ELSEVIER

Journal of Econometrics 91 (1999) 61–87

JOURNAL OF
Econometrics

# Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study

Torben G. Andersen[a], Hyung-Jin Chung[b], Bent E. Sørensen[b,*]

[a] *Department of Finance, Kellogg Graduate School of Management, Northwestern University, USA*

[b] *Department of Economics, Brown University, Box B, Providence, RI 02912, USA*

Received 1 May 1997; received in revised form 1 March 1998

**Abstract**

We perform an extensive Monte Carlo study of efficient method of moments (EMM) estimation of a stochastic volatility model. EMM uses the expectation under the structural model of the score from an auxiliary model as moment conditions. We examine the sensitivity to the choice of auxiliary model using ARCH, GARCH, and EGARCH models for the score as well as nonparametric extensions. EMM efficiency approaches that of maximum likelihood for larger sample sizes. Inference is sensitive to the choice of auxiliary model in small samples, but robust in larger samples. Specification tests and '*t*-tests' show little size distortion. 

*JEL classification:* C15; C22

*Keywords:* Stochastic volatility; GMM; EMM; Monte Carlo

## 1. Introduction

The modeling of return volatility continues to inspire and challenge financial econometricians. The work is spurred on by the overwhelming empirical evidence of strong conditional heteroskedasticity in almost all high-frequency financial return series. Time-varying and highly persistent volatility complicates asset pricing since it implies time-varying risk premiums. Moreover, correct modeling of volatility is essential for valuation of derivative assets like options and

* Corresponding author. E-mail: bs@econ.pstc.brown.edu


PII: S0304-4076(98)00049-9

warrants. In spite of the voluminous literature, a great deal of controversy still surrounds the selection of an appropriate model for volatility and the associated choice of estimation strategy. This stems in large part from the continued failure to rationalize, from economic theory, the existence of such strong and systematic variation in price variability without resorting to a corresponding time variation in underlying (unobserved) economic fundamentals. The lack of theoretical guidance has left the field open to genuine competition among various statistical approaches, each trying to fit and forecast volatility better than the other. The fact that a number of key issues remain hotly debated is testimony to the complexity of the endeavor, which involves fitting highly nonlinear models to the conditional second return moments. This feature sets the field apart from traditional time-series analysis, and has fueled a tremendous amount of new research into alternative tools for inference in nonlinear models.

In light of the above observations, it is desirable to pursue an estimation strategy that accommodates a large class of alternative models, readily affords further generalizations, allows for simple and useful model diagnostics suggestive of the dimensions along which the model may falter, provides a simple overall model specification test, accommodates both continuous- and discrete-time specifications with relative ease, and provides efficient inference. We shall argue that the efficient method of moments (EMM) approach, introduced by Bansal et al. (1993), Bansal et al. (1995) and Gallant and Tauchen (1996), may possess desirable properties across this diverse set of objectives. Unfortunately, the finite sample properties of EMM are largely unknown.

The purpose of this article is to undertake an extensive investigation of the EMM procedure through Monte Carlo techniques. Because one main objective is to gauge the performance of the procedure, not only in an absolute sense, but also relative to alternative approaches, we focus on a setting that affords direct comparison to prior contributions. Specifically, we consider a simple version of the so-called lognormal stochastic volatility model which has served as an unofficial testing ground for such analyses. In fact, the model is so commonly invoked that it is frequently referred to as *the* stochastic volatility model, although it is a special case, both in terms of the functional form and the assumed distributional properties. The model is nonetheless attractive because of its parsimony, and because it provides a reasonable first approximation to the properties of most financial return series. Moreover, it retains the fundamental inference problem associated with the presence of an unobserved latent volatility factor, which is absent in ARCH models.

The lognormal stochastic volatility (SV) model has been estimated by a variety of means, including simple moment matching (MM) (Taylor, 1986), generalized method of moments (GMM) (Melino and Turnbull, 1990), simulated method of moments (SMM) (Duffie and Singleton, 1989), quasi-maximum likelihood (QML) (Harvey et al., 1994), Bayesian Markov-Chain Monte Carlo analysis (MCMC) (Jacquier et al., 1994; henceforth JPR), indirect inference principles (Gouriéroux et al., 1993), efficient method of moments (Gallant et al., 1997), Bayesian importance-sampling Monte Carlo (Geweke, 1994), a unified Markov-Chain Monte-Carlo sampling-based framework for Bayesian and maximum likelihood inference (Kim et al., 1998), simulation-based maximum likelihood (SML) (Danielsson, 1994; Danielsson and Richard, 1993), maximum likelihood Monte Carlo (MCL) (Sandmann and Koopman, 1996), and direct maximum likelihood through recursive numerical integration (ML) (Fridman and Harris, 1998). Although MM, GMM, and QML are simple to implement, the more elaborate and computationally intensive procedures may well be justified, as the associated efficiency gains have been shown to be substantial in many cases.

A number of Monte Carlo studies have explored the small sample properties of these estimators. Andersen and Sørensen (1996), henceforth AS, perform an extensive study of GMM,[1] while Harvey et al. (1994) examine the QML estimator. These simple procedures are about equally efficient, with the relative performance being dependent on the specific parameter values, see also Andersen and Sørensen (1997). More limited Monte Carlo studies have been undertaken for the more computationally intensive techniques such as MCMC, ML, MCL, and SML. The original study is JPR, who show that the MCMC estimator strongly dominates GMM and QML for estimating the stochastic volatility model. Subsequently, several alternative techniques have been found to match the benchmark efficiency established by JPR, see, e.g., Fridman and Harris (1998) and Sandmann and Koopman (1996). Moreover, it is now apparent that these methods effectively provide a likelihood-based inference. Consequently, while the procedures differ in terms of ease and speed of implementation, the issue of efficient inference is resolved for this particular model. Hence, attention should probably turn towards the issues raised earlier, such as whether the alternative estimation techniques readily accommodate relevant extensions to the functional form and distributional assumptions, whether they provide useful model diagnostics, and whether they are flexible enough to allow inference in both continuous-time and discrete-time settings. This is important because, although the 'plain-vanilla' stochastic volatility model may provide a good first approximation to the returns process, it is typically far from perfect. The EMM approach has already proven effective along a number of dimensions. Thus, our interest in EMM is not driven strictly by estimation performance, but also by the wide applicability. Nonetheless, reliable inference in the stochastic volatility environment provides an important litmus test for the approach.

GMM is relatively inefficient due to the largely arbitrary choice of unconditional moments that can be computed in closed form, while likelihood-based

[1] For the remainder of this paper, we will use the term GMM to refer to the GMM implementation of Melino and Turnbull (1990), that was studied by JPR and AS.

procedures achieve the Cramer–Rao efficiency bound by relying on the optimal moments (scores) within a restricted parametric setting. EMM seeks efficiency improvements, while maintaining the general flexibility of GMM, by letting the data guide the choice of an auxiliary quasi-likelihood which serves to generate an efficient set of moments. By construction, the efficiency of EMM is thus likely to fall somewhere between GMM and the (often infeasible) likelihood-based procedures.[2] Our study investigates the relative efficiency of EMM, as well as its sensitivity to the specific features of the implementation in the finite-sample context.

The paper is organized as follows. Section 2 introduces the lognormal stochastic volatility model which generates the simulated data. Section 3 describes the EMM procedure, while Section 4 specifies the Monte Carlo simulation design. Section 5 presents the results, and Section 6 concludes.

## 2. The stochastic volatility model

The univariate lognormal stochastic autoregressive volatility model for the return series $y_t$ is[3]

$$y_t = \sigma_t z_t, \tag{1}$$

$$\ln \sigma_t^2 = \alpha + \beta \ln \sigma_{t-1}^2 + \sigma_u u_t, \tag{2}$$

where $t = 1, \dots, T$. $T$ is the sample size, $\rho = (\alpha, \beta, \sigma_u)$ constitutes the parameter vector, and $\{z_t, u_t\}$ is i.i.d. N(0, $I$). We impose the inequality constraints $-1 < \beta < 1$ and $\sigma_u > 0$, which ensure that $y_t$ is stationary and ergodic, and that the parameters are uniquely identified. The latent volatility process $\sigma_t^2$ follows an AR(1) in logarithms and induces higher-order moment dependence in $y_t$. The parameter $\beta$ measures the volatility persistence and is typically estimated to be less than, but relatively close to, unity in empirical studies. Finally, the model generates a leptokurtic unconditional distribution that is consistent with the prevalence of outliers in financial data.

The system defined by Eqs. (1) and (2) and the (true) parameter vector, $\rho_0$, determines the probabilistic structure of the observed data, $y_t$. We refer to this

[2] It is worth noting that the set of models, for which direct likelihood-based inference is feasible, is expanding rapidly due to the computational revolution documented in the current statistics literature, see, e.g., Kim et al. (1998) for illustration and discussion.

