, \ldots, T
\end{aligned}
$$

This model only differs from the SARMAV specification that the ϵ_{t-1} and η_t are correlated. This allows for asymmetric behaviour or *skewness* as in EGARCH models. Trivially, the SARMAV model is a special case of the ASARMAV model with $\lambda = 0$. As for standard ARMA models, stationarity is governed by the ρ parameters, so for stationarity the roots of 1-$\rho_1 z^1 - \rho_2 z^2 - \ldots - \rho_p z^p$ must lie outside the unit circle.

Stochastic volatility models arise in *option pricing*. The issue of stability testing may be thought of as one were financial institutions are calculating option prices from the estimates of a stochastic volatility model (see Hull & White (1987)). With the arrival of new data there is a dilemma. Should the financial institution estimate the stochastic volatility model again in order the obtain more up to date option prices or not? This question may be answered by the above computationally attractive test statistics.

The data are daily data of the S&P500. In the next subsections the score generator, the data, some practical issues on the implementation of EMM and the results are successively presented.

5.1 Score generator

As explained in section 2, the class of auxiliary models are taken from the SNP class of models of Gallant & Nychka (1987), building on earlier work of Phillips (1983). It is reported in the literature that the performance of the SNP model is very much enhanced in case a parametric model is used as a leading term. The parametric model is taken from the EGARCH class of models (Nelson (1991)). With GARCH models (Bollerslev (1986)) a problem arises concerning stationarity. On itself this is not a problem see Nelson & Cao (1991) and Kleibergen & Van Dijk (1993), but for EMM it is, since EMM hinges on concepts as ergodicity and stationarity, it must be guaranteed that the simulated series generates a stable score-generator. For parameter values outside the most stringent parameter space this cannot be guaranteed. For a more extensive discussion of this issue, reference is made to Andersen & Lund (1997) and Van der Sluis (1997a). Moreover, for GARCH models parameter estimates will often turn out to be on this most stringent IGARCH boundary. On this IGARCH boundary the error variances become infinite. Because of this property a comparison of a long simulated IGARCH series with real financial data shows that IGARCH models are rather far from realistic. Pure ARCH models (Engle (1982)) were not considered because often many ARCH parameters are needed to capture all the persistent heteroskedasticity in the data. This plethora of parameters may cause the same problems as reported in Andersen & Sørensen (1996). In this paper stochastic volatility models are estimated by means of GMM. Andersen & Sørensen report that the use of too many moments drastically deteriorates GMM estimation.

Another argument for the use of the EGARCH model is the fact that it is in some sense a closer approximation to a stochastic volatility model, see Nelson & Foster (1994). The last argument is the fact that EGARCH model is a richer model consequently there are more directions of misspecification captured by tests that are based on these models. For example the EGARCH model on itself is able to capture asymmetry, the GARCH model is not. Therefore changes in asymmetry are not picked up by tests based on a GARCH score generator, whereas the EGARCH score generator might pick up some of the changes in asymmetry.

The EGARCH class of models is defined as

$$
\begin{aligned}
&y_t = \sigma_t \epsilon_t \\
&\sigma_t^2 = \exp\{\alpha_0 + \textstyle\sum_{i=1}^{p} \gamma_i \ln \sigma_{t-i}^2 + (1 + \Sigma_{i:=1}^{q} \alpha_i L^i)[\kappa_1 z_{t-1} + \kappa_2(|z_{t-1}| - \mathsf{E}|z_{t-1}|)]\} \\
&\epsilon_t \sim \mathsf{NID}(0, 1) \\
&t = 1, \ldots, T
\end{aligned}
$$

Extensive analysis of these type of models can be found in Bollerslev *et al.* (1994). The stationarity conditions can be derived in exactly the same way as in the ARMA models à la Box-Jenkins, so only the γ parameters matter for stationarity. For more complicated stochastic volatility models than the one considered here EGARCH models were successfully used as a leading term in the score generator in Andersen & Lund (1996 & 1997).

The metaphor *window* for a likelihood function of Poirier (1991) may also be appropriate in this respect. The EGARCH(p, q)-H(K_x, K_z) model serves as a window by which we view the world. Here *world* is a methaphor for a SARMAV model, which is of course a rather narrow view on the world. If the window is soiled, we need to wash or abandon it. The choice of the window determines our perspective of the world. In the case of an EGARCH(p, q)-H(K_x, K_z) the following panorama is given: The γ parameters give us a idea of the persistence in the data, in particular $\sum_{i=1}^{p} \gamma_i$ will serve as a measure of persistence. The α_0 parameter is a measure of the unconditional variance. The term $1 + \sum_{i=1}^{q} \alpha_i$ reflects the short-run sensitivity to innovations. The term κ_1 reflects the asymmetric volatility effect that is often present in financial time series. The κ_2 parameter measures the changes in z_t in reaction with the conditional heteroskedasticity. In case $K_x = 0$, letting $K_z > 0$ induces a time-homogeneous non-Gaussian error structure. The case $K_x > 0$ induces heterogeneous innovation densities beyond the EGARCH model. Since we believe the EGARCH model captures all this heterogeneity, $K_x > 0$ will not be necessary. This was also very much supported by our empirical findings. To the author's knowledge there are no guidelines provided yet for assessment of the individual coefficients of the Hermite polynomial.

5.2 The data

The data under investigation are raw daily end of day quotes of the S&P500 index. The breakpoint is not set at, but two weeks after Black Monday in October 1987. This is done because inspection of the data reveals that in the aftermath of Black Monday the market was still very volatile. Taking Black Monday as a breakpoint for structural stability tests may be interesting for various reasons. The awareness that a crash of this size can occur may change the *animal spirits* of the investors. Less hypothetically: the Black Monday crash induced a change in the institutions of the trading mechanisms and a revision of many computer programs for automatic trading. Revision of the assumptions underlying program trading may be the main source for a structural break. In the early 1980s computerized programs for automatic trading became rather popular. Program trading is used for three types of investments: portfolio *insurance* (hedging), stock index *arbitrage* and *speculative trading*. These types of programs surely made the market more efficient, however some assumptions underlying program trading were proven wrong on Black Monday. *Insurance, Arbitrage and Speculation* are based on the fact that stock index futures follow actual stock prices. During Black Monday this did not happen. Small gaps between futures and stock prices are used for the so-called *riskless arbitrage*. Big gaps may seem to be a dream come true for arbitrageurs and speculators, however for arbitrage or opposite trading, an order on the futures market and on the stock market must be executed simultaneously. This was not possible with a very high volume, moreover conditions were changing constantly, therefore many arbitrageurs and speculators ceased trading. In 1988 several regulations were set on *program trading*.

Data were available up to December 1993. Our post-sample thus consists of 1322 data points, starting date: beginning of November 1987 end date: end of December 1993. The sample was chosen to be of the same size as the post-sample, ranging from early September 1981 till two weeks after Black Monday 1987. Both data sets were first differenced in 100*log(data). The series are plotted in figure 1. Some preliminary data analysis is provided in table 1. The skewness and kurtosis are both much higher for the sample than they are for the post-sample. This is very much due to Black Monday and its aftermath. As we see from the correlograms in figure 2a and 2b, there may be some autocorrelation of the y_t be present in the data. However, only a very few autocorrelations are marginally significant on a 5% level. However for samples of this size a 5% significance level may be too high. In addition to this, the only lag that is marginally significant in both sample and post-sample is lag number five. This lag may correspond with the day of the week effect. There are at least three possible ways to deal with this issue. The first one is the take this effect into account in the model. This means an increase of the dimension of the problem with two, which will substantially increase the computing time. The second option may be to filter all the correlation out. However this may cause other unwanted problems. The third one is to do nothing. Such behaviour is supported by Theoretical Finance and considering the correlation is very small, the effect on the volatility process may very well be futile. Simulations from a pure stochastic volatility model show also the same kind of autocorrelograms for y_t. The test should therefore be whether the z_t are autocorrelated not the y_t. This will be investigated in section 5.4.

