On a simple two-stage closed-form estimator for a stochastic volatility in a general linear regression *

Jean-Marie Dufour † Pascale Valéry ‡
Université de Montréal HEC-Montréal

First version: September 2001
Revised: May 2002, September 2004
This version: October 2004
Compiled: June 24, 2005, 2:08pm

---

* The authors thank Silvia Gonçalves, Christian Gouriéroux, Nour Meddahi, an anonymous referee, and the Editors Thomas Fomby and Dek Terrell for several useful comments. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 (514) 343 2400; FAX: 1 (514) 343 5831; e-mail: jean.marie.dufour@umontreal.ca . Web page: http://www.fas.umontreal.ca/SCECO/Dufour

‡ Service de l'enseignement de la finance, École des Hautes Études Commerciales de Montréal (HEC-Montréal), 3000 chemin de la Côte-Sainte-Catherine Montréal, Québec), Canada H3T 2A7. TEL: 1 (514) 340-7004, FAX: (514) 340-5632. email: pascale.valery@hec.ca

ABSTRACT

In this paper, we consider the estimation of volatility parameters in the context of a linear regression where the disturbances follow a stochastic volatility model of order one with Gaussian log-volatility. The linear regression represents the conditional mean of the process and may have a fairly general form, including for example finite-order autoregressions. We provide a computationally simple two-step estimator available in closed form which can be described as follows: (1) the conditional mean model is first estimated by a simple consistent procedure that does take into account the stochastic volatility structure; (2) using residuals from this preliminary regression, the volatility parameters are then evaluated by a method-of-moment estimator based on three moments for which a simple closed-form expression is derived. Under general regularity conditions, we show that this two-step estimator is asymptotically normal with covariance matrix that does not depend on the distribution of the conditional mean estimator. We then study its statistical properties by simulation and compare it with alternative GMM estimators based on more moments. Finally, we provide an application to the Standard and Poor's Composite Price Index (1928-87).

Key words: econometrics; finance; time series; autoregressive model; stochastic volatility; method of moments; asymptotic distribution.

Journal of Economic Literature classification: C13, C22, G1.

RÉSUMÉ

Dans cet article, nous étudions l'estimation des paramètres de volatilité dans le contexte d'une régression linéaire où les perturbations suivent un modèle de volatilité stochastique d'ordre un avec log-volatilité gaussienne. La régression linéaire représente la moyenne conditionnelle du processus et peut prendre une forme générale, incluant par exemple des autorégressions d'ordre fini. Dans ce contexte, nous fournissons un estimateur simple en deux étapes peu coûteux en temps de calcul, disponible sous une forme explicite: (1) le modèle de moyenne conditionnelle est d'abord estimé par une procédure convergente simple, qui tient compte de la structure de volatilité stochastique; (2) en utilisant les résidus de cette régression préliminaire, les paramètres du modèle de volatilité stochastique sont ensuite évalués par un estimateur des moments fondé sur trois moments (2S-3M) pour lequel une forme simple est fournie. Sous des conditions de régularité générales, nous montrons que cet estimateur en deux étapes est asymptotiquement normalement distribué, avec une matrice de covariance qui ne dépend pas de la distribution de l'estimateur de moyenne conditionnelle. Nous étudions ensuite ses propriétés statistiques au moyen de simulations et comparons l'estimateur proposé avec des estimateurs de moments généralisés fondés sur un plus grand nombre de moments. Finalement, la procédure d'estimation est appliquée à des données quotidiennes sur l'indice composite du Standard and Poor's (1928-87).

Mots-clés: économétrie; finance; série chronologique; modèle autorégressif; volatilité stochastique; méthode des moments; distribution asymptotique.

Classification du Journal of Economic Literature: C13, C22, G1.

Contents

List of Tables

List of Assumptions, Propositions and Theorems

1. Introduction

Modelling conditional heteroskedasticity is one of the central problems of financial econometrics. The two main families of models for that purpose consist of GARCH-type processes, originally introduced by Engle (1982), and stochastic volatility (SV) models proposed by Taylor (1986). Although the latter may be more attractive – because they are directly connected to diffusion processes used in theoretical finance – GARCH models are much more popular because they are relatively easy to estimate; for reviews, see Gouriéroux (1997) and Palm (1996). In particular, evaluating the likelihood function of GARCH models is simple compared to stochastic volatility models for which it is very difficult to get a likelihood in closed form; see Shephard (1996), Mahieu and Schotman (1998) and the review of Ghysels, Harvey and Renault (1996). Due to the high dimensionality of the integral defining the likelihood function, this is a general feature of almost all nonlinear latent variable models. As a result, maximum likelihood methods are prohibitively expensive from a computational viewpoint, and alternative methods appear to be required for applying such models.

Since the first discrete-time stochastic volatility models were proposed by Taylor (1986) as an alternative to ARCH models, much progress has been made regarding the estimation of nonlinear latent variable models in general and stochastic volatility models in particular. The methods suggested include quasi maximum likelihood estimation [see Nelson (1988), Harvey, Ruiz and Shephard (1994), Ruiz (1994)], generalized method-of-moments (GMM) procedures [Melino and Turnbull (1990), Andersen and Sørensen (1996)], sampling simulation-based techniques – such as simulated maximum likelihood [Danielsson and Richard (1993), Danielsson (1994)], indirect inference and the efficient method of moments [Gallant, Hsieh and Tauchen (1997), Andersen, Chung and Sørensen (1999)] – and Bayesian approaches [Jacquier, Polson and Rossi (1994), Kim, Shephard and Chib (1998), Wong (2002*a*, 2002*b*)]. Note also that the most widely studied specification in this literature consists of a stochastic volatility model of order one with Gaussian log-volatility and zero (or constant) conditional mean. The most notable exception can be found in Gallant et al. (1997) who allowed for an autoregressive conditional mean and considered a general autoregressive process on the log-volatility. It is remarkable that all these methods are highly nonlinear and computer intensive. Implementing them can be quite complicated and get more so as the number of parameters increases (*e.g.*, with the orders of the autoregressive conditional mean and log-volatility).

In this paper, we consider the estimation of stochastic volatility parameters in the context of a linear regression where the disturbances follow a stochastic volatility model of order one with Gaussian log-volatility. The linear regression represents the conditional mean of the process and may have a fairly general form, which includes for example finite-order autoregressions. Our objective is to develop a computationally inexpensive estimator that can be easily exploited within a simulation-based inference procedures, such as Monte Carlo and bootstrap tests.[1] So we study here a simple two-step estimation procedure which can be described as follows: (1) the conditional mean model is first estimated by a simple consistent procedure that does take into account the stochastic volatility structure; for example, the parameters of the conditional mean can be estimated by ordinary least squares (although other estimation procedures can be used); (2) using residuals from

[1]This feature is exploited in a companion paper [Dufour and Valéry (2005)] where various simulation-based test procedures are developed and implemented.

this preliminary regression, the parameters of the stochastic volatility model are then evaluated by a method-of-moment estimator based on three moments (2S-3M) for which a simple closed-form expression can be derived. Under general regularity conditions, we show the two-stage estimator is asymptotically normally distributed. Following recent results on the estimation of autoregressive models with stochastic volatility [see, for example, Gonçalves and Kilian (2004, Theorem 3.1)], this entails that the result holds for such models.

An interesting and potentially useful feature of the asymptotic distribution stems from the fact its covariance matrix does not depend on the distribution of the conditional mean estimator, *i.e.*, the estimation uncertainty on the parameters of the conditional mean does not affect the distribution of the volatility parameter estimates (asymptotically). The properties of the 2S-3M estimator are also studied in a small Monte Carlo experiment and compared with GMM estimators proposed in this context. We find that the 2S-3M estimator has quite reasonable accuracy with respect to the GMM estimators: indeed, in several cases, the 2S-3M estimator has the lowest root mean-square error. With respect to computational efficiency, the 2S-3M estimator always requires less than a second while GMM estimators may take several hours before convergence obtains (if it does). Finally, the proposed estimator is illustrated by applying it to the estimation of a stochastic volatility model on the Standard and Poor's Composite Price Index (1928-87, daily data).

The paper is organized as follows. Section 3 sets the framework and the main assumptions used. The closed-form estimator studied is described in section 3. The asymptotic distribution of the estimator is established in section 4. In section 5, we report the results of a small simulation study on the performance of the estimator. Section 6 presents the application to the Standard and Poor's Composite Price Index return series. We conclude in section 7. All proofs are gathered in the Appendix.