[3] Besides lognormal stochastic autoregressive volatility (SARV) (Andersen, 1994), the model is also labeled autoregressive random variance (Taylor, 1994) and stochastic variance (Harvey et al., 1994).

data generating process as the *structural model*. Our parameterization of the structural model ignores the possibility of a non-zero, and potentially time-varying, mean as in Engle et al. (1987). Further, the independence of $z_t$ and $u_t$ precludes the asymmetric 'leverage effect' of Black (1976), who argues that future volatility may be negatively correlated with returns. Such features are readily handled by EMM, but for ease of comparison with earlier studies, we focus on the simpler version of the model.

## 3. EMM estimation

EMM is a method of moment procedure that often provides a viable approach to estimation when maximum likelihood is computationally intensive or infeasible. EMM is particularly appealing in the context of dynamic latent variable models, where evaluation of the likelihood involves integration over the (partially) unobserved realization of the state vector. For example, the log-likelihood for the stochastic volatility model is readily expressed conditional on the realization of the volatility process, $\sigma_t^2$, but since the volatility is not observed, this serially correlated latent factor must be integrated out of the likelihood. The dimension of the associated integral is equivalent to sample size, so direct evaluation of the likelihood is extremely cumbersome, if not infeasible. This explains why several authors resort to simulation-based approximations to the likelihood or avoid direct dependence on the likelihood altogether. EMM opts for the second solution, but still seeks to mimic the efficiency of likelihood-based inference. The key insight is that a careful selection of moment conditions, guided by the characteristics of the observed data, will allow for efficient estimation via a standard GMM procedure.

### 3.1. The EMM estimation procedure

Maximum likelihood may itself be interpreted as a method of moment procedure with the derivative of the log-likelihood function, *the score vector*, providing the (exactly identifying) moment conditions. Since an analytical expression for the likelihood is not available for the stochastic volatility model, EMM employs an auxiliary model, or score generator, that allows for a closed-form expression for the associated (quasi-) score vector. The auxiliary model is given by a conditional density parameterized by the auxiliary parameter vector $\eta$. In this context the conditional density may be expressed as $f(y_t \mid Y_{t-1}, \eta)$, where $Y_t = \{y_t, \ldots, y_1\}$ denotes the complete return history and thus constitutes the relevant information set for the model.

The initial EMM step is to estimate $\eta$ by quasi-maximum likelihood, which ensures that the quasi-maximum likelihood (QML) estimator, $\hat{\eta}_T$ satisfies the associated first-order conditions,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\eta}\ln f(y_t \mid Y_{t-1},\hat{\eta}_T)=\frac{1}{T}\sum_{t=1}^{T} s_f(Y_t,\hat{\eta}_T)=0, \tag{3}$$

where $s_f(Y_t,\hat{\eta}_T)=(\partial/\partial\eta)\ln f(y_t \mid Y_{t-1},\hat{\eta}_T)$ denotes the quasi-score function. Even if the auxiliary model is misspecified, standard QML theory, cf. White (1994), implies that, under suitable regularity, $\hat{\eta}_T \to \eta_0$, where the limiting value, $\eta_0$, is denoted the quasi-true value of $\eta$.

Next, EMM inverts the auxiliary parameter estimate, or rather the associated score function in Eq. (3), to obtain a consistent estimate of the structural parameter, $\rho$, in a second GMM-based step. The left-hand side of Eq. (3) is simply the sample average of the quasi-score function evaluated at $\hat{\eta}_T$ and thus provides an estimate of the expected value of the auxiliary score. EMM uses the corresponding (population) expectation under the structural model of the score from the auxiliary model as moment conditions. Since the expected quasi-score is defined under the probability measure induced by the structural model, $P(Y_t \mid \rho)$, the moments depend directly on the structural parameters. Identification requires that the dimension of the quasi-score, i.e., the number of parameters in $\eta$, $n_\eta$, exceeds that of the structural parameter vector, $n_\rho$, but otherwise the auxiliary model need not have anything to do with the structural model. However, as with any GMM-based procedure, the choice of moments is critical for efficiency. We discuss the efficiency issue further in Section 3.2.

The population moments that identify the structural parameters are

$$m(\rho,\eta_0)=\mathrm{E}_\rho[s_f(Y_t,\eta_0)]=\int s_f(Y_t,\eta_0)\,\mathrm{d}P(Y_t,\rho). \tag{4}$$

The presence of a latent variable renders an analytical expression for this moment condition infeasible. Consequently, the sample moments are computed by Monte Carlo integration. Hence, the second EMM step is effectively an application of the Simulated Method of Moments of Duffie and Singleton (1993). A simulated series $\hat{y}_n(\rho)$, $n=1,\dots,N$, is generated from the structural model for a given $\rho$ and used in evaluating the sample moments at the fixed QML estimate, $\hat{\eta}_T$:

$$m_N(\rho,\hat{\eta}_T)=\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\eta}\ln f(\hat{y}_n(\rho) \mid \hat{Y}_{n-1}(\rho),\hat{\eta}_T). \tag{5}$$

As $N\to\infty$, $m_N(\rho,\hat{\eta}_T)\to m(\rho,\hat{\eta}_T)$ almost surely. Thus, for a large enough simulated sample, the Monte Carlo error becomes negligible, and we ignore this error in the following derivation. Since $\hat{\eta}_T$ is available from the QML step and the quasi-score is given in analytic form, the evaluation of Eq. (5) is straightforward. The dimension of the score vector typically exceeds that of the structural parameter vector, so that the score vector cannot be forced to zero. Instead, *the* GMM criterion in the moment vector is minimized to obtain the EMM estimator of $\rho$:

$$\hat{\rho}_T = \underset{\rho}{\operatorname{argmin}}\ [m_N(\rho, \hat{\eta}_T)' \hat{I}_T^{-1} m_N(\rho, \hat{\eta}_T)], \tag{6}$$

where $\hat{I}_T$ denotes a consistent estimator of the asymptotic covariance matrix, $I$, of the sample quasi-score vector, i.e., the quasi-information matrix. If the auxiliary model is expanded to the point where it accommodates all main systematic features of the data, likelihood theory implies that the quasi-scores constitute a (near) martingale difference sequence, and a convenient estimator of the quasi-information matrix is obtained from the outer product of the scores:

$$\hat{I}_T = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \eta} \ln f(y_t \mid Y_{t-1}, \hat{\eta}_T) \frac{\partial}{\partial \eta} \ln f(y_t \mid Y_{t-1}, \hat{\eta}_T)'. \tag{7}$$

Notice that $\hat{I}_T$ may be obtained directly from the first QML step, avoiding the need for computation of the weighting matrix during the second GMM-based estimation step.

Gallant and Tauchen (1996) show that, under suitable regularity, the EMM estimator is consistent and asymptotically normal. Specifically,

$$\sqrt{T}(\hat{\rho}_T - \rho_0) \overset{\text{as}}{\to} \mathrm{N}\{0, [D_\rho' I^{-1} D_\rho]^{-1}\}, \tag{8}$$

where $D_\rho = (\partial/\partial\rho')m(\rho_0, \eta_0)$. Moreover, the asymptotic variance-covariance matrix may be estimated consistently by

$$\mathrm{Cov}(\hat{\rho}_T) = \frac{1}{T}\left[\frac{\partial m_N(\hat{\rho}_T, \hat{\eta}_T)'}{\partial \rho} \hat{I}_T^{-1} \frac{\partial m_N(\hat{\rho}_T, \hat{\eta}_T)'}{\partial \rho'}\right]^{-1}. \tag{9}$$

The only new quantity to estimate after the second estimation step is the derivative of the quasi-score with respect to the structural parameter, which is readily done by numerical means. This formula now provides the basis for inference regarding the structural parameter.

As usual in GMM, a test of the over-identifying restrictions may be obtained directly from the criterion function. Under the null hypothesis of correct model specification, $T$ times the minimized value of the EMM objective function is distributed $\chi^2$ with $n_\eta - n_\rho$ degrees of freedom. If the test rejects, the individual elements of the score vector may provide useful information regarding the dimensions in which the structural model fails to accommodate the data. These model diagnostics are based on the standard $t$-statistics of the individual elements of the score vector, $m_N(\hat{\rho}_T, \hat{\eta}_T)$, see, e.g., Tauchen (1996) for details.

### *3.2. Choice of auxiliary model*

As noted above, EMM delivers consistent estimates of the structural parameter vector under weak conditions on the choice of the auxiliary model. However, extrapolating from the GMM evidence, one may suspect that the choice of moments (auxiliary model) is critical for estimation efficiency. For example, it is natural to conjecture that the quality of inference may hinge on how well the auxiliary model approximates the salient features of the observed data. This intuition can be formalized. Gallant and Long (1997) show that a judicious selection of the auxiliary model, ensuring that the quasi-scores asymptotically span the true score vector, will result in full asymptotic efficiency. Effectively, as the score generator approaches the true conditional density, the estimated covariance matrix for the structural parameter approaches that of maximum likelihood. This result embodies one of the main advantages of EMM. It prescribes a systematic approach to the derivation of efficient moment conditions for estimation in a general parametric setting.