5.3 Implementation

Coding was done in Ox 1.11 (Doornik (1996))[7] and in C/C++. The code is extensively described and made available in Van der Sluis (1997b). The code was executed on a P5-90, a P5-133 and on several nodes of a SP2 at SARA[8]. The SP2 consists of 76 nodes where each node represents one IBM RS/6000 processor. With the current implementation of Ox for the RS/6000, the speed of one of such a node can be compared to twice a P5-133 for some particular Ox benchmark program. Since each node consists of 512 Mb RAM, less *paging* is needed, which speeds up the process. In principle one node can be used exclusively. For this problem it was very important to *hard code* the bulk of the Ox procedures in C. Linking them to Ox by means of a dynamic link library on Windows based machines. Static linking had to be used on the SP2. The reported computing times in this paper are based on the OX/C program EmmPack 1.0 (Van der Sluis (1997b)). It is worth mentioning that the C code on the RS/6000 turned out to be less efficient than the C code on the Pentium machines. This makes that for Ox/C programs a P5-133 and a RS/6000 are comparable in speed. It is also worth mentioning that parallellisation techniques on the SP2 are under investigation.

In the Monte Carlo computation of the expectation of the score generator, a variance reduction technique called *antithetic variables* was used, see Ross (1990) for a general context and Andersen & Lund (1996a) for a specific context. This means that $y_1(\theta)$ and $y_2(\theta)$ are simulated using antithetic variables, e.g. ϵ and $-\epsilon$. Also in the calculation of the test-statistics this technique has been employed. A dramatic decrease of the variance is obtained. The precision of the expectation in the antithetic variables case, is currently under investigation. Using antithetic variables leads to a dramatic decrease of computing time and memory allocation. Computing time which is still very long. For example, with $N = 2 \times 50,000$, which may be somewhat economical, one BFGS[9] iteration, including numerical derivatives, takes, for the simplest model SARMAV$(1,0)$ model, with an

[7] Some prelimary data analysis was done in PcGive (Hendry & Doornik (1996)) and in Gauss 3.14.
[8] Stichting Academisch Rekencentrum Amsterdam
[9] Quasi-Newton method developed by Broyden-Fletcher-Goldfarb-Shanno.

EGARCH$(1,2)$-H$(5,0)$ auxiliary model, takes approximately fifteen minutes on a P5-133. Depending on the starting values usually more than 15 iterations are needed[10]. Typically $N = 2 \times 50,000$ is at least necessary to obtain four or five digit accuracy in the estimates, this means roughly about 4 hours on one RS/6000 for a complete maximization.. An extra half an hour for the computation of all the numerical derivatives ($\mathcal{M}$) matrices and test-statistics, makes it indeed a very demanding estimation method. .

Opposed to the generic case in Gallant & Tauchen (1996) in this specific case no starting values are needed to start up the stochastic volatility model. Parallel to ARMA time series models, the ASARMAV model is started up as[11]

$$\begin{aligned} y_0 &= \epsilon_0 \sigma_0, \\ \ln \sigma_0^2 &\sim N(\omega/(1 - \sum_{i=1}^{p} \rho_i), \sigma_\eta^2 (1 + \sum_{j=1}^{q} \zeta_j^2)(1 - \sum_{i=1}^{p} \rho_i^2)^{-1}) \\ \epsilon_0 &\sim N(0,1),\ \mathsf{Corr}(\epsilon_0, \ln \sigma_0^2) = \lambda / \sqrt{(1 + \sum_{j=1}^{q} \zeta_j^2)(1 - \sum_{i=1}^{p} \rho_i^2)^{-1}} \end{aligned}$$

Since stationarity is assumed[12] and for this model the marginal distribution of y_0 is known, there is no need to use starting values here. Antithetic variables are also used for determining (y_0, σ_0).

Another practical issue is the fact that the absolute function in the EGARCH model causes severe numerical problems. This is caused by the fact that the absolute function is non-differentiable in zero. The function $|z|$ is therefore replaced by a function which gives a very close approximation to the absolute function and is smooth around zero[13]. In the EGARCH model the term $\mathsf{E}|z_t|$ is approximated by $\sqrt{\frac{2}{\pi}}$, treating z_t as Gaussian, this approximation is futile.

Besides the first-order approximationsm, the other ”computationally attractive” estimators were obtained from the *integration sampler* of Kim *et al.* (1996) using the OX/C program as documented in Shephard (1996). For the sample sizes in this paper computing times for these estimators are typically in terms of minutes rather than hours or days as they are currently for EMM. This may be due to the fact that the evaluation of the EGARCH-H score vector is probably more complex than the evaluations needed in one of the Shephard (1996) procedures. The rationale of using the mode of the Bayesian posteriors as a root-n consistent estimator *in a classical sense* can be found in Barndorff-Nielsen and Cox (1994)[14]. For details on the procedures the readers is advised to consult Kim *et al.* (1996) and Shephard (1996) and the references therein. In the application here their settings are closely followed and the *integration sampler* is used as it generally has lowest *inefficiency* measure[15].

[10] One technique the author found useful is to use in the first say, twenty, iterations less replications, say $N = 2 \times 10,000$ and then to use a higher number of replications, say $N = 2 \times 50,000$ when more precision is needed.

[11] The case $\lambda = 0$, is trivial from ordinary time-series models. The case $\lambda \neq 0$ is straightforward algebra. Note that in the stationary region it is always the case that $\sqrt{(1 + \sum_{j=1}^{q} \zeta_j^2)(1 - \sum_{i=1}^{p} \rho_i^2)^{-1}} \geq 1$. Taking care of the fact that $|\mathsf{Corr}(\epsilon_0, \ln \sigma_0^2)|$ should be ≤ 1.

[12] Parallel to the stationarity condition for the ARMA and EGARCH model, the stationarity condition for the ASARMAV model is that the roots of the polynomial $1 - \rho_1 z - \ldots - \rho_p z^p$ must lie outside the unit circle.

[13] A good choice is given in Andersen & Lund (1996a) here $|z|$ is replaced by $b(z)$ where $b(z) = |z|$ for $|z| \geq \pi/2K$ and $b(z) = (\pi/2 - \cos(Kz))/K$ for $|z| < \pi/2K$ and set e.g. $K = 100$.

[14] One can even argue that these modes of the Bayesian procedures have Maximum Likelihood properties in a classical sense. However, we choose to stay on the safe side and take only the property of root-n consistency as this a sufficient condition.

[15] Here *inefficiency* has a different, namely Bayesian, meaning than it has in the rest of the paper. The interested reader should consult the original papers.