2. Framework

We consider here a regression model for a variable y_t with disturbances that follow a stochastic volatility process, which is described below following a notation similar to the one used by Gallant et al. (1997). $\mathbb{N}_0$ refers to the nonnegative integers.

Assumption 2.1 LINEAR REGRESSION WITH STOCHASTIC VOLATILITY. *The process* $\{y_t : t \in \mathbb{N}_0\}$ *follows a stochastic volatility model of the type:*

$$y_t = x_t'\beta + u_t \,, \tag{2.1}$$

$$u_t = \exp(w_t/2) r_y z_t \,, \quad w_t = \sum_{j=1}^{L_w} a_j w_{t-j} + r_w v_t \,, \tag{2.2}$$

where x_t *is a* $k \times 1$ *random vector independent of the variables* $\{x_{\tau-1},\, z_\tau,\, v_\tau,\, w_\tau : \tau \leq t\}$, *and* β, r_y, $\{a_j\}_{j=1}^{L_w}$, r_w *are fixed parameters.*

Typically y_t denotes the first difference over a short time interval, a day for instance, of the log-price of a financial asset traded on security markets. The regression function $x_t'\beta$ represents the conditional mean of y_t (given the past) while the stochastic volatility process determines a varying conditional variance. A common specification here consists in assuming that $x_t'\beta$ has an autoregressive form as in the following restricted version of the model described by Assumption 2.1.

Assumption 2.2 AUTOREGRESSIVE MODEL WITH STOCHASTIC VOLATILITY. *The process $\{y_t : t \in \mathbb{N}_0\}$ follows a stochastic volatility model of the type:*

$$y_t - \mu_y = \sum_{j=1}^{L_y} c_j(y_{t-j} - \mu_y) + u_t \ , \tag{2.3}$$

$$u_t = \exp(w_t/2) r_y z_t \ , \quad w_t = \sum_{j=1}^{L_w} a_j w_{t-j} + r_w v_t \ , \tag{2.4}$$

where β, $\{c_j\}_{j=1}^{L_y}$, r_y, $\{a_j\}_{j=1}^{L_w}$ and r_w are fixed parameters.

We shall refer to the latter model as an AR-SV$(L_y,\ L_w)$ model. The lag lengths of the autoregressive specifications used in the literature are typically short, e.g.: $L_y = 0$ and $L_w = 1$ [Andersen and Sørensen (1996), Jacquier et al. (1994), Andersen et al. (1999)], $0 \leq L_y \leq 2$ and $0 \leq L_w \leq 2$ [Gallant et al. (1997)]. In particular, we will devote special attention to the AR-SV$(1,\ 1)$ model:

$$y_t - \mu_y = c(y_{t-1} - \mu_y) + \exp(w_t/2) r_y z_t \ , \quad |c| < 1 \ , \tag{2.5}$$

$$w_t = a w_{t-1} + r_w v_t \ , \quad |a| < 1 \ , \tag{2.6}$$

so that

$$cov(w_t, w_{t+\tau}) = a^\tau \gamma \tag{2.7}$$

where $\gamma = r_w^2/(1 - a^2)$. The basic assumptions described above will be completed by a Gaussian distributional assumption and stationarity condition.

Assumption 2.3 GAUSSIAN NOISE. *The vectors $(z_t,\ v_t)'$, $t \in \mathbb{N}_0$, are i.i.d. according to a* $\mathrm{N}[0,\ I_2]$ *distribution.*

Assumption 2.4 STATIONARITY. *The process $s_t = (y_t,\ w_t)'$ is strictly stationary.*

The process defined above is Markovian of order $L_s = max(L_y,\ L_w)$. Under these assumptions, the AR-SV$(L_y,\ L_w)$ is a parametric model with parameter vector

$$\delta = (\mu_y,\ c_1,\ \ldots\ ,\ c_{L_y},\ r_y,\ a_1,\ \ldots\ ,\ a_{L_w},\ r_w)'. \tag{2.8}$$

Due to the fact that the model involves a latent variable (w_t), the joint density of the vector of observations $y(T) = (y_1,\ \ldots\ ,\ y_T)$ is not available in closed form because the latter would involve evaluating an integral with dimension equal to the whole path of the latent volatilities.

3. Closed-form method-of-moments estimator

In order to estimate the parameters of the volatility model described in the previous section, we shall consider the moments of the residual process in (2.1), which can be estimated relatively easily from regression residuals. Specifically, we will focus on stochastic volatility of order one ($L_w = 1$). Set

$$\theta = (a,\, r_y,\, r_w)' \,, \tag{3.1}$$

$$v_t(\theta) \equiv \exp(\frac{aw_{t-1} + r_w v_t}{2}) r_y z_t \,,\ \forall t \,. \tag{3.2}$$

Model (2.1) - (2.2) may then be conveniently rewritten as the following identity:

$$y_t - x_t'\beta \, = v_t(\theta) \,,\ \forall t \,. \tag{3.3}$$

The estimator we will study is based on the moments of the process $u_t \equiv v_t(\theta)$. The required moments are given in the following lemma.[2]

Lemma 3.1 MOMENTS AND CROSS-MOMENTS OF THE VOLATILITY PROCESS. *Under the assumptions 2.1, 2.3 and 2.4 with $L_w = 1$, the moments and cross-moments of $u_t = \exp(w_t/2) r_y z_t$ are given by the following formulas: for k, l and $m \in \mathbb{N}_0$,*

$$\begin{aligned} \mu_k(\theta) \equiv \mathsf{E}(u_t^k) &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp\left[\frac{k^2}{8} r_w^2/(1-a^2)\right], \quad \textit{if } k \textit{ is even,} \\ &= 0\,, \quad \textit{if } k \textit{ is odd,} \end{aligned} \tag{3.4}$$

$$\begin{aligned} \mu_{k,l}(m|\theta) &\equiv \mathsf{E}(u_t^k u_{t+m}^l) \\ &= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp\left[\frac{r_w^2}{8(1-a^2)}(k^2 + l^2 + 2kla^m)\right] \end{aligned} \tag{3.5}$$

if k and l are even, and $\mu_{k,l}(m|\theta) = 0$ if k or l is odd.

On considering $k = 2$, $k = 4$, or $k = l = 2$ and $m = 1$, we get:

$$\mu_2(\theta) = \mathsf{E}(u_t^2) = r_y^2 \exp[r_w^2/2(1-a^2)] \,, \tag{3.6}$$

$$\mu_4(\theta) = \mathsf{E}(u_t^4) = 3r_y^4 \exp[2r_w^2/(1-a^2)] \,, \tag{3.7}$$

$$\mu_{2,2}(1|\theta) = \mathsf{E}[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1-a)] \,. \tag{3.8}$$

An important observation here comes from the fact that the above equations can be explicitly solved for a, r_y and r_w. The solution is given in the following lemma.

[2] Expressions for the autocorrelations and autocovariances of u_t^2 were derived by Taylor (1986, Section 3.5) and Jacquier et al. (1994). The latter authors also provide the higher-order moments $E[|u_t^m|]$, while general formulas for the higher-order cross-moments of a stochastic volatility process are reported (without proof) by Ghysels et al. (1996). For completeness, we give a relatively simple proof in the Appendix.

Lemma 3.2 MOMENT EQUATIONS SOLUTION. *Under the assumptions of Proposition* **3.1**, *we have:*

$$a = \left[\frac{\log[\mu_{2,2}(1|\theta)] + \log\left[\mu_4(\theta)/\big(3\mu_2(\theta)^4\big)\right]}{\log\left[\mu_4(\theta)/\big(3\mu_2(\theta)^2\big)\right]} \right] - 1\,, \tag{3.9}$$

$$r_y = \frac{3^{1/4}\mu_2(\theta)}{\mu_4(\theta)^{1/4}}\,, \tag{3.10}$$

$$r_w = \Big[(1-a^2)\log\left[\mu_4(\theta)/\big(3\mu_2(\theta)^2\big)\right]\Big]^{1/2}\,. \tag{3.11}$$

From lemmas **3.1** and **3.3**, it is easy to derive higher-order autocovariance functions. In particular, for later reference, we will find useful to spell out the second and fourth-order autocovariance functions.