Although the selection of the auxiliary model in principle is important for estimation performance, the literature has not explored the relevance of this issue in a systematical manner. We consequently parameterize the auxiliary model in several ways, including both fully parametric and some seminonparametric specifications, and investigate the resulting finite-sample efficiency. Our fully parametric score generators are all conditionally Gaussian. Allowing for a non-zero mean, $\mu$, and defining $z_t = (y_t - \mu)/\sigma_t = \varepsilon_t/\sigma_t$, they take the general form

$$f(y_t \mid Y_{t-1}, \eta) \propto \frac{1}{\sigma_t} \exp\left( -\frac{1}{2} z_t^2 \right). \tag{10}$$

The distinguishing feature of the stochastic volatility model is, of course, the conditional heteroskedasticity. We allow for time-dependence in the auxiliary model by letting $\sigma_t^2$ follow several popular ARCH-type specifications from the extant literature. Specifically, we investigate the parameterizations: ARCH($q$) (Engle, 1982), GARCH(1,1) (Bollerslev, 1986), and EGARCH(1,1) (Nelson, 1991), as well as restricted versions of these models. The ARCH($q$) has been used extensively in the EMM literature, but the use of the non-Markovian GARCH and EGARCH specifications has only recently been justified by Gallant and

Long (1997), and successfully employed in Andersen and Lund (1997a,b). The functional forms are

$$\text{ARCH}(q)\text{:}\quad \sigma_t^2 = \omega + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2, \tag{11}$$

$$\text{GARCH(1,1):}\quad \sigma_t^2 = \omega + b\sigma_{t-1}^2 + a\varepsilon_{t-1}^2, \tag{12}$$

$$\text{EGARCH(1,1):}\quad \ln \sigma_t^2 = \omega + b \ln \sigma_{t-1}^2 + (1 + aL)\left[\phi z_{t-1} + \gamma\left\{|z_{t-1}| - \sqrt{\frac{2}{\pi}}\right\}\right]. \tag{13}$$

Gallant and Tauchen (1996) suggest incorporating the semi-nonparametric (SNP) density of Gallant and Nychka (1987) within the auxiliary model, and we perform a number of experiments exploring the use of such SNP representations. An effective approach is to include a leading parametric term to account for the bulk of the dependency in the conditional mean and variance, and then allow a squared Hermite polynomial (SNP-) expansion to accommodate any remaining non-Gaussianity and time series structure in the innovation process. Care must be taken, however, to avoid overparameterization of the auxiliary model, as convergence problems may arise if the quasi-score is extended to the point where it begins to fit the purely idiosyncratic noise in the data.

The SNP model is given by the following parameterization:

$$f_K(y_t \,|\, \Omega_{t-1}, \eta) = \frac{1}{\sigma_t} \frac{[P_K(z_t)]^2\, \phi(z_t)}{\int_{-\infty}^{\infty} [P_K\ (u)]^2 \phi(u)\, \mathrm{d}u}, \tag{14}$$

where $z_t = \varepsilon_t/\sigma_t$, $\phi(\cdot)$ denotes the standard normal density, and the normalization factor $\int_{-\infty}^{\infty}[P_K(u)]^2\phi(u)\,\mathrm{d}u$ ensures that the SNP density integrates to unity. We explore cases where $\sigma_t$ follows one of the ARCH models described above, and also the case where $\sigma_t$ is constant and all heteroskedasticity thus is modeled nonparametrically. The Hermite polynomial is given by

$$P_K(z_t) = \sum_{i=0}^{K_z} a_i z_t^i, \tag{15}$$

where $a_0 = 1$ for identification purposes. $K_z$ denotes the order of the polynomial expansion that controls the extent to which the tails deviate from normality. If $K_z = 0$, the SNP reduces to the normal density. Meanwhile, the conditional mean and variance is governed by the underlying parameterization of $\varepsilon_t$ and $\sigma_t$. Additional dynamic features may be accommodated by letting the $a_i$-coefficients be polynomials in the variables of the information set, as explained in Gallant and Tauchen (1996). For example, we may allow the immediately preceding return to impact the conditional distribution via

$$a_i(y_{t-1}) = \sum_{j=0}^{K_x} a_{ij} y_{t-1}^{j}. \tag{16}$$

When $K_x = 0$, the innovations $\{z_t\}$ are homogeneous, as the conditional density is independent of $Y_{t-1}$. For $K_x > 0$, we effectively multiply the innovations by functions of past observations.

## 4. Monte Carlo setup

The simulations are performed using the Gauss and C++ computer languages on RISC 6000 workstations and Pentium PCs. For the first step, the quasi-likelihood is maximized using the BHHH and Newton algorithm, employing several different starting values to avoid local optima. The latter constitute a serious concern in higher-order SNP models, so a judicious choice of initial conditions is critical to conserve on computing time in the simulation setting. In the second EMM estimation step, a long simulated series, $\hat{Y}_N$ is generated from the structural lognormal stochastic autoregressive volatility model, and then used to evaluate the quasi-score vector at the fixed QML estimate, $\hat{\eta}_T$. The parameter vector is given by $\rho = (\alpha, \beta, \sigma_u)$ and one efficient starting value is simply the true value of $\rho$, denoted $\rho_0$.[4] We use the BFGS algorithm to minimize the EMM criterion function.

The (simulated) data were generated using the parameter values from JPR and AS, in order to compare our results to each of these studies. A majority of the simulations is based on $\beta = 0.90$, but we also include several designs with $\beta = 0.98$. The parameter vectors are $(\alpha, \beta, \sigma_u) = (-0.736, 0.90, 0.363)$ and $(\alpha, \beta, \sigma_u) = (-0.147, 0.98, 0.166)$.

The first parameter constellation is calibrated to typical results for weekly return series, while the second is more reflective of findings at the daily frequency. Given these values, we generate samples of length $T = 500$, $T = 1000$, $T = 2000$, and $T = 4000$, with the initial value of the volatility process for each sample being drawn from the stationary distribution of $\sigma_t^2$. We perform 500 Monte Carlo simulations for each combination of score generator and $T$.

[4] For low-dimensional SNP models, starting values affect mainly the convergence time, but for higher-dimensional SNP score generators, poorly chosen starting values often imply that the algorithm fails to converge.

The Monte Carlo integration required in equation (6) was generally performed using the average sample score vectors obtained from two simulated series of the structural model, each of length $N = 20{,}000$. The two series are identical, except that the first calculates $\hat{Y}_N$ using the innovation sequence $\{z_n, u_n\}$, while the second relies on $\{-z_n, -u_n\}$. This antithetic variates technique induces negative correlation in the two estimates of the integral and, as noted in Andersen and Lund (1997a), is highly efficient in reducing the associated Monte Carlo error.

## 5. Results

To compare our findings to prior research, we focus on the design with the lower degree of volatility persistence, $\beta = 0.90$. Tables 1–3 display the mean EMM-parameter estimates and associated root mean squared errors (RMSEs) for sample sizes between 500 and 4000 and alternative auxiliary models. For convenience, corresponding results from the most successful GMM estimation

Table 1
$T = 500$, Simulated mean and root mean square error

| Score generator | $\alpha$ | $\beta$ | $\sigma_u$ |
|---|---|---|---|
| (1) EGARCH(1,0) (no const.) (3 moments) | − 0.96 (0.70) | 0.87 (0.09) | 0.40 (0.21) |
| (2) GARCH(1,1) (4 moments) | − 0.91 (0.60) | 0.88 (0.08) | 0.38 (0.20) |
| (3) GARCH(1,1) (diagonal weight) (4 moments) | − 0.79 (0.69) | 0.89 (0.10) | 0.37 (0.36) |
| (4) EGARCH(1,0) (4 moments) | − 0.95 (0.63) | 0.87 (0.08) | 0.39 (0.20) |
| (5) EGARCH(1,1) (6 moments) | − 1.31 (4.42) | 0.85 (0.15) | 0.39 (0.28) |
| (6) GMM (crashes ignored) (14 moments) | − 0.62 (0.59)* | 0.92 (0.08)* | 0.24 (0.17)* |
| (7) Infeasible GMM (true weight) (24 moments) | − 1.13 (1.15) | 0.85 (0.14) | 0.39 (0.13) |

True parameters $(\alpha, \beta, \sigma_u) = (-0.736, 0.90, 0.363)$, 500 Monte Carlo iterations.
For each score generator, we report the mean and the root mean square error in parentheses.
The GMM results are from Andersen and Sørensen (1996), Tables 5 and 3, resp.
*The bias and RMSE in line (6) are calculated for 1000 converged simulations, with 342 crashes ignored. Hence, these RMSEs are not comparable to the other rows.
'Diagonal weight' refers to simulations where the off-diagonal terms in the weighting matrix are set to 0.
'True weight' refers to simulations where the true long run weighting matrix has been used.