5.4 Results

At first, an auxiliary model must be specified to the sample, the post-sample and the combined sample and post-sample. In table 2 and 3 several model selection criteria are given. As discussed in section 2 our main guide is the BIC. As argued in section 5.1 an EGARCH model was taken as a leading term. For the sample the EGARCH$(1,2)$-H$(5,0)$ model was chosen as a score-generator, estimates for this optimal model are given in table 4. For the post-sample the results for the EGARCH$(1,1)$ and EGARCH$(1,2)$ leading terms are given in table 5 and 6. The BIC is in favour of an EGARCH$(1,1)$-H$(8,0)$ model. The estimates that were obtained for the optimal auxiliary model of the post-sample are given in figure 7. For the combination of sample and post-sample the model selection criteria are displayed in table 8 and table 9. The estimates are given in table 10. Here the BIC is in favour of the EGARCH$(1,2)$-H$(8,0)$ model. We observe that the persistence parameter γ_1 in the post-sample is higher than the persistence in the sample. The asymmertry, governed by κ_1, is slightly higher in the sample than it is in the post-sample. The κ_2 parameter is much higher in the sample than in the post-sample. It is hard to assess the properties of the SNP terms here, we mention that many terms are significantly different from zero so we have estimated a density that is different from the Gaussian density.

The structural model was specified according to the following strategy: first some preliminary analysis was done using series of $N = 2*10,000$ in the Monte Carlo evaluation. After that the model was estimated with $N = 2*50,000$ and these estimates are reported. The first structural model that was considered is a SARMAV(1,0) model. The J$-$test as given in table 13 is too high. However in the EMM literature it is very often the case that the J$-$test is too high. Van der Sluis (1997a) is an exception on this rule. In this paper the stochastic volatility model is, according to the J$-$test, accepted for exchange rates using the same types of auxiliary models. Liu and Zhang (1996) identify a possible cause of the fact that the overidentifying tests is too high and give a remedy. From the individual elements of the score vector, as given in table 15, we learn that almost every term is significantly different from zero.

The next model that was considered is the SARMAV(1,0)-t model[16]. This model converged badly as was also reported for about the same model in Gallant *et al.* (1994). The number of degrees of freedom seemed to converge to infinity. This model is still under investigation by the author but is abandoned in this paper. The next model is the ASARMAV(1,0) model which can incorporate asymmetry. The J$-$test is lower than that of the SARMAV(1,0) but still too high as can be seen from table 13. Also observe from table 15 and 16 that some of the individual elements of the J test that correspond to the EGARCH parameters are significantly different from zero for the SARMAV(1,0) model, but are not significantly different from zero for the ASARMAV(1,0) model. This highlights the diagnostic capabilities of EMM-based test procedures. The ASARMAV(1,0) model still does not pass the test because of the nonparametric part. Finally, the *non-markovian* SARMAV(1,1) gave no substantial decrease of the Hansen J-test. We therefore decided not to consider the ASARMAV(1,0) and SARMAV(1,1) in the rest of the paper and to postpone a study of these types of asymmetric and non-markovian models to a later stage[17]. Therefore we decided to consider only the SARMAV(1,0) model with respect to the tests for structural stability. This is also convenient because the Bayesian procedure that we will use below are only implemented for the SARMAV(1,0).

[16] The author simultaneously estimated $(\omega, \rho, \sigma_\eta, v)$. The error terms ϵ_t were scaled to have a variance approximately equal to one. Such was achieved by dividing the ϵ_t by their theoretical variance $\frac{\nu}{\nu-2}$.

[17] Note assumptions 1 and 2 in Gallant & Long (1997) allow for the use non-markovian structural models.

Parameter estimates[18] and $t-$values for the SARMAV(1,0) estimated by EMM are

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \ln \sigma_t^2 &= \underset{(1.12)}{-.006} + \underset{(50.0)}{.975} \ln \sigma_{t-1}^2 + \underset{(2.08)}{.101} \eta_t \end{aligned}$$

Figure 3 provides a plot of two antithetic simulated series of the same size as the sample in the optimum. Comparison of the simulated series and the real sample tells us that this model cannot generate such extreme values as Black Monday. Also several longer series of 100,000 (about 400 years of daily data!) were plotted and it is revealed that in all these 400 years an extreme value such as Black Monday is not even by far generated. Likewise a pure EGARCH model at the estimated parameter values will never be able to generate such an outlier. In the EGARCH case the outliers must be generated by the SNP part of the error-term. This explains partly why the individual elements in the J$-$test as displayed in table 15 and 16, are significantly different from their neutral values.

More importantly from table 12 we can see that the SARMAV(1,0) does not capture the excess kurtosis and skewness that are present in the data (see table 1). In table 12 theoretical and where needed simulated characteristics of the models in the optimum are provided. For the SARMAV(p,q) and SARMAV(p,q) $-$ t_v the theoretical values of the mean, standard deviations, skewness and kurtosis are known, see appendix B. These theoretical values are respectively 0, $\exp\{\frac{\omega}{2(1-\rho)} + \frac{\sigma_\eta^2}{4(1-\rho^2)}\}$ $(= .9398)$,0 and $3\exp\{\frac{2\omega}{1-\rho} + \frac{2\sigma_\eta^2}{(1-\rho^2)}$ $(= 4.485)$ for the SARMAV$(1,0)$ model in the optimum.

The integration sampler of Kim *et al.* (1996) provides us[19] with relatively fast parameter estimates of the post-sample. These posterior modes are

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \ln \sigma_t^2 &= -.007 + .979 \ln \sigma_{t-1}^2 + .136 \eta_t \end{aligned}$$

Note that the Bayesian estimates for the post-sample are almost equal to the EMM estimates of the sample. Only the stochastic volatility is higher which may be reasonable considering the volatility on the market after Black Monday. The first order approximations are given in table 11. Note that the estimates for σ_η are slightly higher than in the sample, but lower than the Bayesian estimates for the post-sample. The other parameters are all about the same, except for number IV. There is also an indication that the unconditional volatility $\omega/(1-\rho)$ is about the same over sample and post-sample. The stochastic component in the volatility becomes a little bigger over the post-sample.

The tests for structural stability are given in tables 14, 17 and 18. The tests JSS like and PSP tests all give rejections at any reasonable level, whereas the Wald like tests are rather low and will give acceptance at some reasonable level. Only in the LR case the outcome of the test is dependent on the estimator for the post-sample that is used. The LR test based on the Bayesian procedure reject whereas the others reject at a 0.05 level. This is clearly due to the higher estimate of σ_η the Bayesian procedure generated. High values for the J, JSS and PSP tests seem to be usual for models estimated by EMM, probably due to the rather stringent nature of the overidentifying restrictions, but also due to the fact that the EGARCH$(1,2)$-H$(5,0)$ model with 11 parameters probably describes this data set much better than a SARMAV$(1,0)$ model with only 3 parameters.

Individual $t-$values of the components of the test-statistics have been given in table 15. We observe that for the J test almost every element is significantly different from zero. For the PSP test the outcomes are not that

[18] Using $N = 2 * 50,000$

[19] Ox-procedures were kindly provided by Neil Shephard.

clear. The general conclusion is that for the S&P500 data set the SV models for both sample and post-sample are misspecified. This source of mispecification is mainly detected by the moment based tests and not by the tests for parameter variation. More importantly this source of misspecification is not due to the extreme Black Monday outlier but is also present in the more ”decent” post-sample. This can also be seen from a comparison of the values of the auxiliary parameters of the sample and post-sample. Further research in this direction will certainly be needed.

6 Conclusion

After the papers quoted in the above, this paper also offers again a successful application of the EMM technique of Gallant & Tauchen (1996). The success is very much due to recent results in Gallant & Long (1997) were the use of non-Markovian score-generators is justified. This result is crucial for a successful application of EMM in the context of financial time-series with high persistence in the volatility. Markovian score-generators have too many dimensions and would thus make a practical application unsuccessful. See Andersen & Sørensen (1996) for numerical evidence in a GMM context. Other fine-tuning of the EMM technique was achieved by the use of antithetic variables, as was also advocated in Andersen & Lund (1997), and the use of knowledge of the marginal distribution of the ASARMAV models, omitting the need for long simulated starting values. This knowledge cannot be used for level ASARMAV models, since for these types of models the stationarity conditions are yet unknown. Then it is needed to simulate a long series of starting values to ensure stationarity at the particular parameter values.