Lemma 3.3 HIGHER-ORDER AUTOCOVARIANCE FUNCTIONS. *Under the assumptions of Proposition* **3.1**, *let* $X_t = (X_{1t},\, X_{2t},\, X_{3t})'$ *with*

$$X_{1t} = u_t^2 - \mu_2(\theta),\ X_{2t} = u_t^4 - \mu_4(\theta),\ X_{3t} = u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)\,. \tag{3.12}$$

Then the covariances $\gamma_i(\tau) = \mathsf{Cov}(X_{i,t},\, X_{i,t+\tau})$, $i = 1, 2, 3$, *are given by:*

$$\gamma_1(\tau) = \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1], \tag{3.13}$$

$$\gamma_2(\tau) = \mu_4^2(\theta)[\exp(4\gamma a^\tau) - 1]\,,\ \forall\tau \ge 1, \tag{3.14}$$

$$\gamma_3(\tau) = \mu_{2,2}^2(1|\theta)[\exp(\gamma(1+a)^2 a^{\tau-1}) - 1]\,,\ \forall\tau \ge 2, \tag{3.15}$$

where $\gamma = r_w^2/(1-a^2)\,.$

Suppose now we have a preliminary estimator $\hat{\beta}$ of β. For example, for the autoregressive model (2.3) - (2.4), estimation of the equation (2.3) yields consistent asymptotically normal estimators of β; see Gonçalves and Kilian (2004, Theorem 3.1) and Kuersteiner (2001). Of course, other estimators of the regression coefficients may be considered. Given the residuals

$$\hat{u}_t = y_t - x_t'\hat{\beta}\,,\ t = 0, 1, \ldots, T\,, \tag{3.16}$$

it is then natural to estimate $\mu_2(\theta)$, $\mu_4(\theta)$ and $\mu_{2,2}(1|\theta)$ by the corresponding empirical moments:

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\,, \quad \hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4\,, \quad \hat{\mu}_2(1) = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2\,.$$

This yields the following estimators of the stochastic volatility coefficients:

$$\hat{a} = \left[\frac{\log[\hat{\mu}_2(1)] + \log\left[\hat{\mu}_4/\big(3\hat{\mu}_2^4\big)\right]}{\log\left[\hat{\mu}_4/\big(3\hat{\mu}_2^2\big)\right]} \right] - 1\,, \tag{3.17}$$

$$\hat{r}_y = \frac{3^{1/4}\hat{\mu}_2}{\hat{\mu}_4^{1/4}} = \left(\frac{3\hat{\mu}_2^4}{\hat{\mu}_4}\right)^{1/4}, \tag{3.18}$$

$$\hat{r}_w = \Big[(1-\hat{a}^2)\log\big[\hat{\mu}_4/\big(3\hat{\mu}_2^2\big)\big]\Big]^{1/2}. \tag{3.19}$$

Clearly, it is straightforward to compute the latter estimates as soon as the estimator $\hat{\beta}$ used to compute the residuals $\hat{u}_t = y_t - x_t'\hat{\beta}$ is easy to obtain (*e.g.*, $\hat{\beta}$ could be a least squares estimator).[3]

4. Asymptotic distribution

We will now study the asymptotic distribution of the moment estimator defined in (3.17) - (3.19). For that purpose, it will be convenient to view the latter as a special case of the general class of estimators obtained by minimizing a quadratic form of the type:

$$M_T(\theta) = [\bar{g}_T(\hat{U}_T) - \mu(\theta)]'\hat{\Omega}_T[\bar{g}_T(\hat{U}_T) - \mu(\theta)] \tag{4.1}$$

where $\mu(\theta)$ is a vector of moments, $\bar{g}_T(\hat{U}_T)$ is the corresponding vector of empirical moments based on the residual vector $\hat{U}_T = (\hat{u}_1, \ldots, \hat{u}_T)'$, and $\hat{\Omega}_T$ is a positive-definite (possibly random) matrix. Of course, this estimator belongs to the general family of moment estimators, for which a number of general asymptotic results do exist; see Hansen (1982), Gouriéroux and Monfort (1995*b*, Volume 1, Chapter 9) and Newey and McFadden (1994). However, we need to account here for two specific features, namely: (1) the disturbances in (2.1) follow a stochastic volatility model, and the satisfaction of the relevant regularity conditions must be checked; (2) the two-stage nature of the procedure where the estimator of the parameter β of the conditional mean equation is obtained separately and may not be based on the same objective function as the one used to estimate θ. In particular, it is important to know whether the estimator of the conditional mean parameter β has an effect on the asymptotic distribution of the estimator of θ. It is worth noting at this stage that Andersen and Sørensen (1996) did refer to the asymptotic distribution of the usual GMM estimator as derived in Hansen (1982), but without checking the suitable regularity conditions for the SV model, what we will do here.

To spell out the properties of the estimator $\hat{\theta}_T(\hat{\Omega}_T)$ obtained by minimizing $M_T(\theta)$, we will consider first the following *generic assumptions*, where θ_0 denotes the "true" value of the parameter vector θ.

Assumption 4.1 ASYMPTOTIC NORMALITY OF EMPIRICAL MOMENTS.

$$\sqrt{T}\big[\bar{g}_T(U_T) - \mu(\theta_0)\big] \xrightarrow{D} \mathrm{N}[0,\ \Omega_*] \tag{4.2}$$

[3] Andersen and Sørensen (1996) did also consider a moment estimator based on a different set of 3 moments, namely $\mathsf{E}(|u_t|)$, $\mathsf{E}(u_t^2)$ and $\mathsf{E}(|u_t u_{t-1}|)$. A closed-form solution for this estimator was not provided.

where $U_T \equiv (u_1, \ldots, u_T)'$ *and*

$$\Omega_* = \lim_{T\to\infty} \mathsf{E}\{T\left[\bar{g}_T(U_T) - \mu(\theta_0)\right]\left[\bar{g}_T(U_T) - \mu(\theta_0)\right]'\}. \tag{4.3}$$

Assumption 4.2 ASYMPTOTIC EQUIVALENCE FOR EMPIRICAL MOMENTS. *The random vector* $\sqrt{T}\left[\bar{g}_T(\hat{U}_T) - \mu(\theta_0)\right]$ *is asymptotically equivalent to* $\sqrt{T}\left[\bar{g}_T(U_T) - \mu(\theta_0)\right]$, i.e.

$$\operatorname*{plim}_{T\to\infty}\left\{\sqrt{T}\left[\bar{g}_T(\hat{U}_T) - \mu(\theta_0)\right] - \sqrt{T}\left[\bar{g}_T(U_T) - \mu(\theta_0)\right]\right\} = 0. \tag{4.4}$$

Assumption 4.3 ASYMPTOTIC NONSINGULARITY OF WEIGHT MATRIX. $\operatorname*{plim}_{T\to\infty}(\hat{\Omega}_T) = \Omega$ *where* $\det(\Omega) \neq 0$.

Assumption 4.4 ASYMPTOTIC NONSINGULARITY OF WEIGHT MATRIX. $\mu(\theta_0)$ *is twice continuously differentiable in an open neighborhood of* θ_0 *and the Jacobian matrix* $P(\theta_0)$ *has full rank, where* $P(\theta) = \frac{\partial\mu'}{\partial\theta}$.

Given these assumptions, the asymptotic distribution of $\hat{\theta}_T(\hat{\Omega}_T)$ is determined by a standard argument on method-of-moments estimation.

Proposition 4.1 ASYMPTOTIC DISTRIBUTION OF METHOD-OF-MOMENTS ESTIMATOR. *Under the assumptions 4.1 to 4.4,*

$$\sqrt{T}\left[\hat{\theta}_T(\Omega) - \theta_0\right] \xrightarrow{D} \mathrm{N}[0,\, V(\theta_0|\Omega)] \tag{4.5}$$

where

$$V(\theta|\Omega) = \left[P(\theta)\Omega P(\theta)'\right]^{-1} P(\theta)\Omega\Omega_*\Omega P(\theta)' \left[P(\theta)\Omega P(\theta)'\right]^{-1} \tag{4.6}$$

$P(\theta) = \frac{\partial\mu'}{\partial\theta}$. *If, furthermore, (i)* $P(\theta)$ *is a square matrix, or (ii)* Ω_* *is nonsingular and* $\Omega = \Omega_*^{-1}$, *then*

$$V(\theta|\Omega) = \left[P(\theta)\Omega_*^{-1}P(\theta)'\right]^{-1} \equiv V_*(\theta). \tag{4.7}$$

As usual, $V_*(\theta_0)$ is the smallest possible asymptotic covariance matrix for a method-of-moments estimator based on $M_T(\theta)$. The latter, in particular, is reached when the dimensions of μ and θ are the same, in which case the estimator is obtained by solving the equation

$$\bar{g}_T(\hat{U}_T) = \mu\left(\hat{\theta}_T\right).$$

Consistent estimators $V(\theta_0|\Omega)$ and $V_0(\theta_0)$ can be obtained on replacing θ_0 and Ω_* by consistent estimators.