Table 2
$T = 1000$, Simulated mean and root mean square error

| Score generator | $\alpha$ | $\beta$ | $\sigma_u$ |
|---|---|---|---|
| (1) ARCH(1) (3 moments) | − 1.33 (2.17) | 0.84 (0.20) | 0.48 (0.59) |
| (2) GARCH(1,1) (4 moments) | − 0.81 (0.35) | 0.89 (0.05) | 0.37 (0.12) |
| (3) EGARCH(1,0) (4 moments) | − 0.83 (0.38) | 0.89 (0.05) | 0.38 (0.13) |
| (4) ARCH(2) (4 moments) | − 1.07 (1.18) | 0.86 (0.15) | 0.40 (0.32) |
| (5) EGARCH(1,1) (6 moments) | − 0.83 (0.44) | 0.89 (0.06) | 0.34 (0.16) |
| (6) GARCH(1,1) − Kz(2) (6 moments) | − 0.80 (0.33) | 0.89 (0.04) | 0.33 (0.12) |
| (7) ARCH(5) (7 moments) | − 1.05 (0.96) | 0.86 (0.10) | 0.39 (0.27) |
| (8) GARCH(1,1) − Kz(4) (8 moments) | − 0.84 (0.36) | 0.89 (0.05) | 0.35 (0.10) |
| (9) GMM (crashes ignored) (14 moments) | − 0.57 (0.41)* | 0.92 (0.06)* | 0.25 (0.15)* |
| (10) Infeasible GMM (true weight) (24 moments) | − 0.93 (0.66) | 0.87 (0.08) | 0.38 (0.09) |

True parameters $(\alpha, \beta, \sigma_u) = (-0.736, 0.90, 0.363)$, 500 Monte Carlo iterations.
For each score generator, we report the mean and the root mean square error in parentheses.
The GMM results are from Andersen and Sørensen (1996), Tables 5 and 3, resp.
*The bias and RMSE in line (9) are calculated for 1000 converged simulations, with 77 crashes ignored. These RMSEs are not comparable to the other rows.

procedures in AS are also included at the bottom of the tables. AS find that GMM frequently fails to converge for low sample sizes, and that these convergence problems are caused by imprecise estimates of the long-run covariance matrix (and therefore of the estimated GMM-weighting matrix). These problems disappear if an (in practice infeasible) accurate approximation to the true weighting matrix is utilized. In order to highlight the extent to which the difference between GMM and EMM results is due to an imprecise GMM weighting matrix we further include results for the infeasible GMM estimator. Figs. 1–3 provide complementary evidence on the distribution of EMM-$\beta$ parameter estimates for different auxiliary models and sample sizes. Table 4 reports on the case of higher volatility persistence, $\beta = 0.98$, while Table 5 compares results across a wider range of estimation procedures, including some recently developed techniques that provide (near) likelihood-based inference. Finally, Figs. 4 and 5 characterize the quality of the EMM-specification test for goodness-of-fit and EMM-based inference regarding the volatility persistence

Table 3
$T = 4000$, simulated mean and root mean square error

| Score generator | $\alpha$ | $\beta$ | $\sigma_u$ |
|---|---|---|---|
| (1) EGARCH(1,0) (4 moments) | − 0.764 (0.158) | 0.896 (0.021) | 0.368 (0.049) |
| (2) GARCH(1,1) (4 moments) | − 0.764 (0.153) | 0.896 (0.020) | 0.371 (0.050) |
| (3) EGARCH(1,1) (6 moments) | − 0.754 (0.158) | 0.898 (0.021) | 0.362 (0.047) |
| (4) GARCH(1,1) − Kz(2) (6 moments) | − 0.739 (0.153) | 0.899 (0.020) | 0.352 (0.051) |
| (5) GARCH(1,1) − Kz(4) (8 moments) | − 0.769 (0.135) | 0.896 (0.018) | 0.363 (0.033) |
| (6) EGARCH(1,0) − Kz(2) − Kx(1) (9 moments) | − 0.814 (0.684) | 0.890 (0.092) | 0.340 (0.117) |
| (7) GMM (crash free) (14 moments) | − 0.835 (0.219) | 0.887 (0.029) | 0.342 (0.051) |
| (8) Infeasible GMM (true weight) (14 moments) | − 0.786 (0.175) | 0.893 (0.024) | 0.373 (0.050) |

True parameters $(\alpha, \beta, \sigma_u) = (-0.736, 0.90, 0.363)$, 500 Monte Carlo iterations.
For each score generator, we report the mean and the root mean square error in parentheses.
The GMM results are from Andersen and Sørensen (1996), Tables 7 and 3, resp.

parameter, $\beta$. The following subsections provide a brief summary and discussion of our findings along the various dimensions.

## 5.1. Comparison to GMM

We initially compare our small-sample EMM results to those for GMM reported by AS. In Table 1, for $T = 500$, EMM clearly dominates the GMM implementation of AS which, even in the best of cases, fails to converge in about 30% of the simulations. AS document that these failures invariably are associated with samples where no optimum for the criterion function exists within the (open) parameter space. In contrast, we do not encounter any convergence problems for the EMM-procedure, when using relatively parsimonious score generators. Even the GMM results (for $T = 500$) using the 'true' weighting matrix yield RMSEs that are dominated by the present EMM results: although the EMM-RMSE's for $\sigma_u$ are somewhat larger than for the infeasible GMM procedure, the preferred EMM procedures lower the RMSE for $\alpha$ and $\beta$ by about one-half. The sample size $T = 1000$ produces qualitatively similar findings. The corresponding distributions of the estimates for the important volatility persistence parameter, $\beta$, are also displayed in Fig. 1. Again, it must be kept

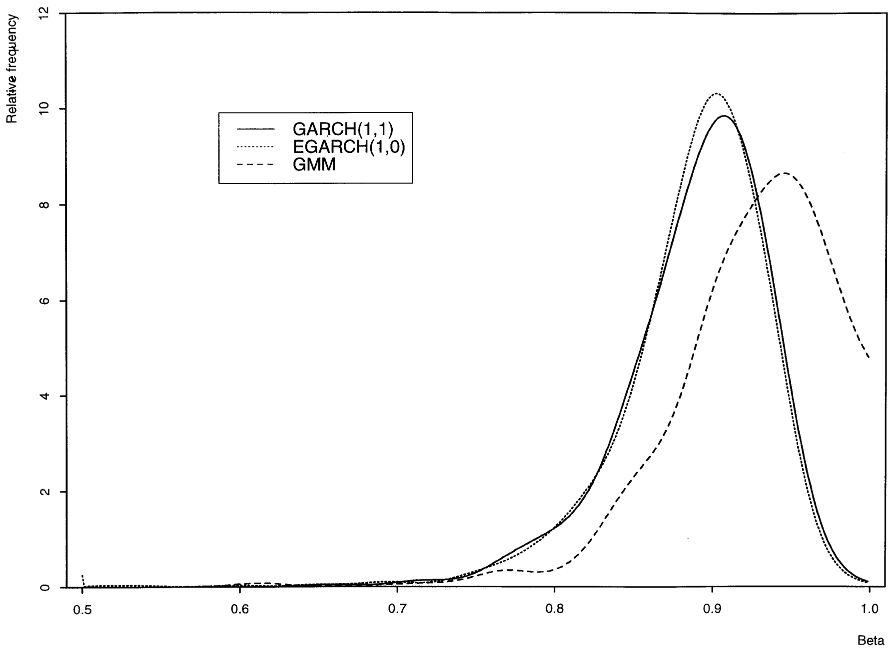


Fig. 1. Distribution of beta for GMM and EMM with GARCH and EGARCH score generators ($T = 1000$).

in mind that over 13% of the GMM estimates are omitted, as they were converging towards the infeasible upper bound on $\beta$ of unity, and then 'crashed'. These results support the notion that EMM improves efficiency via the superior selection of moment conditions, and not simply because of problems associated with the GMM weighting matrix.

For $T = 4000$, AS show that GMM becomes operational, i.e., the estimation procedure converges for all simulated series. From Table 3, it is evident that all method of moments procedures provide nearly unbiased estimates for $\beta$ in this setting. However, while the EMM estimates of the mean volatility parameter, $\alpha$, are slightly downward biased, EMM displays little bias for $\sigma_u$. Both parameters are significantly downward biased under GMM. Furthermore, the more satisfactory EMM-implementations typically display a reduction of about 30% in the RMSE relative to GMM.

The evidence is unambiguous. The EMM selection of moment conditions generates a very substantial improvement in estimation efficiency relative to the use of simple unconditional return moments, as implemented in AS.