Apart from estimation this paper concentrates on tests for structural stability in the context of EMM. In a related paper Van der Sluis (1997a) a prediction test was derived and applied to stochastic volatility models. By definition this prediction test do not require computationally intensive estimators for the post-sample. In this paper other tests for structural stablity were investigated. Modifications were made to keep the same optimality procedures while again no computationally intensive estimators are needed for the post-sample, only root$-n$ consistency is required. Root$-n$ consistent first order approximations of the EMM estimators were thus derived. These are plugged into the test statistics. Also other root-n consistent estimators than the first order approximations proposed here, such as the ones of Kim *et al.* (1996), were plugged in. In the context of stochastic volatility models some other estimators that may be plugged in are the Kalman Filter estimator of Harvey *et al.* (1994), the Simulated Maximum Likelihood estimator of Danielsson (1994), the Bayesian Gibbs estimator of Jacquier *et al.* (1994) and the GMM estimator of Andersen & Sørensen (1996).

From a theoretical viewpoint it is better to calculate every computationally intensive EMM estimates for the structural parameters of the sample, post-sample and combination, and then perform the exact stability tests as also provided in the paper. However, in a real-life situation where option traders are every minute or even second confronted with the arrival of new data a quick assessment of the question whether *to update or not to update* the parameters in option pricing models is desirable.

The take-away of the empirical example here is that for this particular data-set the structural model is misspecified and the auxiliary model may be better in modeling the data. In case the Wald type tests for structural stability are used as a general specification test this source of misspecification will not be identified. The JSS and PSP tests do identify this source of misspecification indicating on a deeper source of misspecification. This misspecification is not due to the outlier at Black Monday in the sample, the same misspecification is present in the post-sample. The unconditional volatily seems to be stable over the sample and post-sample. The idiosyncratic stochastic volatility component seems to be bigger over the post-sample than it is over the sample.

Maybe JSS tests should be considered for subsets of the moment conditions only since these test may become too stringent in the context of EMM. This is also true for the J test and PSP test. Results in this paper show that using an expanding SNP model with any root-n consistent estimator, these Wald/LM/LR tests are uniformly most powerful. For the SARMAV$(1, 0)$ a decomposition of the noncentrality parameters for this particular model may be interesting: structural stability tests may have low power for other directions of misspecification, such as ω and ρ on the line $\omega/(1 - \rho)$, where the unconditional variance is equal for the sample and post-sample but the ω and ρ are different in the sample and post-sample.

It should be noted that the complexity of the estimation problem is rather high. The complexity of the numerical problem may be even too high for the given state of computer technology. For this reason some of the conclusions in the paper are rather preliminary, Monte Carlo studies are needed to pass judgement on this in the future. It is the author's believe that this will take at least a decade before a full scale Monte Carlo study can be done in the context of EMM.

References

[1] Abramowitz, M. and I.A. Stegun (1972), *Handbook of Mathematical Functions*. New York: Dover.

[2] Ahn, S. C., (1995), Model specification testing based on root-t consistent estimators. Working paper, Arizona State University.

[3] Akaike, H, (1973), Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*, 267-281. Budapest: Akademia Kiado.

[4] Andersen, T.G. and J. Lund (1997), Estimating continuous time stochastic volatility models of the short term interest rate. *Journal of Econometrics* **77**, 343-377.

[5] Andersen, T.G. and J. Lund (1996), Stochastic volatility and mean drift in the short rate diffusion: sources of steepness, level and curvature in the yield curve. Working paper No.214 Northwestern University.

[6] Andersen, T.G. and B.E. Sørensen (1996), Estimation of a stochastic volatility model: A Monte Carlo Study. Forthcoming *Journal of Business and Economic Statistics* .

[7] Andrews, D.W.K. and R.C. Fair (1988), Inference in econometric models with structural change. *Review of Economic Studies* **55**, 615-640.

[8] Bansal, R., A.R. Gallant, R. Hussey and G.E. Tauchen (1995), Nonparametric estimation of structural models for high-frequency currency market data. *Journal of Econometrics* **66**, 251-287.

[9] Barndorff-Nielsen, O.E. and D.R. Cox (1994), *Inference and Asymptotics*, London: Chapman & Hall.

[10] Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.

[11] Bollerslev, T., R. F. Engle and D. B. Nelson (1994), ARCH models. Ch. 49 in *Handbook of Econometrics* volume IV.

[12] Danielsson, J. (1994), Stochastic volatility in asset prices: estimation with simulated maximum likelihood, *Journal of Econometrics* **61**, 375-400.

[13] Doornik, J.A. (1996), *Object-Oriented Matrix Programming using Ox*, London: International Thomson Business Press and Oxford: http://www.nuff.ox.ac.uk/Users/Doornik/

[14] Eastwood, B. (1991), Asymptotic normality and consistency of seminonparametric regression estimators using an upward F-test rule, *Journal of Econometrics* **48**, 151-183.

[15] Fenton, V.M. and A.R. Gallant (1996a), Qualitative and asymptotic performance of SNP density estimators. Working Paper, University of North Carolina.

[16] Fenton, V.M. and A.R. Gallant (1996b), Convergence rates of SNP density estimators. *Econometrica* **64**, 719-727.

[17] Gallant. A.R., D.A. Hsieh and G.E. Tauchen (1991), On fitting a recalcitrant series: the pound/dollar exchange rate, 1974-1983, in W.A. Barnett, J. Powell and G.E. Tauchen (eds), *Nonparametric and semiparametric methods in econometrics and statistics*, Fifth International Symposium in Economic Theory and Econometrics, 199-240, Cambridge: Cambridge University Press.

[18] Gallant. A.R., D.A. Hsieh and G.E. Tauchen (1994), Estimation of stochastic volatility models with diagnostics. Forthcoming in *Journal of Econometrics*

[19] Gallant, A.R. and J.R.Long (1997), Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Square, *Biometrika* **84**, 125-141.

[20] Gallant, A.R. and D.W. Nychka (1987), Semi-nonparametric maximum likelihood estimation, *Econometrica* **55**, 363-390.

[21] Gallant, A. R. and G.E. Tauchen (1996), Which moments to match? *Econometric Theory* **12**, 657-681.

[22] Ghysels, E. and A. Hall (1990a), A test for structural stability of Euler conditions parameters estimated via the generalized method of moments estimator. *International Economic Review*, **31**, 355-64

[23] Ghysels, E. and A. Hall (1990b), Are consumption-based intertemporal capital asset pricing models structural? *Journal of Econometrics* **45**, 121-139.

[24] Ghysels, E., A. Harvey and E. Renault (1995), Stochastic Volatility. Forthcoming in *Handbook of Statistics*, vol 14: Statistical Methods in Finance.

[25] Gourieroux, C. , A. Monfort and E. Renault (1993), Indirect Inference, *Journal of Applied Econometrics* **8**, S85-S119.

[26] Hall, A. and A. Sen (1996), Structural stability testing in models estimated by generalized method of moments. Working paper, North Carolina State University.

[27] Hannan, E.J. and B.G. Quinn (1979), The determination of the order of an autoregression, *Journal of the Royal Statistical Society*, **B 41**, 190-195.