A consistent estimator of Ω_* can easily be obtained [see Newey and West (1987)] by a Bartlett kernel estimator, *i.e.*:

$$\hat{\Omega}_* = \hat{\Gamma}_0 + \sum_{k=1}^{K(T)} \left(1 - \frac{k}{K(T)+1}\right)(\hat{\Gamma}_k + \hat{\Gamma}_k') \tag{4.8}$$

where

$$\hat{\Gamma}_k = \frac{1}{T}\sum_{t=k+1}^{T}[g_{t-k}(\hat{u}) - \mu(\theta)][g_t(\hat{u}) - \mu(\theta)]' \tag{4.9}$$

with θ replaced by a consistent estimator $\tilde{\theta}_T$ of θ. The truncation parameter $K(T) = \tilde{c}T^{1/3}$ is allowed to grow with the sample size such that:

$$\lim_{T\to\infty}\frac{K(T)}{T^{1/2}} = 0 \; ; \tag{4.10}$$

see White and Domowitz (1984). A consistent estimator of $V_*(\theta_0)$ is then given by

$$\hat{V}_* = \left[P(\hat{\theta}_T)\hat{\Omega}_*^{-1}P(\hat{\theta}_T)'\right]^{-1} . \tag{4.11}$$

The main problem here consists in showing that the relevant regularity conditions are satisfied for the estimator $\hat{\theta} = (\hat{a},\, \hat{r}_y,\, \hat{r}_w)'$ given by (3.17)-(3.19) for the parameters of a stochastic volatility model of order one. In this case, we have $\mu(\theta) = [\mu_2(\theta),\, \mu_4(\theta),\, \mu_{2,2}(1|\theta)]'$,

$$\bar{g}_T(\hat{U}_T) = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{U}_T) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 \\ \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4 \\ \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2 \end{pmatrix} , \tag{4.12}$$

$$\bar{g}_T(U_T) = \frac{1}{T}\sum_{t=1}^{T} g_t(U_T) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T}u_t^2 \\ \frac{1}{T}\sum_{t=1}^{T}u_t^4 \\ \frac{1}{T}\sum_{t=1}^{T}u_t^2 u_{t-1}^2 \end{pmatrix} , \tag{4.13}$$

where $g_t(\hat{U}_T) = [\hat{u}_t^2,\, \hat{u}_t^4,\, \hat{u}_t^2\hat{u}_{t-1}^2]'$, and $g_t(U_T) = [u_t^2,\, u_t^4,\, u_t^2 u_{t-1}^2]'$.

Since the number of moments used is equal to the number of parameters (three), the moment estimator can be obtained by taking $\hat{\Omega}_T$ equal to an identity matrix so that Assumption 4.3 automatically holds. The main problem then consists in showing that the assumptions 4.1 and 4.2 are satisfied.

Assumption 4.5 EXISTENCE OF MOMENTS.

$$\underset{T\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T} x_t x_t' = \sigma_{2,x}(0) \; , \tag{4.14}$$

$$\underset{T\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T} x_t u_t^2 x_t' = \sigma_{2,x,u}(0,0) \; , \tag{4.15}$$

$$\underset{T\to\infty}{\text{plim}}\frac{1}{T}\sum_{t=1}^{T} x_t u_{t-1}^2 x_t' = \sigma_{2,x,u}(0,1) \; , \tag{4.16}$$

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t^2x_{t-1}' = \sigma_{2,x,u}(1,0)\ , \tag{4.17}$$

where the $k\times k$ matrices $\sigma_{2,x}(0)$, $\sigma_{2,x,u}(0,0)$, $\sigma_{2,x,u}(0,1)$ and $\sigma_{2,x,u}(1,0)$ are bounded.

Proposition 4.2 ASYMPTOTIC DISTRIBUTION FOR EMPIRICAL MOMENTS. *Under the assumptions 2.1, 2.3 and 2.4 with $L_w = 1$, we have:*

$$\sqrt{T}\big[\bar{g}_T(U_T)-\mu(\theta_0)\big] \xrightarrow{D} \mathrm{N}[0,\ \Omega_*] \tag{4.18}$$

where $\bar{g}_T(U_T)=\sum_{t=1}^{T}g_t/T$, $g_t=[u_t^2,\ u_t^4,\ u_t^2u_{t-1}^2]'$, and

$$\Omega_* = V[g_t] = \mathsf{E}[g_tg_t'] - \mu(\theta_0)\mu(\theta_0)'\,. \tag{4.19}$$

Proposition 4.3 ASYMPTOTIC EQUIVALENCE FOR EMPIRICAL MOMENTS. *Suppose the assumptions 2.1, 2.3, 2.4 and 4.5 hold with $L_w = 1$, let $\hat{\beta}$ be an estimator of β such that*

$$\sqrt{T}(\hat{\beta}-\beta)\ \textit{is asymptotically bounded,} \tag{4.20}$$

and let $\hat{u}_t = y_t - x_t'\hat{\beta}$. Then $\sqrt{T}\big[\bar{g}_T(\hat{U}_T)-\mu(\theta_0)\big]$ is asymptotically equivalent to $\sqrt{T}\big[\bar{g}_T(U_T)-\mu(\theta_0)\big]$.

The fact that condition (4.20) is satisfied by the least squares estimator can be easily seen from earlier published results on the estimation of regression models with stochastic volatility; see Gonçalves and Kilian (2004, Theorem 3.1) and Kuersteiner (2001). Concerning equation (4.14) it holds in particular for the AR(p) case with $x_t = Y_{t-1} = (y_{t-1},\ldots,y_{t-p})'$; see the proofs of Gonçalves and Kilian (2004, Theorems 3.1) .

On assuming that the matrices Ω_* and $P(\theta_0)$ have full rank, the asymptotic normality of $\hat{\theta}_T$ follows as described in Proposition **4.1**. Concerning the latter, it is interesting and potentially useful to note that this asymptotic distribution does not depend on the asymptotic distribution of the first-step estimator of the autoregressive coefficient $(\hat{\beta})$ in the conditional mean equation.

5. Simulation study

In this section, we study by simulation the properties the 2S-3M estimator in terms of bias, variance and root mean square error (RMSE). We consider two different sets of parameters: one set with low serial dependence in the autoregressive dynamics of both processes ($c = 0.3$, $a = 0$), while the other one has high dependence ($c = 0.95$ and $a = 0.95$). For both sets, the scale parameters are fixed at $r_y = 0.5$ and $r_w = 0.5$. Under these designs, the 2S-3M estimator is compared with the GMM estimators of Andersen and Sørensen (1996) based on 5 and 24 moments. The number of replications used is 1000. The results are presented in tables 1 - 3.

Globally, there is no uniform ranking between the different estimators. In several cases, the simple 3-moment estimator exhibits smaller bias and RMSE than the computationally expensive GMM estimator based on 24 moments. The GMM estimator with 5 moments is also clearly dominated by the 2S-3M estimator. In terms of variance, the GMM estimator with 24 moments performs better than the 2S-3M estimator, but its bias is higher.

It is interesting to note that Andersen and Sørensen (1996) studied the choice of the number of moments to include in the overidentified estimation procedure and found that it depends critically on sample size. According to these authors, one should exploit additional moment restrictions when the sample size increases. However, this advice does not appear to be compelling in view of the fact that the 2S-3M estimator can exhibit a better performance even for large samples, such as $T = 1000,\ 2000,\ 5000$. This may be related to theoretical and simulation findings on IV-based inference suggesting that large numbers of instruments may adversely affect estimator precision and test power; see Buse (1992), Chao and Swanson (2000) and Dufour and Taamouti (2003). In particular, overidentification increases the bias of IV and GMM estimators in finite samples. When 24 moments are used, one needs to estimate $24(24+1)/2$ separate entries in the weighting matrix along with the sample moments, and the GMM estimator becomes computationally cumbersome. Furthermore, when the values of the autoregressive parameters get close to the boundaries of the domain, this creates numerical instability in estimating the weight matrix, and the situation gets worse in small samples ($T = 100, 200$). When the sample size is small ($T = 100,\ 200$), RMSE is critically high especially for the autoregressive parameter a; this may be due to the poor behavior of sample moments in small samples.