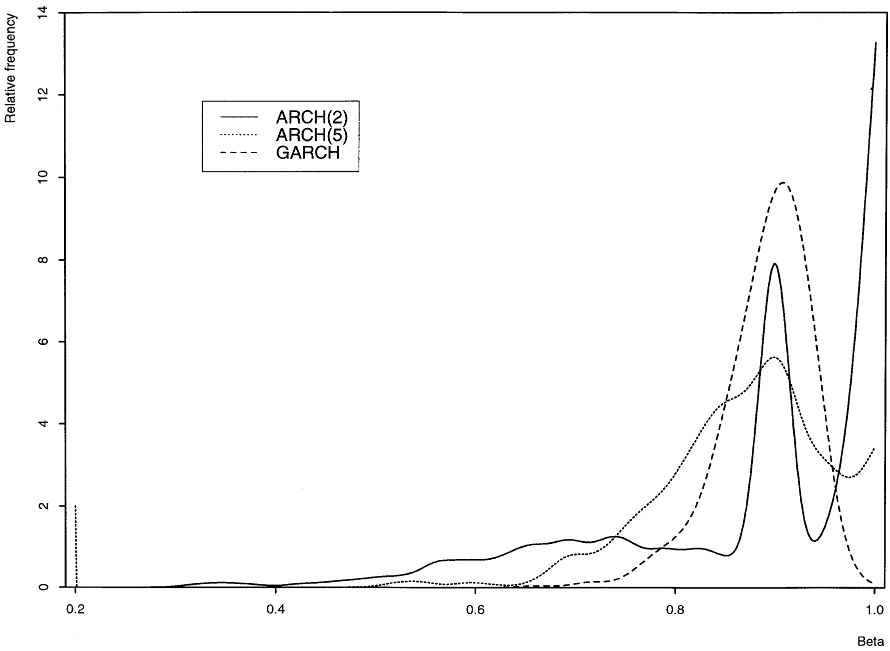


Fig. 2. Distribution of beta for different score generators ($T = 1000$).

### 5.2. Choice of score generator

The efficiency of EMM should approach that of maximum likelihood if the score generator provides a close approximation to the true conditional density, and thus effectively nests the structural model. This suggests that the choice of auxiliary model is an important practical concern. In prior applied work, a number of approaches have been used. At the extreme ends of the spectrum, Gallant et al. (1997) employ the purely non-parametric SNP generator to estimate a long series of 16,127 daily stock returns, while Andersen and Lund (1997a) rely on an elaborate EGARCH specification as the leading term in the SNP expansion. Substantial efficiency gains may be feasible in small samples using such a parametric leading term, relative to a (usually higher-dimensional) 'pure' SNP generator, as also argued in, e.g., Tauchen (1996).

$T=500$. For $T=500$, this intuition is strongly supported by our results. In fact, for this limited sample size, we are typically unable to achieve convergence for *any* model containing non-parametric SNP terms (i.e. $K_z > 0$). We stress that, by 'no convergence', we do not refer to the AS problem in the GMM context, where an optimum does not exist within the parameter space. Instead, we refer to a scenario where QML estimation of the score generator model is

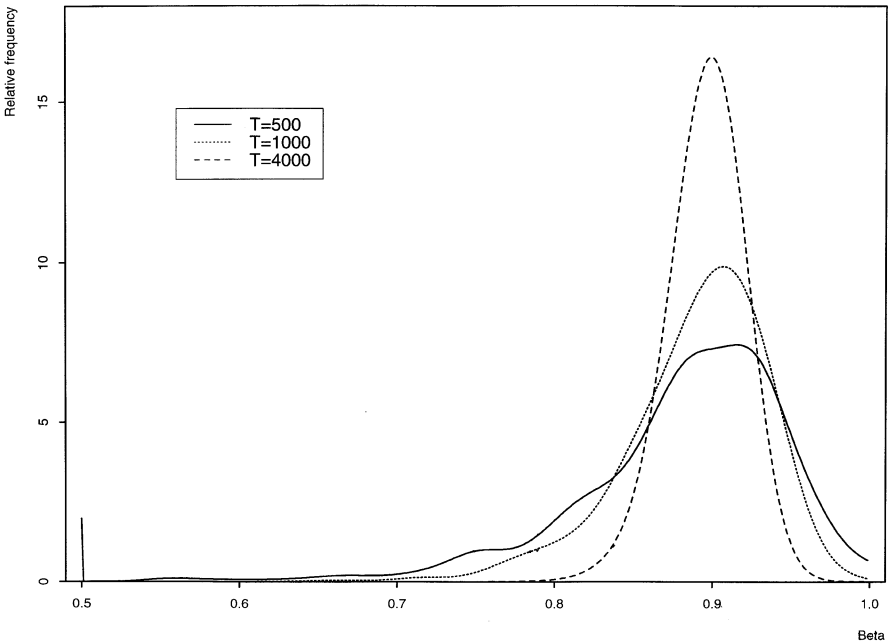


Fig. 3. Distribution of beta for GARCH(1,1) score generator.

numerically unstable, resulting in singularity crashes and/or prohibitively long convergence times.[5] For brevity, we refer to models that are numerically unstable in this sense as 'non-convergent' or 'numerically unstable' in the sequel.

We have experimented with a variety of parametric score generators for $T=500$, and generally find high-dimensional ARCH models to be numerically unstable. More tightly parameterized score generators are typically better behaved. Table 1 displays results for five alternative score generators of the GARCH and EGARCH variety. Since the EGARCH models and the lognormal stochastic volatility model share the identical continuous-record diffusion limit (see Nelson, 1990), one may a priori expect EGARCH score generators to be preferable. Specifically, we explore the parsimonious EGARCH(1,0) model with the asymmetry parameter, $\phi$, set to zero, as well as a less restrictive

[5] We do not rule out that a study employing more sophisticated optimization algorithms or modifications of the SNP model will be able to obtain convergence. Our goal is to identify EMM implementations that are well behaved using standard Newton-type optimization methods. Furthermore, our subsequent results strongly suggest that extensions relying on highly parameterized models in small samples are unlikely to provide improved inference, and we consequently terminated our experimentation along such lines.

Table 4
Simulated mean and root mean square error for $\beta = 0.98$

| Score generator | $\alpha$ | $\beta$ | $\sigma_u$ |
|---|---|---|---|
| (1) GARCH(1,1) ($T = 500$) (4 moments) | − 0.282 (0.329) | 0.944 (0.162) | 0.132 (0.164) |
| (2) GARCH(1,1) ($T = 1000$) (4 moments) | − 0.207 (0.132) | 0.972 (0.018) | 0.149 (0.108) |
| (3) GARCH(1,1)* ($T = 4000$) (4 moments) | − 0.161 (0.046) | 0.978 (0.0062) | 0.169 (0.020) |
| (4) GARCH(1,1) − Kz(2)** ($T = 4000$) (4 moments) | − 0.163 (0.048) | 0.978 (0.0064) | 0.163 (0.028) |
| (5) GARCH(1,1) − Kz(4)** ($T = 4000$) (6 moments) | − 0.162 (0.045) | 0.978 (0.0060) | 0.165 (0.016) |
| (6) GMM ($T = 2000$) (14 moments) (810 crashes ignored) | − 0.140 (0.112)* | 0.981 (0.015)* | 0.125 (0.068)* |
| (7) GMM ($T = 4000$) (14 moments) (399 crashes ignored) | − 0.121 (0.089)* | 0.984 (0.012)* | 0.126 (0.063)* |

True parameters ($\alpha$, $\beta$, $\sigma_u$) = ( − 0.147, 0.98, 0.166), 500 Monte Carlo iterations.
For each score generator, we report the mean and the root mean square error in parentheses.
*The GMM results are from Andersen and Sørensen (1996), Table 9. Due to ignored crashes, these RMSEs are not comparable to other rows.
The score-generators marked with ** are simulated using 100,000 antithetic draws in the Monte Carlo integration.

EGARCH(1,1) model. Many ARCH models, however, provide consistent volatility filters for the Nelson diffusion, suggesting that the choice of auxiliary model may not be clear cut, and thus motivating our exploration of alternative specifications.

It is evident from Table 1 that a tightly parameterized score generator is needed. The lowest RMSEs are obtained for the GARCH(1,1) model. EGARCH(1,0) performs somewhat worse due to a few extreme parameter estimates. A natural conjecture is that these outliers are caused by poor estimates of the quasi-information (weighting) matrix. We thus performed a set of simulations using the inverse of the diagonal of the information matrix as the weighting matrix (AS found that such a short-cut worked well), but the results deteriorated, and we did not experiment any further with diagonal weighting. In general, it is clear that, for this small sample size, including a large number of parameters in the score generator leads to a deterioration of the results. In particular, the asymmetric EGARCH generator with six parameters is much inferior to the four-parameter symmetric EGARCH generator ($a = 0$ and $\phi = 0$).