[28] Harvey A.C. E. Ruiz and N. Shephard (1994), Multivariate stochastic volatility models, *Review of Economic Studies* **61**, 247-264.

[29] Hendry, D.F. and J.A. Doornik (1996), *Empirical econometric modelling using PcGive for Windows*. London: International Thomson Business Press.

[30] Hoffman, D. and A. Pagan (1989), Post-sample prediction tests for generalized method of moments estimators. *Oxford Bulletin of Economics and Statistics*, **51**, 331-43

[31] Hull, J. and A. White (1987), The pricing of options on assets with stochastic volatilities, *Journal of Finance* **42**, 281-300.

[32] Jacquier, E., N.G. Polson and P.E. Rossi (1994), Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business and Economics Statistics* **12**, 371-417.

[33] Kim, S., N. Shephard and S. Chib (1996), Stochastic volatility: optimal likelihood inference and comparison with ARCH model. Discussion paper, Nuffield College, Oxford.

[34] Kleibergen, F. and H. K. van Dijk (1993), Non-stationarity in GARCH models: a Bayesian analysis, *Journal of Applied Econometrics* **8**, S41-S61.

[35] Liu, M. and H.H. Zhang (1996), Specification tests in the efficient method of moments framework with application to the stochastic volatility models. Working paper The Chinese University of Hong Kong and Carnegie Mellon University.

[36] Lütkepohl, H. (1990), *Introduction to multiple time series analysis*. Berlin: Springer Verlag

[37] Nelson, D.B. (1991), Conditional heteroskedasticity in asset returns: a new approach, *Econometrica* **59**, 347-370.

[38] Nelson D. B., and C. Q. Cao (1992), Inequality constraints in the univariate GARCH model. *Journal of Business & Economics Statistics* **10**, 229-235

[39] Nelson, D.B., and D.P. Foster (1994), Asymptotic filtering theory for univariate ARCH models, *Econometrica* **62**, 1-41.

[40] Newey, W. K. (1985), Generalized method of moments specification testing, *Journal of Econometrics* **29**, 229-256.

[41] Phillips, P.C.B. (1983), ERA's: A new approach to small sample theory, *Econometrica* **51**, 1505-1527.

[42] Poirier, D.J. (1991), Window washing: A Bayesian perspective on diagnostic checking. In *Volume in Honor of Arnold Zellner*, eds D.Berry, K.Chalon and J.Geweke.

[43] Quinn, B.G. (1980), Order determination for a multivariate autoregression, *Journal of the Royal Statistical Society*, **B 42**, 182-185.

[44] Ross, S.M. (1990), *A course in simulation*, New York: MacMillan.

[45] Rothenberg, T. J. (1973), *Efficient Estimation with A Priori Information*. New Haven: Yale University Press.

[46] Schwarz, G. (1978), Estimating the dimension of a model. *Annals of Statistics,* **6**, 461-464.

[47] Shephard, N. (1996), Ox software for volatility models, Discussion paper, Nuffield College, Oxford.

[48] Silverman, B.W. (1986), *Density estimation for statistics and data-analysis,* London: Chapman & Hall.

[49] Sowell, F. (1996), Test for violations of moment conditions. Manuscript, Graduate School of Industrial Administration, Carnegie Mellon University.

[50] Tauchen, G.E. (1996), New minimum chi-square methods in empirical finance, invited paper at Seventh World Congress of the Econometric Society, Tokyo Japan, August 1995.

[51] Tauchen, G.E., H.H. Zhang and M. Liu (1996), Volume, volatility and leverage: a dynamic analysis, *Journal of Econometrics* **74**, 177-208.

[52] Van der Sluis, P. J. (1997a), Post-sample prediction tests for EMM. Discussion paper TI97-054/4 Tinbergen Institute.

[53] Van der Sluis, P.J. (1997b), EmmPack 1.0: C code for use with Ox for estimation of univariate stochastic volatility models with the efficient method of moments, Unpublished working paper and freeware, University of Amsterdam. Paper presented at CEF 97, June 29-July 2 1997, Stanford University.

[54] White, H. (1994), *Estimation, inference and specification analysis*, Cambridge: Cambridge University Press.

A Appendix

This appendix contains proofs of the theorems set forth in the paper

Proof of theorem 1

Let $m_{12} := [m^{'}(\theta,\beta_1)\ m^{'}(\theta,\beta_2)]^{'}$, $\mathcal{I}_{1,2} := \mathrm{diag}(\mathcal{I}_1\ \mathcal{I}_2)$, $\mathcal{M}_{1,2} := [\mathcal{M}_1^{'}(\theta,\beta_1)\ \mathcal{M}_2^{'}(\theta,\beta_2)]^{'}$ and consider

$$\widetilde{\theta}_{n_1+n_2} = \arg\min_{\theta\in\Theta} m_{12}^{'}\mathcal{I}_{1,2}m_{12}$$

we can write this as

$$\widetilde{\theta}_{n_1+n_2} = \arg\min_{\theta\in\Theta}[m^{'}(\theta,\beta_1)\mathcal{I}_1^{-1}m(\theta,\beta_2) + m_2^{'}(\theta,\beta_2)\mathcal{I}_2^{-1}m_2(\theta,\beta_2)]$$

we linearise this problem around the consistent estimators $\widehat{\theta}_{n_1}$ and $\widehat{\theta}_{n_2}$ and consider

$$\begin{aligned}\widehat{\theta}^{\mathrm{I}}_{n_1+n_2} \quad : \quad &= \arg\min_{\theta\in\theta}\{[m^{'}(\widehat{\theta}_{n_1},\beta_1) + \mathcal{M}_1(\widehat{\theta}_{n_1},\beta_1)(\theta-\widehat{\theta}_{n_1})]^{'}\mathcal{I}_1^{-1}[m^{'}(\widehat{\theta}_{n_1},\beta_1) + \mathcal{M}_1(\widehat{\theta}_{n_1},\beta_1)(\theta-\widehat{\theta}_{n_1})] + \\ &[m^{'}(\widehat{\theta}_{n_2},\beta_2) + \mathcal{M}_2(\widehat{\theta}_{n_2},\beta_2)(\theta-\widehat{\theta}_{n_2})]^{'}\mathcal{I}_2^{-1}[m^{'}(\widehat{\theta}_{n_2},\beta_2) + \mathcal{M}_2(\widehat{\theta}_{n_2},\beta_2)(\theta-\widehat{\theta}_{n_2})]\}\end{aligned}$$