6. Application to Standard and Poor's price index

In this section, we apply our moment estimator to the daily data on the Standard and Poor's Composite Price Index (SP), over the period 1928-87. This time series was used by Gallant et al. (1997) to estimate a standard stochastic volatility model with the efficient method of moments. The data comprise 16127 daily observations $\{\tilde{y}_t\}_{t=1}^{16,127}$ on adjusted movements of the Standard and poor's Composite Price Index, 1928-87. The raw series is the Standard and Poor's Composite Price Index (SP), daily, 1928-87. The raw series is converted to a price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and then adjusted for systematic calendar effects, that is, systematic shifts in location and scale due to different trading patterns across days of the week, holidays, and year-end tax trading. This yields a variable we shall denote y_t.

The unrestricted estimated value of δ from the data is:

$$\hat{\delta}_T = (0.129,\ 0.926,\ 0.829,\ 0.427)'\ ,$$

$$\hat{\sigma}_T = [0.007,\ 2.89,\ 1.91,\ 8.13]'\ ,$$

where the method-of-moments estimated value of a corresponds to $\hat{a}_T = 0.926$. We may conjecture that there is some persistence in the data during the period 1928-87 what has been statistically checked by performing the three standard tests in a companion paper [see Dufour and Valéry (2005)].

Table 1. Moment estimators: bias

$c = 0.3$, $a = 0$, $r_y = 0.5$, $r_w = 0.5$									
	$T = 100$			$T = 200$			$T = 500$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	-0.2106	-0.0767	0.0780	-0.1554	-0.0522	0.0901	-0.0805	-0.0233	0.0717
r_y	0.0047	-0.0117	-0.0152	0.0044	-0.0021	-0.0064	0.0023	0.0017	-0.0012
r_w	-0.2988	-0.4016	-0.3315	-0.2384	-0.3643	-0.3070	-0.1360	-0.3210	-0.2218
	$T = 1000$			$T = 2000$			$T = 5000$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	-0.0332	0.0052	0.0186	-0.0204	0.0149	0.0186	-0.0062	0.0191	0.0186
r_y	0.0012	0.0026	0.0009	0.0006	0.0019	0.0009	0.0003	0.0012	0.0009
r_w	-0.0685	-0.3097	-0.0485	-0.0328	-0.3026	-0.0485	-0.0127	-0.2074	-0.0485
$c = 0.95$, $a = 0.95$, $r_y = 0.5$, $r_w = 0.5$									
	$T = 100$			$T = 200$			$T = 500$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	-0.2490	-0.2904	-0.3400	-0.1576	-0.2652	-0.1327	-0.0921	-0.3209	-0.0257
r_y	0.2063	0.0801	0.0178	0.1754	0.0422	0.0339	0.1379	0.0124	0.0284
r_w	-0.1240	-0.3307	-0.3024	-0.0817	-0.2240	-0.3146	-0.0687	-0.0843	-0.3215
	$T = 1000$			$T = 2000$			$T = 5000$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	-0.0610	-0.3391	-0.0156	-0.0480	-0.3593	0.0071	-0.0299	-0.3813	0.0256
r_y	0.1149	0.0056	0.0253	0.0890	0.0061	0.0262	0.0639	0.0141	0.0305
r_w	-0.0746	-0.0104	-0.3105	-0.0583	0.0676	-0.2856	-0.0683	0.1988	-0.2461

Table 2. Moment estimators: variance

$c = 0.3,\ a = 0,\ r_y = 0.5,\ r_w = 0.5$									
	$T = 100$			$T = 200$			$T = 500$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.6482	0.3712	0.2914	0.5434	0.3819	0.2986	0.3346	0.3373	0.2947
r_y	0.0019	0.0056	0.0024	0.0010	0.0018	0.0008	0.0005	0.0004	0.0003
r_w	0.0572	0.0423	0.0360	0.0593	0.0557	0.0321	0.0436	0.0827	0.0233
	$T = 1000$			$T = 2000$			$T = 5000$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.1686	0.2103	0.0354	0.0862	0.1027	0.0354	0.0276	0.0304	0.0354
r_y	0.0002	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
r_w	0.0200	0.1119	0.0030	0.0092	0.1432	0.0030	0.0029	0.1252	0.0030
$c = 0.95,\ a = 0.95,\ r_y = 0.5,\ r_w = 0.5$									
	$T = 100$			$T = 200$			$T = 500$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.1796	0.3538	0.3019	0.0751	0.3217	0.1634	0.0343	0.3339	0.0426
r_y	0.1184	0.0815	0.0691	0.0647	0.0458	0.0497	0.0284	0.0177	0.0225
r_w	0.1574	0.0607	0.0633	0.1679	0.0979	0.0481	0.1649	0.1254	0.0325
	$T = 1000$			$T = 2000$			$T = 5000$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.0210	0.3336	0.0414	0.0143	0.3309	0.0172	0.0093	0.2911	0.0003
r_y	0.0143	0.0089	0.0115	0.0073	0.0047	0.0056	0.0040	0.0020	0.0021
r_w	0.1522	0.1484	0.0213	0.1432	0.1546	0.0189	0.1312	0.1709	0.0108

Table 3. Moment estimators: RMSE

$c = 0.3,\ a = 0,\ r_y = 0.5,\ r_w = 0.5$									
	$T = 100$			$T = 200$			$T = 500$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.8318	0.6138	0.5459	0.7530	0.6205	0.5536	0.5837	0.5818	0.5475
r_y	0.0439	0.0759	0.0513	0.0320	0.0434	0.0295	0.0226	0.0203	0.0199
r_w	0.3827	0.4512	0.3822	0.3408	0.4335	0.3555	0.2491	0.4313	0.2694
	$T = 1000$			$T = 2000$			$T = 5000$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.4118	0.4590	0.4759	0.2942	0.3211	0.3561	0.1662	0.1754	0.1891
r_y	0.0155	0.0140	0.0137	0.0113	0.0101	0.0098	0.0078	0.0070	0.0068
r_w	0.1571	0.4559	0.2000	0.1014	0.4852	0.1393	0.0556	0.4100	0.0732
$c = 0.95,\ a = 0.95,\ r_y = 0.5,\ r_w = 0.5$									
	$T = 100$			$T = 200$			$T = 500$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.4914	0.6617	0.6459	0.3159	0.6351	0.4252	0.2069	0.6607	0.2079
r_y	0.4010	0.2964	0.2634	0.3089	0.2180	0.2255	0.2178	0.1338	0.1527
r_w	0.4155	0.4123	0.3933	0.4176	0.3847	0.3835	0.4116	0.3638	0.3686
	$T = 1000$			$T = 2000$			$T = 5000$		
	3 mm.	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**	**3 mm.**	**5 mm.**	**24 mm.**
a	0.1573	0.6696	0.2041	0.1291	0.6780	0.1314	0.1014	0.6605	0.0312
r_y	0.1659	0.0944	0.1102	0.1234	0.0686	0.0797	0.0900	0.0460	0.0553
r_w	0.3970	0.3852	0.3431	0.3828	0.3988	0.3170	0.3685	0.4586	0.2673

7. Conclusion

In this paper, we have provided a computationally simple moment estimator available in closed form and derived its asymptotic distribution for the parameters of a stochastic volatility in a linear regression. Compared with the GMM estimator of Andersen and Sørensen (1996), it demonstrates good statistical properties in terms of bias and RMSE in many situations. Further, it casts doubt on the advice that one should use a large number of moments. In this respect, our just identified estimator underscores that one should not include too many instruments increasing thereby the chance of including irrelevant ones in the estimation procedure. This assertion is documented in the literature on asymptotic theory; see for example, Buse (1992), Chao and Swanson (2000). In particular, overidentification increases bias of IV and GMM estimators in finite samples. Concurring evidence based on finite-sample optimality results and Monte Carlo simulations is also available in Dufour and Taamouti (2003). Further, our closed-form estimator is especially convenient for use in the context of computationally costly inference techniques, such as simulation-based inference methods when asymptotic approximations do not provide reliable inference; see Dufour and Valéry (2005).