The general tradeoff, discussed in AS, between the amount of information included in estimation via incorporation of additional moments and the associated deterioration of the precision of the GMM objective function is also relevant here. For $T = 500$, the more parsimonious representations are strongly

Table 5
Comparison of QML, GMM, EMM, JPR, ML, and MCL

| | $T = 500$ | | | $T = 2000$ | | |
|---|---|---|---|---|---|---|
| | $\alpha$ | $\beta$ | $\sigma_u$ | $\alpha$ | $\beta$ | $\sigma_u$ |
| | − 0.736 | 0.90 | 0.363 | − 0.736 | 0.90 | 0.363 |
| QML[a] | − 1.4 | 0.81 | 0.45 | − 0.853 | 0.88 | 0.383 |
| | (1.60) | (0.22) | (0.27) | (0.46) | (0.06) | (0.11) |
| GMM[b] | − 0.62 | 0.92 | 0.24 | − 0.59 | 0.92 | 0.28 |
| | (0.59)* | (0.08)* | (0.17)* | (0.31) | (0.04) | (0.12) |
| EMM[c] | − 0.91 | 0.88 | 0.38 | − 0.793 | 0.893 | 0.359 |
| | (0.60) | (0.08) | (0.20) | (0.224) | (0.030) | (0.049) |
| JPR[d] | − 0.87 | 0.88 | 0.35 | − 0.762 | 0.896 | 0.359 |
| | (0.34) | (0.05) | (0.07) | (0.15) | (0.02) | (0.034) |
| ML[e] | − 0.87 | 0.88 | 0.37 | | NA | |
| | (0.43) | (0.05) | (0.08) | | NA | |
| MCL[f] | − 0.60 | 0.90 | 0.37 | | NA | |
| | (0.27) | (0.04) | (0.08) | | NA | |

RMSE in parentheses.
[a]Jacquier et al. (1994).
[b]Andersen and Sørensen (1996).
*For $T = 500$ GMM crashed 342 times in 1342 iterations. Mean and RMSE for converged values only. These RMSEs are not comparable to other rows.
[c]GARCH(1,1) for $T = 500$. GARCH(1,1)-Kz(4) for $T = 2000$.
[d]Jacquier et al. (1994).
[e]Fridman and Harris (1996).
[f]Sandmann and Koopman (1996).

favored. This does not imply that a minimal number of moments is preferable: indeed, the (exactly identified) model reported in the first row of Table 1 is inferior to the corresponding model allowing for a mean parameter. This is perhaps mildly surprising since the structural model has mean zero, but it is simply a reflection of the fact that the auxiliary model does not nest the true model, so that allowing for an extra degree of freedom in the score generator may be beneficial.

$T = 1000$. For a sample size of $T = 1000$, the class of auxiliary models that are estimable without convergence problems expands. The more parsimonious models, GARCH(1,1) and EGARCH(1,0), still perform well, but including a few tail parameters in an SNP expansion ($K_z > 0$) is now not only feasible, but actually beneficial: the GARCH(1,1)-$K_z$(2) provides the uniformly lowest RMSEs. However, any further expansion in the SNP terms leads to a deterioration in estimation performance. We did not obtain convergence when including seminonparametric heteroskedasticity terms ($K_x > 0$). The ARCH(5) score perform poorly, as indicated by the high RMSEs, which reflect pronounced

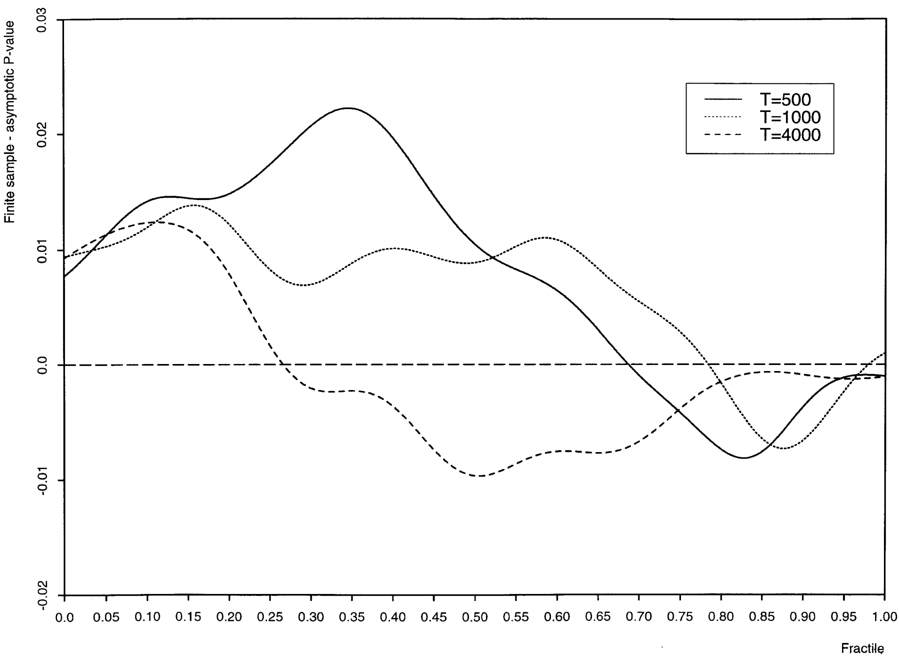


Fig. 4. $P$-value discrepancy plot for the chi-square specification test, GARCH(1,1).

left-skewness and kurtosis. This again illustrates the tradeoff between over-parameterization and finite-sample efficiency.[6] Of course, parsimony is only a virtue if the score generator picks up the main features of the data. This is clearly illustrated by the ARCH(1) and ARCH(2) models: they are parsimonious, but provide inferior approximations to the structural model, resulting in noisy inference as indicated by the RMSEs. The latter are truly poor for the ARCH(1), and still roughly triple those of the preferred models for ARCH(2). In addition, both are inferior to the ARCH(5) score. Fig. 2 graphically documents the disastrous performance of the ARCH-scores relative to the pure GARCH(1,1).

$T=2000$. For the sake of brevity, our results for $T=2000$ are not tabulated. We briefly summarize the findings. The parsimonious EGARCH and GARCH models continue to perform well, but the preferred model is GARCH(1,1)-$K_z(4)$. Including additional parameters in the EGARCH model induces a small deterioration in the RMSE, but the impact is minor. We did not obtain convergence with a pure SNP-generator, but we obtain parameter estimates for 100

[6] An experiment with an ARCH(10) model was aborted after 100 iterations. The tentative results were inferior to the ARCH(5) and convergence was slow.

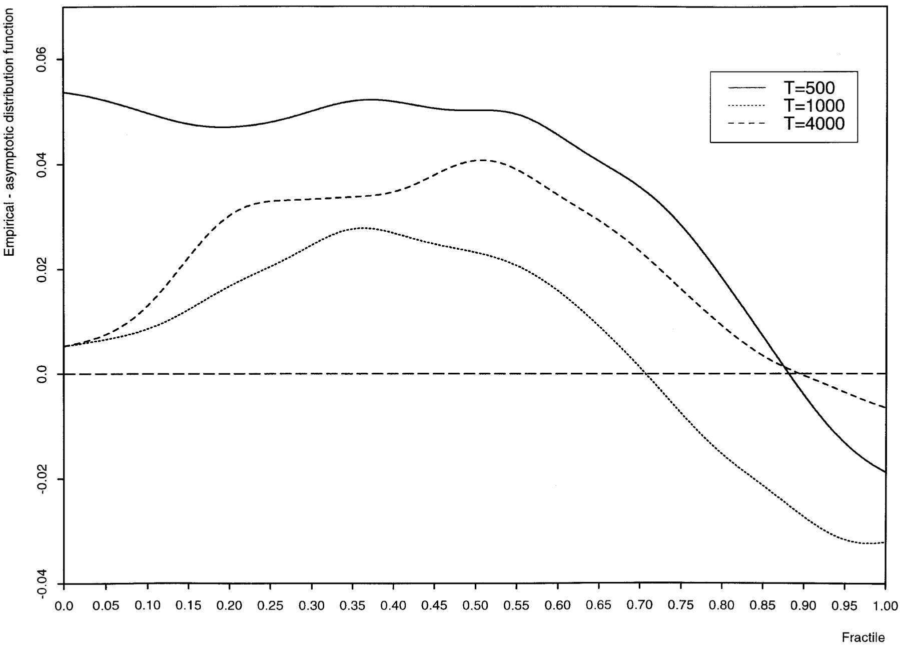


Fig. 5. $P$-value discrepancy plot for $t$-statistics, GARCH(1,1).

simulated series using an EGARCH(1,0)-$K_z$(2)-$K_x$(1) score generator.[7] However, convergence is typically slow and the results are marred by some extreme outliers, so it is clearly no serious contender. The findings confirm that poor convergence properties of the auxiliary model are indicative of poor inference in the overall EMM-procedure as well.