Note that we are back to the well-known case of pooling GLS estimators and therefore

$$\begin{aligned}\widehat{\theta}^{\mathrm{I}}_{n_1+n_2} \quad = \quad & \left(\begin{bmatrix}\mathcal{M}_1(\widehat{\theta}_{n_1},\beta_1)\\ \mathcal{M}_2(\widehat{\theta}_{n_2},\beta_2)\end{bmatrix}^{'}\begin{bmatrix}\mathcal{I}_1 & 0\\ 0 & \mathcal{I}_2\end{bmatrix}^{-1}\begin{bmatrix}\mathcal{M}_1(\widehat{\theta}_{n_1},\beta_1)\\ \mathcal{M}_2(\widehat{\theta}_{n_2},\beta_2)\end{bmatrix}\right)^{-1}\cdot \\ & \begin{bmatrix}\mathcal{M}_1(\widehat{\theta}_{n_1},\beta_1)\\ \mathcal{M}_2(\widehat{\theta}_{n_2},\beta_2)\end{bmatrix}^{'}\begin{bmatrix}\mathcal{I}_1 & 0\\ 0 & \mathcal{I}_2\end{bmatrix}^{-1}\begin{bmatrix}-m(\widehat{\theta}_{n_1},\beta_1) + \mathcal{M}_1(\widehat{\theta}_{n_1},\beta_1)\widehat{\theta}_{n_1}\\ -m(\widehat{\theta}_{n_2},\beta_2) + \mathcal{M}_2(\widehat{\theta}_{n_2},\beta_2)\widehat{\theta}_{n_2}\end{bmatrix}\end{aligned}$$

and thus

$$\begin{aligned}\widehat{\theta}^{\mathrm{I}}_{n_1+n_2} \quad = \quad & [\mathcal{V}_1(\widehat{\theta}_{n_1}) + \mathcal{V}_2(\widehat{\theta}_{n_2})]^{-1}\cdot \\ & [\mathcal{V}_1(\widehat{\theta}_{n_1})\widehat{\theta}_{n_1} + \mathcal{V}_2(\widehat{\theta}_{n_2})\widehat{\theta}_{n_2} - \mathcal{M}_1^{'}(\widehat{\theta}_{n_1},\beta_1)(\mathcal{I}_1)^{-1}m(\widehat{\theta}_{n_1},\beta_1) - \mathcal{M}_2^{'}(\widehat{\theta}_{n_2},\beta_2)(\mathcal{I}_2)^{-1}m(\widehat{\theta}_{n_2},\beta_2)]\end{aligned}$$

By standard arguments on Taylor expansions we have $\widehat{\theta}^{\mathrm{I}}_{n_1+n_2}$ is asymptotically equivalent with $\widetilde{\theta}_{n_1+n_2}$ ■

Proof of theorem 2

The same as the previous proof except a Taylor expansion is taken around $\widehat{\theta}_{n_1}$ ■

Proof of theorem 3

The same as the previous proof except a Taylor expansion is taken around $\widehat{\theta}_{n_1}$ and $\widehat{\theta}^*_{n_2}$ ■

Proof of theorem 4

The same as the previous proof except a Taylor expansion is taken around $\widehat{\theta}^*_{n_2}$ ■

B Appendix

This appendix somewhat generalizes the results of appendix A in Jacquier, Polson and Rossi (1994). The results are based on moment generating functions as found in e.g. Abramowitz & Stegun (1970). Let $\phi := \omega/(1 -$

$\sum_{i=1}^{p} \rho_i)$, and $\tau^2 := \sigma_\eta^2(1+\sum_{j=1}^{q} \zeta_j^2)(1-\sum_{i=1}^{p} \rho_i^2)^{-1}$. For the $\mathsf{ASARMAV}(p,q)$ model we have for $\lambda = 0$:

$$
\begin{aligned}
\mathsf{E}y_t^i &= 0 \text{ for } i \text{ odd} \\
\mathsf{E}y_t^2 &= \mathsf{E}(\sigma_t^2\epsilon_t^2) = \mathsf{E}(\sigma_t^2)\mathsf{E}(\epsilon_t^2) = \exp\{\phi + \frac{1}{2}\tau^2) \\
\mathsf{E}y_t^4 &= \mathsf{E}(\sigma_t^4\epsilon_t^4) = \mathsf{E}(\sigma_t^4)\mathsf{E}(\epsilon_t^4) = 3\exp\{2\phi + 2\tau^2) \\
\mathsf{E}y_t^6 &= \mathsf{E}(\sigma_t^6\epsilon_t^6) = \mathsf{E}(\sigma_t^6)\mathsf{E}(\epsilon_t^6) = 15\exp\{3\phi + \frac{9}{2}\tau^2)
\end{aligned}
$$

Unfortunately for the $\mathsf{ASARMAV}(p,q)$ model for $\lambda \neq 0$ the author does not have a tractable solution yet. For the $\mathsf{SARMAV}(p,q)$-t_ν model we find

$$
\begin{aligned}
\mathsf{E}y_t^i &= 0 \text{ for } i \text{ odd} \\
\mathsf{E}y_t^2 &= \mathsf{E}(\sigma_t^2\epsilon_t^2) = \mathsf{E}(\sigma_t^2)\mathsf{E}(\epsilon_t^2) = \exp\{\phi + \frac{1}{2}\tau^2\}, \nu > 2 \\
\mathsf{E}y_t^4 &= \mathsf{E}(\sigma_t^4\epsilon_t^4) = \mathsf{E}(\sigma_t^4)\mathsf{E}(\epsilon_t^4) = \frac{\nu-2}{\nu-4}9\exp\{2\phi + 2\tau^2), \nu > 4 \\
\mathsf{E}y_t^6 &= \mathsf{E}(\sigma_t^6\epsilon_t^6) = \mathsf{E}(\sigma_t^6)\mathsf{E}(\epsilon_t^6) = \frac{(v-2)^2 15^2}{(v-4)(v-6)}\exp\{3\phi + \frac{9}{2}\tau^2), \nu > 6
\end{aligned}
$$

Therefore for the $\mathsf{SARMAV}(p,q)$ model the kurtosis equals $3e^{\tau^2}$ and for the $\mathsf{SARMAV}(p,q)-\mathrm{t}_\nu$ model it equals $9\frac{\nu-2}{\nu-4}e^{\tau^2}$.