A. Appendix: Proofs

PROOF OF LEMMA **3.1** If $U \sim \mathrm{N}(0,\,1)$, then $\mathsf{E}(U^{2p+1}) = 0$, $\forall p \in \mathbb{N}$ and $\mathsf{E}(U^{2p}) = (2p)!/[2^p p!]\ \forall\ p \in \mathbb{N}$; see Gouriéroux and Monfort (1995*a*, Volume 2, page 518). Under Assumption 2.1,

$$\begin{aligned}\mathsf{E}(u_t^k) &= r_y^k \mathsf{E}(z_t^k)\mathsf{E}[\exp(kw_t/2)] \\ &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!}\exp\left[\frac{k^2}{4}r_w^2/2(1-a^2)\right] \\ &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!}\exp\left[\frac{k^2}{8}r_w^2/(1-a^2)\right] \qquad \text{(A.1)}\end{aligned}$$

where the second equality uses the definition of the Gaussian Laplace transform of $w_t \sim \mathrm{N}[0,\, r_w^2/(1-a^2)]$ and of the moments of the $\mathrm{N}(0,\,1)$ distribution. Let us now calculate the cross-product:

$$\begin{aligned}\mathsf{E}[u_t^k u_{t+m}^l] &= \mathsf{E}\left[r_y^{k+l} z_t^k z_{t+m}^l \exp(k\frac{w_t}{2} + l\frac{w_{t+m}}{2})\right] \\ &= r_y^{k+l}\mathsf{E}(z_t^k)\mathsf{E}(z_{t+m}^l)\mathsf{E}[\exp(k\frac{w_t}{2} + l\frac{w_{t+m}}{2})] \\ &= r_y^{k+l}\frac{k!}{2^{(k/2)}(k/2)!}\frac{l!}{2^{(l/2)}(l/2)!}\exp\left[\frac{r_w^2}{8(1-a^2)}(k^2+l^2+2kla^m)\right] \qquad \text{(A.2)}\end{aligned}$$

where $\mathsf{E}(w_t) = 0$, $\mathsf{Var}(w_t) = r_w^2/(1-a^2)$ and

$$\begin{aligned}\mathsf{Var}\left(k\frac{w_t}{2} + l\frac{w_{t+m}}{2}\right) &= \frac{k^2}{4}\mathsf{Var}(w_t) + \frac{l^2}{4}\mathsf{Var}(w_{t+m}) + 2\frac{k}{2}\frac{l}{2}\mathsf{Cov}(w_t,\, w_{t+m}) \\ &= \frac{r_w^2}{4(1-a^2)}(k^2+l^2+2kla^m)\ . \qquad \text{(A.3)}\end{aligned}$$

□

PROOF OF LEMMA **3.2** By equations (3.6) and (3.7), we get:

$$\frac{\mathsf{E}(u_t^4)}{[\mathsf{E}(u_t^2)]^2} = 3\exp[r_w^2/(1-a^2)]\ , \qquad \text{(A.4)}$$

hence

$$r_w^2/(1-a^2) = \log\left(\frac{\mathsf{E}(u_t^4)}{3(\mathsf{E}(u_t^2))^2}\right) \equiv Q\ . \qquad \text{(A.5)}$$

Inserting $Q \equiv r_w^2/(1-a^2)$ in (3.6) yields

$$r_y = \left(\frac{\mathsf{E}(u_t^2)}{\exp(Q/2)}\right)^{1/2} = \frac{3^{1/4}\mathsf{E}(u_t^2)}{\mathsf{E}(u_t^4)^{1/4}} \ . \tag{A.6}$$

From equation (3.8), we have

$$\exp(\frac{r_w^2}{(1-a)}) = \frac{\mathsf{E}[u_t^2 u_{t-1}^2]}{r_y^4} \tag{A.7}$$

which, after a few manipulations, yields

$$1 + a = \frac{[\log(\mathsf{E}[u_t^2 u_{t-1}^2]) - 4\log(r_y)]}{Q} \tag{A.8}$$

and

$$a = \frac{[\log(\mathsf{E}[u_t^2 u_{t-1}^2]) - \log(3) - 4\log(\mathsf{E}[u_t^2]) + \log(\mathsf{E}[u_t^4])]}{\log\left(\frac{\mathsf{E}[u_t^4]}{3(\mathsf{E}[u_t^2])^2}\right)} - 1 \ . \tag{A.9}$$

Finally, from (A.5), we get:

$$r_w = \left[(1-a^2)\log\left(\frac{\mathsf{E}[u_t^4]}{3(\mathsf{E}[u_t^2])^2}\right)\right]^{1/2} \ . \tag{A.10}$$

□

PROOF OF LEMMA **3.3** Here we derive the covariances of the components of $X_t = (X_{1t},\, X_{2t},\, X_{3t})'$:

$$\begin{aligned}
\gamma_1(\tau) &= \mathsf{Cov}(X_{1t},\, X_{1,t+\tau}) = \mathsf{E}\{[u_t^2 - \mu_2(\theta)][u_{t+\tau}^2 - \mu_2(\theta)]\} \\
&= \mathsf{E}(u_t^2 u_{t+\tau}^2) - \mu_2^2(\theta) = r_y^4 \mathsf{E}[\exp(w_t + w_{t+\tau})] - \mu_2^2(\theta) \\
&= r_y^4 \exp\left[\frac{r_w^2}{1-a^2}(1+a^\tau)\right] - \mu_2^2(\theta) = \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1] \ ,
\end{aligned} \tag{A.11}$$

where $\gamma = r_w^2/(1-a^2)$. Similarly,

$$\begin{aligned}
\gamma_2(\tau) &= \mathsf{Cov}(X_{2t},\, X_{2,t+\tau}) = \mathsf{E}\{[u_t^4 - \mu_4(\theta)][u_{t+\tau}^4 - \mu_4(\theta)]\} \\
&= \mathsf{E}(u_t^4 u_{t+\tau}^4) - \mu_4^2(\theta) = 9 r_y^8 \mathsf{E}\{\exp[2(w_t + w_{t+\tau})]\} - \mu_4^2(\theta) \\
&= 9 r_y^8 \exp[4\frac{r_w^2}{1-a^2}(1+a^\tau) - \mu_4^2(\theta)] = \mu_4^2(\theta)[\exp(4\gamma a^\tau) - 1] \ .
\end{aligned} \tag{A.12}$$

Finally,

$$\gamma_3(\tau) = \mathsf{Cov}(X_{3t},\, X_{3,t+\tau}) = \mathsf{E}\{[u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)][u_{t+\tau}^2 u_{t+\tau-1}^2 - \mu_{2,2}(1|\theta)]\}$$

$$
\begin{aligned}
&= \mathsf{E}[u_t^2 u_{t-1}^2 u_{t+\tau}^2 u_{t+\tau-1}^2] - \mu_{2,2}^2(1|\theta) \\
&= r_y^8 \mathsf{E} \exp(w_{t+\tau} + w_{t+\tau-1} + w_t + w_{t-1}) - \mu_{2,2}^2(1|\theta) \\
&= r_y^8 \exp[2(1+a)\gamma] \exp[\gamma(a^{\tau-1} + 2a^\tau + a^{\tau+1})] - \mu_{2,2}^2(1|\theta) \\
&= \mu_{2,2}^2(1|\theta)\{\exp[\gamma(a^{\tau-1} + 2a^\tau + a^{\tau+1})] - 1\} \\
&= \mu_{2,2}^2(1|\theta)\{\exp[\gamma(1+a)^2 a^{\tau-1}] - 1\} \qquad \text{(A.13)}
\end{aligned}
$$

for all $\tau \geq 2$. □

PROOF OF PROPOSITION **4.1** The method-of-moments estimator $\hat{\theta}_T(\Omega)$ is solution of the following optimization problem:

$$
\min_\theta M_T(\theta) = \min_\theta [\mu(\theta) - \bar{g}_T(\hat{U}_T)]' \hat{\Omega}_T [\mu(\theta) - \bar{g}_T(\hat{U}_T)] \; . \qquad \text{(A.14)}
$$

The first order conditions (F.O.C) associated with this problem are:

$$
\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}_T[\mu(\hat{\theta}_T) - \bar{g}_T(\hat{U}_T)] = 0 \; . \qquad \text{(A.15)}
$$

An expansion of the F.O.C above around the true value θ yields

$$
\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}_T[\mu(\theta) + P(\theta)'(\hat{\theta}_T - \theta) - \bar{g}_T(\hat{U}_T)] = O_p(T^{-1}) \qquad \text{(A.16)}
$$

where, after rearranging the equation,

$$
\sqrt{T}\big[\hat{\theta}_T(\Omega) - \theta\big] = [P(\theta)\Omega P(\theta)']^{-1} P(\theta)\Omega\sqrt{T}\big[\bar{g}_T(\hat{U}_T) - \mu(\theta)\big] + O_p(T^{-1/2}) \; . \qquad \text{(A.17)}
$$

Using Assumptions 4.1 to 4.4, we get the asymptotic normality of $\hat{\theta}_T(\Omega)$ with asymptotic covariance matrix $V(\Omega)$ as specified in Proposition **4.1**. □