$T=4000$. The largest sample size we explore is $T=4000$. The results are provided in Table 3. The EMM estimator performs extremely well for this relatively large sample. Moreover, the results appear, to a first approximation, independent of the leading term in the score generator. Most of the candidate auxiliary models perform roughly the same, implying that EMM is quite robust. Although this finding may be anticipated given the theoretical results of Fenton and Gallant (1996), the fact that the dimension of the auxiliary model may be increased quite rapidly for realistic sample sizes, at least for the 'plain-vanilla' stochastic volatility model, is encouraging. For $T=4000$, the results – in particular for the volatility-of-volatility parameter $\sigma_u$ – are clearly improved with

[7] This design makes use of the logistic transformation of the lagged returns, suggested in Gallant and Tauchen (1995), as we encounter severe problems without this transformation.

the inclusion of non-parametric 'tail terms' ($K_z > 0$). For example, the GARCH(1,1)-$K_z$(4) score includes eight moments and has the lowest RMSE for all structural parameters. Of course, it is still possible to overparameterize the auxiliary model: we tried to estimate the GARCH(1,1)-$K_z$(8) score, but it was hard to obtain convergence. Moreover, it is now feasible to estimate a model containing nonparametric heteroskedasticity terms ($K_x > 0$), and results for the EGARCH(1,0)-$K_z$(2)-$K_x$(1) score are included in Table 3. The RMSEs for this model are, however, very large, reflecting the presence of large outliers, so the inclusion of nonparametric heteroskedasticity terms is still not beneficial.

In summary, we infer that a non-parsimonious parameterization of the auxiliary model induces large efficiency costs for smaller samples, while parsimony is less critical for larger (but still realistic) sample sizes. Moreover, it is critical to include a suitable parametric leading term in the auxiliary model for all sample sizes considered. Finally, inference improves rapidly as sample size grows. This is vividly illustrated in Fig. 3, which displays the distribution of the EMM-$\beta$ parameter estimates for various sample sizes obtained with the simple GARCH(1,1) score generator.

### 5.3. Higher volatility persistence

Many studies of stochastic autoregressive volatility models obtain estimates of $\beta$ close to unity for daily data. We therefore study the case $\beta = 0.98$, which is summarized in Table 4. AS find that 10,000 observations typically are required to avoid convergence problems for GMM for this parameter-constellation. In contrast, with a parsimonious parameterization of the auxiliary model, EMM does not display any convergence problems, even for small samples. There is a relative high incidence of outliers for $T = 500$, but the results are otherwise entirely unproblematic. The parameters are estimated very precisely for the larger samples, in particular $T = 4000$. The drop in RMSE, relative to the $\beta = 0.90$ case (for $T$ larger than 500), is consistent with AS, JPR, and Harvey and Shephard (1996) who also find that the improved signal-to-noise ratio of the volatility process implies a lower RMSE for larger samples. As before, we find that the inclusion of non-parametric tail parameters improves the precision over the purely parametric score generator for $T = 4000$. Specifically, the results for the GARCH(1,1)-$K_z$(4) model reported in row (5) dominate those of the corresponding GARCH(1,1) in row (3).[8]

[8] We have explored a number of additional designs in an earlier draft of this paper. Specifically, we have results for $\beta = 0.95$, for ARMA models for the conditional mean, and for endogenous selection of the score generator model based on alternative information criteria. Furthermore, skewness and kurtosis measures for all parameter estimates are available. These results may be obtained from the authors upon request.

### 5.4. Monte Carlo integration of the moment conditions

The evaluation of the score moment conditions is performed by Monte Carlo integration of the expected value of the score vector under the true model. As we require thousands of Monte Carlo samples to be drawn from the structural model, we have constrained ourselves to two antithetic series of length 20,000 when evaluating these integrals throughout our simulation study. In an empirical investigation one may prefer a higher number of draws in order to render the simulation error (near) negligible. We consequently conduct a small experiment keeping the original, simulated sample – and thus the estimated score generator model, GARCH(1,1) – fixed while repeating the second EMM-estimation step. All remaining uncertainty is then due to the numerical procedure applied to evaluate the integrals. For a typical estimate of the auxiliary model in the lower persistence case ($\beta = 0.90$), the associated standard error for the ($\alpha$, $\beta$, $\sigma_u$) parameter vector is of the order (0.070, 0.009, 0.022).[9] This standard error is small relative to the overall Monte Carlo variation for the smaller samples, but not negligible relative to the reported RMSE for $T=4000$.[10] For sample sizes of $T=2000$ and smaller, the variation induced by the Monte Carlo integration is not important: re-estimating the GARCH(1,1) model for $T=2000$ using two antithetic series of length 100,000 does not appear to improve on the reported RMSEs. For the high persistence case ($\beta = 0.98$) the standard error from Monte Carlo integration (using two times 20,000 draws) is about (0.033, 0.005, 0.017). For this high persistence and $T=4000$, the parameters are, however, estimated very precisely and the Monte Carlo integration error may therefore add noticeably to the Monte Carlo variation in the estimates. The results for this design are therefore calculated using two antithetic series of length 100,000.

### 5.5. Comparison to maximum likelihood

Table 5 compares RMSEs of our preferred EMM results to GMM, QML, JPR, ML, and MCL for $T=500$ and $T=2000$. For $T=500$, EMM is clearly

[9] Doubling the number of draws to 40,000, the standard errors induced by the numerical integration improve by a factor of $\sqrt{2}$. This suggests that the rate of convergence implied by the standard central limit theorem is relevant for these sample sizes.

[10] To assess the magnitude of the variation induced by the Monte Carlo integration error, assume that the standard error from the Monte Carlo integration is $X$ while the standard error arising from other sources is $Y$ and that the distributions of these two error components are independent (which is likely to be a reasonable approximation). The overall standard error is then $\sqrt{X^2 + Y^2}$. If we ignore the slight bias in the parameter estimates, the reported RMSEs equal $\sqrt{X^2 + Y^2}$, while an approximate RMSE in the absence of Monte Carlo integration error would be given by the square-root of [(reported RMSE)$^2$ − Monte Carlo integration variance]. The minimum RMSE for $\beta$ in Table 4 is 0.018. An estimate of the RMSE that will prevail with zero Monte Carlo integration error is thus $\sqrt{0.018^2 - 0.009^2} = 0.016$.

not as efficient as the likelihood-based methods, but EMM improves substantially on GMM which, for this sample size, is simply not a viable alternative, since the about a quarter of all attempts to estimate the model results in non-convergence. For $T=2000$, the efficiency of EMM falls between GMM and JPR, but it is closer to the JPR estimator. Furthermore, the impressive improvements associated with the longer, but still empirically relevant, sample of $T=4000$ (Table 3) verifies that EMM is a very reliable tool for inference in this setting, and suggests that the efficiency of EMM approaches that of the genuine likelihood-based procedures.[11]

### 5.6. Specification testing

Under correct specification, $T$ times the optimized GMM criterion is distributed $\chi^2(n_\eta - n_\rho)$. This statistic therefore provides a test for the goodness-of-fit of the structural model (known as the test of the overidentifying restrictions), see Hansen (1982). The $P$-values are asymptotically uniform over the fractiles of the $\chi^2$ distribution. We illustrate how well the finite-sample distribution of the test statistic conforms to the asymptotic distribution by plotting the difference between the empirical distribution for the $P$-values of the finite-sample test statistic and the corresponding asymptotic $P$-values over the entire unit interval in Fig. 4. This is the so-called $P$-value discrepancy plot for the goodness-of-fit test statistic, see Davidson and MacKinnon (1998). If the discrepancies are close to zero, the finite-sample distribution of the test statistic mirrors that of the asymptotic distribution. A simple calculation reveals that the point estimates for $T=1000$ and $T=4000$ indicated in Fig. 4 generally are well within a 95% confidence band around zero. Thus, there is no indication of bias in the specification test as long as the samples are of moderate size and a reasonable score generator has been used.

These results contrasts sharply to the GMM-evidence reported by AS. For GMM an excessive number of moments tends to imply underrejection (inflated $P$-values), whereas too few moments lead to overrejection. In addition, as $T$ increases, a systematic leftward shift takes place in the $P$-value distribution for typical sample sizes. In fact, more than a million observations are needed for the distribution of the specification test statistics to approximate the $\chi^2$ in the GMM setting.

The conclusion is that EMM-inference based on the test for over-identifying restrictions is reliable. This is likely a testament to the improved estimates of the weighting matrix that are obtained from the quasi-score moment vector under

[11] There is no evidence in the literature on the behavior of the likelihood-based procedures for samples of this magnitude.

EMM relative to the highly serially correlated (unconditional) moments used in GMM.