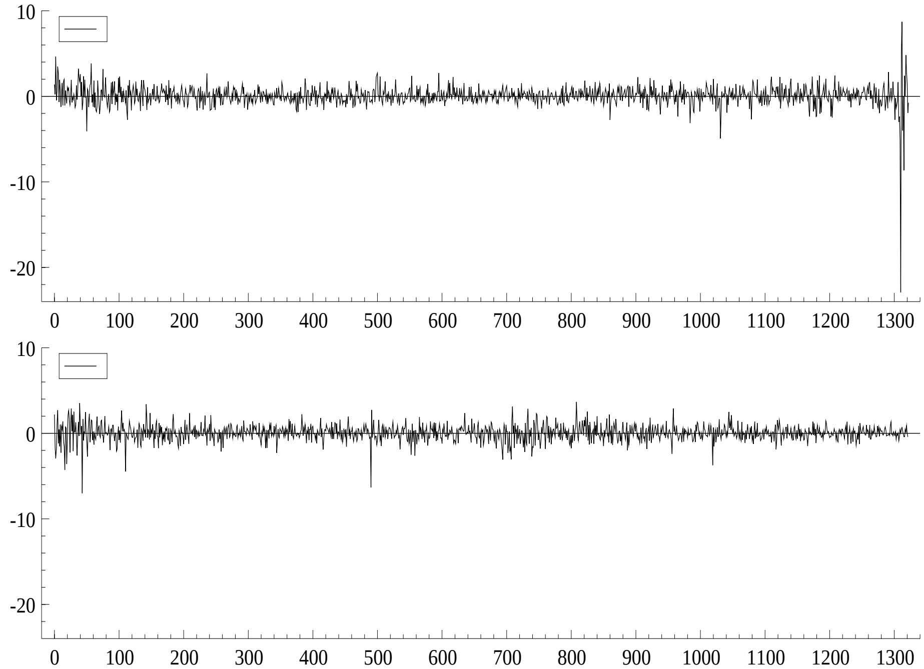

Figure 1: Sample and Post-sample

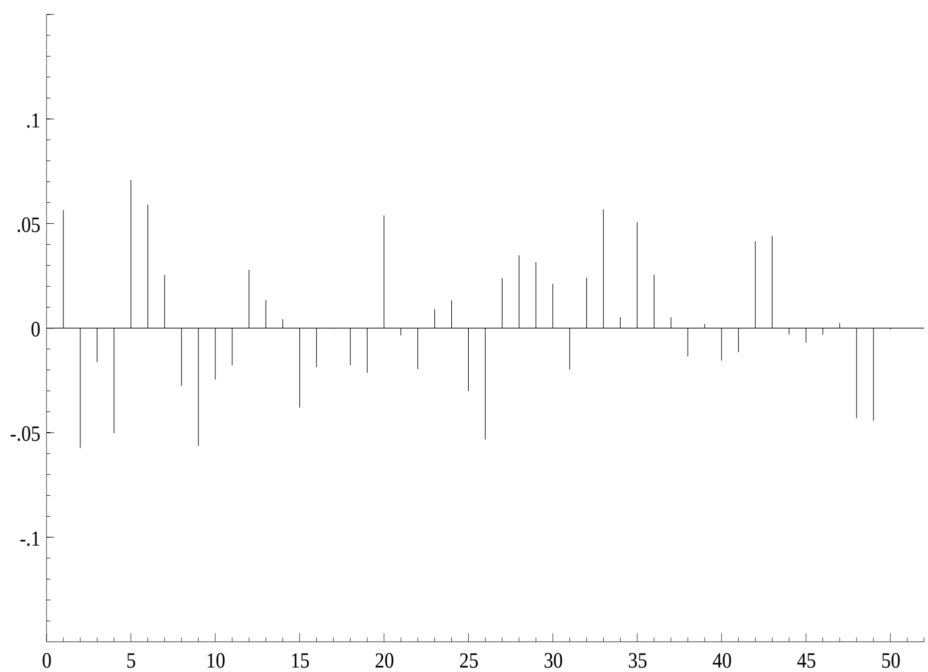

Figure 2a: Correlogram of the sample

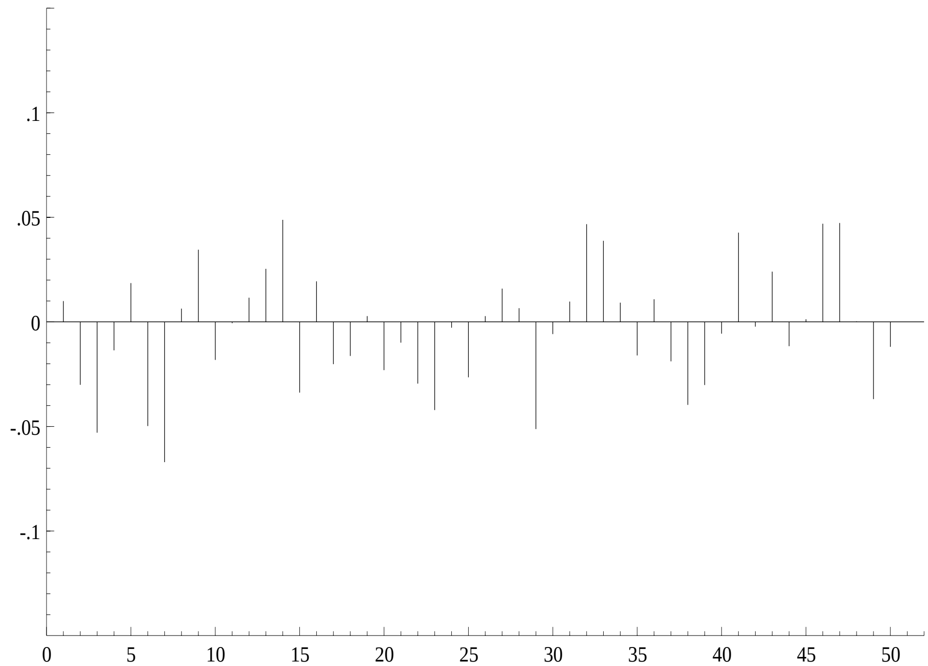

Figure 2b: Correlogram of the post-sample

	sample	post-sample
Mean	$.068$	$.042$
Std.Dv.	1.19	$.934$
Skewness	-5.36	$-.652$
Kurtosis	111	8.50
Minimum	-22.9	-7.01
Maximum	8.71	3.66
Normality Chiˆ2	1894	453

Table 1: Some descriptive statistics of the sample and post-sample

q	K_z	Log Likl	AIC	HQC	BIC
1	1	-1766.0	-1772.0	-1777.8	-1787.5
1	2	-1765.7	-1772.7	-1779.5	-1790.9
1	3	-1748.8	-1756.8	-1764.6	-1777.6
1	4	-1748.6	-1757.6	-1766.3	-1780.9
1	5	-1748.6	-1758.6	-1768.3	-1784.5
1	6	-1721.7	-1732.7	-1743.4	-1761.2
1	7	-1719.6	-1731.6	-1743.3	-1762.7
1	8	-1719.4	-1732.4	-1745.0	-1766.1
1	9	-1757.7	-1771.7	-1785.3	-1808.0
1	10	-1755.6	-1770.6	-1785.1	-1809.5

Table 2: Various model selection criteria for the EGARCH$(1,q)$-H$(K_x,0)$ model for the sample

q	K_z	Log Likl	AIC	HQC	BIC
2	1	-1753.3	-1760.3	-1767.1	-1778.5
2	2	-1753.1	-1761.1	-1768.9	-1781.9
2	3	-1753.4	-1762.4	-1771.1	-1785.7
2	4	-1722.3	-1732.3	-1742.0	-1758.2
2	5	-1718.7	-1729.7	-1740.4	-1758.2
2	6	-1718.4	-1730.4	-1742.1	-1761.5
2	7	-1716.1	-1729.1	-1741.7	-1762.8
2	8	-1711.4	-1725.4	-1739.0	-1761.7
2	9	-1710.6	-1725.6	-1740.2	-1764.5
2	10	-1710.6	-1726.6	-1742.2	-1768.1

Table 3: Various model selection criteria for the EGARCH$(1,q)$-H$(K_z,0)$ model for the sample

	parameters	t-values
α_0	$.028$	5.56
α_1	$-.783$	-14.4
α_2	$.484$	9.57
γ_1	$.959$	77.8
κ_1	$-.149$	-11.9
κ_2	$.336$	18.2
a_{10}	$.049$	3.52
a_{20}	$-.058$	-4.27
a_{30}	$-.004$	$.243$
a_{40}	$.086$	5.75
a_{50}	$-.039$	-2.54

Table 4: Sample estimates for the parameters of the EGARCH$(1,2)$-H$(5,0)$ model.