PROOF OF PROPOSITION **4.2** In order to establish the asymptotic normality of $\sqrt{T}[\bar{g}_T(U_T) - \mu(\theta))]$, we shall use a central limit theorem (C.L.T) for dependent processes [see Davidson (1994, Theorem 24.5, page 385)]. For that purpose, we first check the conditions under which this C.L.T holds. Setting

$$
X_t \equiv \begin{pmatrix} u_t^2 - \mu_2(\theta) \\ u_t^4 - \mu_4(\theta) \\ u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta) \end{pmatrix} = g_t(\theta) - \mu(\theta) \; , \qquad \text{(A.18)}
$$

$$
S_T = \sum_{t=1}^T X_t = \sum_{t=1}^T \big[g_t(\theta) - \mu(\theta)\big] \; , \qquad \text{(A.19)}
$$

and the subfields $\mathcal{F}_t = \sigma(s_t,\, s_{t-1},\, \ldots)$ where $s_t = (y_t,\, w_t)'$, we need to check three conditions:
a) $\{X_t,\, \mathcal{F}_t\}$ is stationary and ergodic,
b) $\{X_t,\, \mathcal{F}_t\}$ is a L_1-mixingale of size -1,

c)

$$\limsup_{T\to\infty} T^{-1/2}\mathsf{E}|S_T| < \infty \tag{A.20}$$

in order to get that $T^{-1/2}S_T = \sqrt{T}[\bar{g}_T(U_T) - \mu(\theta)] \xrightarrow{D} \mathrm{N}[0,\ \Omega_*]$.
a) By propositions 5 and 17 from Carrasco and Chen (2002), we can say that:
(i) if $\{w_t\}$ is geometrically ergodic, then $\{(w_t,\, \ln|v_t|)\}$ is Markov geometrically ergodic with the same decay rate as the one of $\{w_t\}$;
(ii) if $\{w_t\}$ is stationary β-mixing with a certain decay rate, then $\{\ln|v_t|\}$ is β-mixing with a decay rate at least as fast as the one of $\{w_t\}$.
If the initial value v_0 follows the stationary distribution, $\{\ln|v_t|\}$ is strictly stationary β-mixing with an exponential decay rate. Since this property is preserved by any continuous transformation, $\{v_t\}$ and hence $\{v_t^k\}$ and $\{v_t^k v_{t-1}^k\}$ are strictly stationary and exponential β-mixing. We can then deduce that X_t is strictly stationary and exponential β-mixing.
b) A mixing zero-mean process is an adapted L_1-mixingale with respect to the subfields $\mathcal{F}_t$ provided it is bounded in the L_1-norm [see Davidson (1994, Theorem 14.2, page 211)]. To see that $\{X_t\}$ is bounded in the L_1-norm, we note that:

$$\mathsf{E}|v_t^2 - \mu_2(\theta)| \le \mathsf{E}(|v_t^2| + |\mu_2(\theta)|) = 2\mu_2(\theta) < \infty\,, \tag{A.21}$$

$$\mathsf{E}|v_t^4 - \mu_4(\theta)| \le 2\mu_4(\theta) < \infty\,, \tag{A.22}$$

$$\mathsf{E}|v_t^2 v_{t-1}^2 - \mu_{2,2}(1|\theta)| \le 2\mu_{2,2}(1|\theta) < \infty\,. \tag{A.23}$$

We now need to show that the L_1-mixingale $\{X_t,\, \mathcal{F}_t\}$ is of size -1. Since X_t is $\beta-$mixing, it has mixing coefficients of the type $\beta_n = c\rho^n\ ,\ c > 0,\ 0 < \rho < 1$. In order to show that $\{X_t\}$ is of size -1, we need to show that its mixing coefficients $\beta_n = O(n^{-\phi})$, with $\phi > 1$. Indeed,

$$\frac{\rho^n}{n^{-\phi}} = n^{\phi}\exp(n\log\rho) = \exp(\phi\log n)\exp(n\log\rho) = \exp(\phi\log n + n\log\rho)\,. \tag{A.24}$$

It is known that $\lim_{n\to\infty} \phi\log n + n\log\rho = -\infty$ which yields

$$\lim_{n\to\infty} \exp(\phi\log n + n\log\rho) = 0\,. \tag{A.25}$$

This holds in particular for $\phi > 1$, [see Rudin (1976, Theorem 3.20(d), page 57)].
c) By the Cauchy-Schwarz inequality, we have:

$$\mathsf{E}|T^{-1/2}S_T| \le T^{-1/2}\|S_T\|_2 \tag{A.26}$$

so that (A.20) can be proven by showing that $\limsup_{T\to\infty} T^{-1}\mathsf{E}(S_T S_T') < \infty$.
We shall prove that:

$$\limsup_{T\to\infty} T^{-1}\mathsf{E}(S_T S_T') = \limsup_{T\to\infty} \mathsf{Var}\left[\frac{1}{\sqrt{T}}S_T\right] < \infty\ . \tag{A.27}$$

i) First and second components of S_T. Set $S_{T1} = \sum_{t=1}^{T} X_{1,t}$ where $X_{1,t} \equiv u_t^2 - \mu_2(\theta)$. We compute:

$$\begin{aligned}\mathsf{Var}\left[\frac{1}{\sqrt{T}}S_{T1}\right] &= \frac{1}{T}\left[\sum_{t=1}^{T}\mathsf{Var}(X_{1,t}) + \sum_{\substack{t=1\\ s\neq t}}^{T}\mathsf{Cov}(X_{1,s},\, X_{1,t})\right] \\ &= \frac{1}{T}\left[T\gamma_1(0) + 2\sum_{\tau=1}^{T}(T-\tau)\gamma_1(\tau)\right] = \gamma_1(0) + 2\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_1(\tau)\,, \quad \text{(A.28)}\end{aligned}$$

where $\gamma \equiv r_w^2/(1-a^2)$. We must prove that $\sum_{\tau=1}^{T}(1-\frac{\tau}{T})\gamma_1(\tau)$ converge as $T \to \infty$. By Lemma 3.1.5 in Fuller (1976, page 112), it is sufficient to show that $\sum_{\tau=1}^{\infty}\gamma_1(\tau)$ converge. Using Lemma **3.3**, we have:

$$\begin{aligned}\gamma_1(\tau) &= \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1] = \mu_2^2(\theta)\left[1 + \sum_{k=1}^{\infty}\frac{(\gamma a^\tau)^k}{k!} - 1\right] = \mu_2^2(\theta)\left[\gamma a^\tau \sum_{k=1}^{\infty}\frac{(\gamma a^\tau)^{k-1}}{k!}\right] \\ &= \mu_2^2(\theta)\left[\gamma a^\tau \sum_{k=0}^{\infty}\frac{(\gamma a^\tau)^k}{(k+1)!}\right] \leq \mu_2^2(\theta)\gamma a^\tau \sum_{k=0}^{\infty}\frac{(\gamma a^\tau)^k}{k!} = \mu_2^2(\theta)\gamma a^\tau \exp(\gamma a^\tau)\,. \quad \text{(A.29)}\end{aligned}$$

Therefore, the series

$$\begin{aligned}\sum_{\tau=1}^{\infty}\gamma_1(\tau) &\leq \mu_2^2(\theta)\gamma\sum_{\tau=1}^{\infty}a^\tau\exp(\gamma a^\tau) \leq \mu_2^2(\theta)\gamma\exp(\gamma a)\sum_{\tau=1}^{\infty}a^\tau \\ &= \mu_2^2(\theta)\frac{a\gamma\exp(\gamma a)}{1-a} < \infty \quad \text{(A.30)}\end{aligned}$$

converges. We deduce by the Cauchy-Schwarz inequality that

$$\limsup_{T\to\infty} T^{-1/2}\mathsf{E}\left|\sum_{t=1}^{T}\left[u_t^2 - \mu_2(\theta)\right]\right| < \infty\,. \quad \text{(A.31)}$$

The proof is very similar for the second component of S_T.
ii) Third component of S_T. Set $S_{T3} = \sum_{t=1}^{T} X_{3,t}$ where $X_{3,t} \equiv u_t^2u_{t-1}^2 - \mu_{2,2}(1|\theta)$. Likewise, we just have to show that $\sum_{\tau=1}^{\infty}\gamma_3(\tau) < \infty$ in order to prove that

$$\limsup_{T\to\infty} T^{-1/2}\mathsf{E}\left|\sum_{t=1}^{T}\left[u_t^2u_{t-1}^2 - \mu_{2,2}(1|\theta)\right]\right| < \infty\,. \quad \text{(A.32)}$$