### 5.7. Hypothesis tests

To examine inference based on the individual parameter estimates, we explore the studentized distribution of the structural volatility persistence parameter, $\beta$. The studentized parameter is defined as the difference between the true and estimated parameter normalized by the (EMM-) estimate of the standard error. Asymptotically, the studentized parameters are distributed as standard normals, so the mass located in the tail fractiles approximates the size of one-sided tests for equality of the estimated parameters and their true value.

Fig. 5 provides the $P$-value discrepancy plots[12] for the relevant $t$-statistics for inference on $\beta$ across different sample sizes, when estimation is based on the GARCH(1,1) auxiliary model. For the smallest sample, $T=500$, there is a significant positive deviation for the lower fractiles. This reflects a downward bias in the $t$-statistic for $\beta$. The tendency for $\beta$ to be biased downward was, of course, already documented in Fig. 3. Interestingly, the relatively large left tail of the corresponding distribution in Fig. 3 for $T=1000$ does not translate into a similar significant deviation for the $t$-statistic at the lower fractiles. This implies that the bias has been accommodated by sufficiently large estimates for the associated standard errors. If anything, the $t$-statistic for $T=1000$ deviates slightly from normality at the higher fractiles, indicating an upward bias in the EMM estimates of $\beta$, that is not reflected in the EMM standard errors. For $T=4000$, there is only a mild indication of size-distortions among the fractiles in the midst of the unit interval, which are less critical for inference.

In summary, standard hypothesis testing based on individual parameter estimates is much more reliable for all sample sizes than might be expected from the dispersion of the point estimates in Fig. 3. Again, the results are vastly superior to the corresponding GMM results reported by AS.

## 6. Conclusion

A number of broad conclusions emerge from our study. Perhaps most significantly, the EMM procedure performs quite well in comparison with earlier estimation procedures. It provides a very substantial improvement in efficiency relative to simple GMM, as the RMSEs are reduced uniformly across the simulation designs. Further, contrary to GMM, the procedure does not

[12] Fig. 5 shows the difference between the empirical distribution function and the Gaussian distribution function evaluated at the fractiles for the Gaussian distribution function.

encounter any convergence problems for smaller samples when parsimonious score generators are used. Although EMM is generally not quite as efficient as JPR's MCMC, the flexibility of method of moments estimators offers an advantage over direct likelihood-based inference procedures that often are constructed specifically for inference in the stochastic volatility model. Moreover, it is evident that the efficiency of EMM approaches that of the Bayesian likelihood estimator for large, yet empirically relevant, sample sizes.

Several observations regarding the implementation of EMM are worth reiterating. First, it is advisable to include a suitable leading parametric term in the score generator, such that the bulk of the conditional heteroskedasticity in the data series is captured in a parsimonious fashion. Second, although theoretical considerations may suggest that an EGARCH model is particularly well suited as a leading term under the current data generation scheme a GARCH leading term generally performs equally well. Third, it is important to conserve on the number of moment conditions in small samples of 1000 observations or less. In particular, including non-parametric terms in the form of an SNP-expansion is likely to generate numerical instability in smaller samples. However, the inclusion of Hermite polynomial terms improves efficiency considerably in the larger samples. Fourth, in larger samples the efficiency of the procedure is quite robust to the choice of score-generator. Fifth, the test for over-identifying restrictions, which often performs poorly in the standard GMM context, is remarkably reliable for inference in the EMM setting. Finally, inference regarding the parameter values based on standard $t$-statistics is also quite well-behaved under EMM.

While our findings, of course, only apply in a strict sense to the estimation designs investigated, we expect the qualitative conclusions to carry over to a wide range of economic models with serially correlated latent factors. As such, the encouraging EMM results provide a strong incentive to implement the procedure for a variety of dynamic latent models beyond the stochastic volatility setting. The true benefit of EMM is, of course, realized in settings where, unlike the case of the 'plain-vanilla' stochastic volatility model, the procedure is applicable, while the more direct likelihood-based approaches are infeasible.

## Acknowledgements

We are grateful to Hyung-Kwon Chung for assistance and David Tom for programming the EGARCH model in C + + . The second author gratefully acknowledges financial assistance from the Ehrlich Foundation. We thank the editor, Ron Gallant, an associate editor, three anonymous referees, as well as Moshe Buchinsky, Neil Shephard, Doug Steigerwald, and seminar participants at Brown University, Yale University, University of Rochester, the NBER 1997 Summer Institute and the Econometric Society Winter Meetings, Chicago, January 1998, for comments.

## References


Andersen, T.G., 1994. Stochastic autoregressive volatility: A framework for volatility modeling. Mathematical Finance 4, 75–102.

Andersen, T.G., Lund, J., 1997a. Estimating continuous-time stochastic volatility models of the short-term interest rate. Journal of Econometrics 77, 343–377.

Andersen, T.G., Lund, J., 1997b. Stochastic volatility and mean drift in the short rate diffusion: sources of steepness, level and curvature in the yield curve. Working paper no. 214, Department of Finance, Northwestern University.

Andersen, T.G., Sørensen, B.E., 1996. GMM estimation of a stochastic volatility model: A Monte Carlo study. Journal of Business and Economic Statistics 14, 328–352.

Andersen, T.G., Sørensen, B.E., 1997. GMM and QML asymptotic standard deviations in stochastic volatility models: A comment on Ruiz (1994). Journal of Econometrics 76, 397–403.

Bansal, R., Gallant, A.R., Hussey, R., Tauchen, G.E., 1993. Computational aspects of nonparametric simulation estimation. In: Belsley, D.A. (Ed.), Computational Techniques for Econometrics and Economic Analysis. Kluwer Academic Publishers, Boston, MA, pp. 3–22.

Bansal, R., Gallant, A.R., Hussey, R., Tauchen, G.E., 1995. Nonparametric estimation of structural models for high-frequency currency market data. Journal of Econometrics 66, 251–287.

Black, F., 1976. Studies in stock price volatility changes. Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section. American Statistical Association, pp. 177–181.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Danielsson, J., 1994. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. Journal of Econometrics 61, 375–400.

Danielsson, J., Richard, J.F., 1993. Accelerated gaussian importance sampler with application to dynamic latent variable models. Journal of Applied Econometrics 8 (Suppl.), S153–S173.

Davidson, R., MacKinnon, J.G., 1998. Graphical methods for investigating the size and power of hypothesis tests. The Manchester School, forthcoming.

Duffie, D., Singleton, K.J., 1989. Simulated moments estimation of Markov models of asset prices. Working paper, Graduate School of Business, Stanford University.

Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50, 987–1008.

Engle, R.F., Lilien, D.M., Robins, R.P., 1987. Estimating time-varying risk premia in the term structure: the ARCH-M model. Econometrica 55, 391–407.

Fenton, V.M., Gallant, A.R., 1996. Convergence rates of SNP density estimators. Econometrica 64, 719–727.

Fridman, M., Harris, L., 1998. A maximum likelihood approach for Mon-Gaussian stochastic volatility models. Journal of Business and Economic Statistics 16, 284–291.

Gallant, A.R., Hsieh, D.A., Tauchen, G.E., 1997. Estimation of stochastic volatility models with diagnostics. Journal of Econometrics 81, 159–192.

Gallant, A.R., Long, J.R., 1997. Estimating stochastic differential equations efficiently by minimum chi-square. Biometrika 84, 125–142.

Gallant, A.R., Nychka, D.W., 1987. Semi-nonparametric maximum likelihood estimation. Econometrica 55, 363–390.

Gallant, A.R., Tauchen, G.E., 1995. SNP: A program for nonparametric time series analysis. Version 8.4. Mimeo, Duke University.

Gallant, A.R., Tauchen, G.E., 1996. Which moments to match? Econometric Theory 12, 657–681.

Geweke, J., 1994. Bayesian comparison of econometric models. Working Paper, Federal Reserve Bank of Minneapolis, Minnesota.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8 (Suppl.), S85–S118.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies 61, 247–264.

Harvey, A.C., Shephard, N., 1996. Estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business and Economic Statistics 14, 429–434.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 45, 361–393.

Melino, A., Turnbull, S.M., 1990. Pricing foreign currency options with stochastic volatility. Journal of Econometrics 45, 239–265.

Nelson, D.B., 1990. ARCH models as diffusion approximations. Journal of Econometrics 45, 7–38.

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

Sandmann, G., Koopman, S.J., 1996. Maximum likelihood estimation of stochastic volatility models. Working paper, London School of Economics.

Tauchen, G.E., 1996. New minimum chi-square methods in empirical finance. In: Kreps, D., Wallis, K. (Eds.), Advances in Economic Theory and Econometrics, Seventh World Congress, Cambridge University Press, New York.

Taylor, S., 1994. Modeling stochastic volatility: A review and comparative study. Mathematical Finance 4(2), 183–204.

Taylor, S., 1986. Modelling Financial Time Series. Wiley, Chichester.

White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press, New York.