K_z	Log Likl	AIC	HQC	BIC
1	-1706.6	-1712.6	-1718.5	-1728.2
2	-1706.3	-1713.3	-1720.1	-1731.5
3	-1695.1	-1703.1	-1710.9	-1723.9
4	-1660.5	-1669.5	-1678.2	-1692.8
5	-1659.7	-1669.7	-1679.4	-1695.6
6	-1659.6	-1670.6	-1681.3	-1699.1
7	-1650.4	-1662.4	-1674.1	-1693.5
8	-1641.7	$-1654.7.$	-1667.3	-1688.4
9	-1641.3	-1655.3	-1668.9	-1691.6
10	-1641.3	-1656.3	-1670.8	-1695.2

Table 5: Various model selection criteria for the EGARCH$(1,1)$-H$(K_x,0)$ model for the post-sample

K_z	Log Likl	AIC	HQC	BIC
1	-1703.3	-1710.3	-1717.1	-1728.4
2	-1692.9	-1700.9	-1708.7	-1721.6
3	-1691.4	-1700.4	-1709.1	-1723.7
4	-1656.0	-1666.0	-1675.7	-1691.9
5	-1654.9	-1665.9	-1676.6	-1694.4
6	-1652.3	-1664.3	-1675.9	-1695.4
7	-1646.9	-1659.9	-1672.5	-1693.6
8	-1640.1	-1654.1	-1667.7	-1690.4
9	-1639.7	-1654.7	-1669.6	-1693.6
10	-1639.3	-1655.3	-1670.9	-1696.8

Table 6: Various model selection criteria for the EGARCH$(1, 2)$-H$(K_x, 0)$ model for the post-sample

	parameters	t-values
α_0	$.016$	13.0
α_1	$.475$	13.5
α_2	$-.867$	-26.1
γ_1	$.986$	269
κ_1	$-.107$	-5.26
κ_2	$.127$	12.7
a_{10}	$.029$	2.10
a_{20}	$-.142$	-9.85
a_{30}	$-.003$	$-.17$
a_{40}	$.060$	3.94
a_{50}	$.028$	1.69
a_{60}	$-.090$	-5.79

Table 7: Post-sample estimates for the parameters of the EGARCH$(1, 2)$-H$(6, 0)$ model.

K_z	Log Likl	AIC	HQC	BIC
1	-3494.9	-3500.9	-3507.3	-3518.5
2	-3494.4	-3501.4	-3508.9	-3522.0
3	-3481.7	-3489.7	-3498.3	-3513.3
4	-3398.5	-3407.5	-3417.1	-3434.0
5	-3396.2	-3406.2	-3416.9	-3435.6
6	-3396.1	-3407.1	-3418.8	-3439.5
7	-3391.2	-3403.2	-3416.0	-3438.5
8	-3371.5	-3384.5	-3398.3	-3422.7
9	-3371.5	-3385.5	-3400.4	-3426.6
10	-3371.4	-3386.4	-3402.4	-3430.5

Table 8: Various model selection criteria for the EGARCH$(1,1)$-H$(K_x,0)$ model for the combined sample and post-sample

K_z	Log Likl	AIC	HQC	BIC
1	-3484.8	-3491.8	-3499.2	-3512.3
2	-3484.4	-3492.4	-3500.9	-3515.9
3	-3465.0	-3474.0	-3483.6	-3500.4
4	-3391.3	-3401.3	-3412.0	-3430.7
5	-3388.0	-3399.0	-3410.7	-3431.4
6	-3387.3	-3399.3	-3412.1	-3434.6
7	-3376.4	-3389.4	-3403.2	-3427.6
8	-3364.2	-3378.2	-3393.1	-3419.4
9	-3363.9	-3378.9	-3394.9	-3423.0
10	-3363.9	-3379.9	-3396.9	-3426.9

Table 9: Various model selection criteria for the EGARCH$(1,2)$-H$(K_x,0)$ model for the combined sample and post-sample

	parameters	t-values
α_0	$.039$	22.9
α_1	$-.210$	-8.69
α_2	$-.195$	-8.32
γ_1	$.977$	265
κ_1	$-.129$	-12.0
κ_2	$.240$	22.4
a_{10}	$.038$	3.83
a_{20}	$-.164$	-16.1
a_{30}	$-.014$	-1.39
a_{40}	$.103$	10.2
a_{50}	$-.026$	-2.29
a_{60}	$-.033$	-2.91
a_{70}	$-.020$	-1.66
a_{80}	$.049$	4.32

Table 10: EGARCH$(1,2)$-H$(8,0)$ model for the combination of sample and post-sample.

Estimation Technique	ω	ρ	σ_η
Bayesian	$-.007$	$.979$	$.136$
I	$-.008$	$.975$	$.105$
II	$-.008$	$.975$	$.104$
IV	$-.012$	$.974$	$.115$
V	$-.007$	$.969$	$.106$

Table 11: Computationally attractive estimates of the SARMAV$(1,0)$ model

Model	mean	std.dv	skewness	kurtosis	minimum	maximum
SARMAV$(1,0)$	0	.9398	0	3.663	-5.074^{*}	5.821^{*}
ASARMAV$(1,0)$	0	.9909	$-.2938^{*}$	4.881^{*}	-7.666^{*}	5.844^{*}

Table 12: Salient features of the structural models considered in the paper. For the SARMAV(p,q) and ASARMAV(p,q) of the mean, standard deviations, skewness and kurtosis are known, see appendix B, so here the theoretical values at the optimum are reported. Otherwise the statistics are calculated from a two simulated antithetic series of size 50,000, these are marked with an asterisk.

Model	J	P- value	df
SARMAV$(1,0)$	33.8	.000	8
SARMAV$(1,1)$	33.6	.000	7
ASARMAV$(1,0)$	24.0	.001	7

Table 13: Hansen J test for overidentifying restrictions for several structural models

Model	PSP	P- value	df
SARMAV$(1,0)$	32.06	.0014	12

Table 14: PSP-test for several structural models

moments	J	PSP
α_0	$.133$	$.083$
α_1	-1.52	$-.159$
α_2	-2.06	$-.099$
γ_1	-2.30	$-.015$
κ_1	1.85	$.295$
κ_2	$-.883$	$.052$
a_{10}	-3.57	-1.33
a_{20}	-2.46	$.011$
a_{30}	$-.367$	$.124$
a_{40}	-2.40	$-.109$
a_{50}	2.74	$-.168$
a_{60}	-	$.289$

Table 15: t-values for the individual elements of the J test and PSP test for the SARMAV$(1,0)$ model

moments	J
α_0	-1.14
α_1	$.579$
α_2	$-.608$
γ_1	$-.558$
κ_1	$.658$
κ_2	$-.420$
a_{10}	-3.48
a_{20}	-2.79
a_{30}	$-.390$
a_{40}	-1.83
a_{50}	2.56

Table 16: t-values for the individual elements of the J test and PSP test for the ASARMAV$(1,0)$ model

	$\mathsf{JSS}^*(\widehat{\theta}_{n_1})$	$\mathsf{JSS}^*(\widehat{\theta}^{I}_{n_2})$	$\mathsf{JSS}^*(\widehat{\theta}^{II}_{n_2})$	$\mathsf{JSS}^*(\widehat{\theta}^{IV}_{n_2})$	$\mathsf{JSS}^*(\widehat{\theta}^{V}_{n_2})$	$\mathsf{JSS}^*(\widehat{\theta}^{B}_{n_2})$
X	67.3	53.2	54.1	53.2	91.4	235
P value	.000	.000	.000	.000	.000	.000
df	20	20	20	20	20	20

Table 17: Test statistics for structural stability of the optimal stochastic volatility model

$\mathsf{W}^* = \mathsf{LR}^*(\widehat{\theta}_{n_1})$	$\mathsf{LR}^*(\widehat{\theta}^{I}_{n_2})$	$\mathsf{LR}^*(\widehat{\theta}^{II}_{n_2})$	$\mathsf{LR}^*(\widehat{\theta}^{IV}_{n_2})$	$\mathsf{LR}^*(\widehat{\theta}^{V}_{n_2})$	$\mathsf{LR}^*(\widehat{\theta}^{B}_{n_2})$
.218	3.30	3.78	3.30	.070	21.0
.988	.348	.286	.348	.995	.000
3	3	3	3	3	3

Table 18: Test statistics for structural stability of the optimal stochastic volatility model