By lemma **3.3** we have for all $\tau \geq 2$:

$$\gamma_3(\tau) = \mu_{2,2}^2(1|\theta)[\exp(\gamma(1+a)^2a^{\tau-1}) - 1]$$

$$
\begin{aligned}
&= \mu_{2,2}^2(1|\theta)\left\{1+\sum_{k=1}^{\infty}\frac{[\gamma(1+a)^2a^{\tau-1}]^k}{k!}-1\right\} \\
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2a^{\tau-1}]\sum_{k=1}^{\infty}\frac{[\gamma(1+a)^2a^{\tau-1}]^{k-1}}{k!} \\
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2a^{\tau-1}]\sum_{k=0}^{\infty}\frac{[\gamma(1+a)^2a^{\tau-1}]^{k}}{(k+1)!} \\
&\leq \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2a^{\tau-1}]\sum_{k=0}^{\infty}\frac{[\gamma(1+a)^2a^{\tau-1}]^{k}}{k!} \\
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2a^{\tau-1}]\exp[\gamma(1+a)^2a^{\tau-1}]\ , \qquad \text{(A.33)}
\end{aligned}
$$

hence

$$
\begin{aligned}
\sum_{\tau=1}^{\infty}\gamma_3(\tau) &\leq \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\sum_{\tau=2}^{\infty}a^{\tau-1}\exp[\gamma(1+a)^2a^{\tau-1}] \\
&\leq \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\sum_{\tau=2}^{\infty}a^{\tau-1} \\
&= \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\sum_{\tau=1}^{\infty}a^{\tau} \\
&= \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\frac{a}{1-a}<\infty\ . \qquad \text{(A.34)}
\end{aligned}
$$

Since $\limsup_{T\to\infty} T^{-1/2}\mathsf{E}\left|\sum_{t=1}^{T}X_t\right| < \infty$, we can therefore apply Theorem 24.5 of Davidson (1994) to each component S_{Ti} , $i = 1,\ 2,\ 3$ of S_T to state that: $T^{-1/2}S_{Ti} \xrightarrow{D} \mathrm{N}(0,\ \lambda_i)$ and then by the Cramér-Wold theorem, establish the limiting result for the 3×1-vector S_T using the stability property of the Gaussian distribution, *i.e.*,

$$
T^{-1/2}S_T = T^{-1/2}\sum_{t=1}^{T}X_t = \sqrt{T}\left[\bar{g}_T(U_T)-\mu(\theta)\right] \xrightarrow{D} \mathrm{N}_3(0,\ \Omega_*)\ , \qquad \text{(A.35)}
$$

where

$$
\Omega_* = \lim_{T\to\infty}\mathsf{E}\left[\left(T^{-1/2}S_T\right)^2\right] = \lim_{T\to\infty}\mathsf{E}\left\{T\left[\bar{g}_T(U_T)-\mu(\theta)\right]\left[\bar{g}_T(U_T)-\mu(\theta)\right]'\right\}\ .
$$

□

PROOF OF PROPOSITION **4.3** The asymptotic equivalence of

$$\sqrt{T}\left[\bar{g}_T(\hat{U}_T)-\mu(\theta)\right]=\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2-\mu_2(\theta)\right] \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^4-\mu_4(\theta)\right] \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta)\right] \end{pmatrix} \tag{A.36}$$

with $\sqrt{T}\left[\bar{g}_T(U_T)-\mu(\theta)\right]$ can be established by looking at each component.

1. **Component** $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2-\mu_2(\theta)\right]$. We have:

$$\begin{aligned} \hat{u}_t^2-u_t^2 &= (y_t-x_t'\hat{\beta})^2-(y_t-x_t'\beta)^2 \\ &= (\hat{\beta}-\beta)'x_tx_t'(\hat{\beta}-\beta)-2(\hat{\beta}-\beta)'x_tu_t\ . \end{aligned} \tag{A.37}$$

We deduce after aggregation:

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2-\mu_2(\theta)\right] &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^2-\mu_2(\theta)\right]+\frac{1}{\sqrt{T}}\left[\sqrt{T}(\hat{\beta}-\beta)\right]'\left[\frac{1}{T}\sum_{t=1}^{T}x_tx_t'\right]\sqrt{T}(\hat{\beta}-\beta) \\ &\quad -2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_tu_t\ . \end{aligned} \tag{A.38}$$

By (4.20), Assumption 4.5 and the Law of Large Numbers (LLN), we have

$$\frac{1}{T}\sum_{t=1}^{T}x_tu_t\to\mathsf{E}(x_tu_t)=\mathsf{E}(x_t)\mathsf{E}(u_t)=0 \tag{A.39}$$

and equation (A.38) is equivalent to

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2-\mu_2(\theta)\right]=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^2-\mu_2(\theta)\right]+o_p(1) \tag{A.40}$$

asymptotically.

2. **Component** $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^4-\mu_4(\theta)\right]$. Noting that

$$\begin{aligned} \hat{u}_t^4-u_t^4 &= -4(\hat{\beta}-\beta)'x_tu_t^3+6(\hat{\beta}-\beta)'x_tu_t^2x_t'(\hat{\beta}-\beta)-4(\hat{\beta}-\beta)'x_tx_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_tu_t \\ &\quad +(\hat{\beta}-\beta)'x_tx_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_tx_t'(\hat{\beta}-\beta) \end{aligned} \tag{A.41}$$

we get after aggregation:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^4-\mu_4(\theta)\right]=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^4-\mu_4(\theta)\right]+R_T \tag{A.42}$$

where

$$
\begin{aligned}
R_T \equiv & -4\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t u_t^3 + \frac{6}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t u_t^2 x_t'\right]\sqrt{T}(\hat{\beta}-\beta) \\
& -4\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]\sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right] \\
& +\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]\sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]\sqrt{T}(\hat{\beta}-\beta)\ .
\end{aligned}
$$

Since $\sqrt{T}(\hat{\beta}-\beta)$, $\frac{1}{T}\sum_{t=1}^{T} x_t u_t^2 x_t'$ and $\frac{1}{T}\sum_{t=1}^{T} x_t x_t'$ are asymptotically bounded, while by the LLN, $\frac{1}{T}\sum_{t=1}^{T} x_t u_t^3$ and $\frac{1}{T}\sum_{t=1}^{T} x_t u_t$ converge to zero, we can conclude that R_T is an $o_p(1)$-variable, which yields

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^4-\mu_4(\theta)\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^4-\mu_4(\theta)\right] + o_p(1)\ .
$$

3. **Component** $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta)\right]$. Since

$$
\begin{aligned}
\hat{u}_t^2\hat{u}_{t-1}^2 - u_t^2 u_{t-1}^2 = & -2(\hat{\beta}-\beta)'\left[x_t u_t u_{t-1}^2 + x_{t-1}u_t^2 u_{t-1}\right] + (\hat{\beta}-\beta)' x_t u_{t-1}^2 x_t'(\hat{\beta}-\beta) \\
& +(\hat{\beta}-\beta)' x_{t-1}u_t^2 x_{t-1}'(\hat{\beta}-\beta) + 4(\hat{\beta}-\beta)' x_t u_t u_{t-1} x_{t-1}'(\hat{\beta}-\beta) \\
& -2(\hat{\beta}-\beta)' x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)' x_{t-1}u_{t-1} \\
& -2(\hat{\beta}-\beta)' x_{t-1}x_{t-1}'(\hat{\beta}-\beta)(\hat{\beta}-\beta)' x_t u_t \\
& +(\hat{\beta}-\beta)' x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)' x_{t-1}x_{t-1}'(\hat{\beta}-\beta)\ , \qquad \text{(A.43)}
\end{aligned}
$$

we have:

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta)\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^2 u_{t-1}^2-\mu_{2,2}(1|\theta)\right] + R_T \qquad \text{(A.44)}
$$

where

$$
\begin{aligned}
R_T \equiv & -2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}\left[x_t u_t u_{t-1}^2 + x_{t-1}u_t^2 u_{t-1}\right] \\
& +\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t u_{t-1}^2 x_t'\sqrt{T}(\hat{\beta}-\beta) \\
& +\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_{t-1}u_t^2 x_{t-1}'\sqrt{T}(\hat{\beta}-\beta) \\
& +4\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t u_t u_{t-1}x_{t-1}'\sqrt{T}(\hat{\beta}-\beta)
\end{aligned}
$$

$$
\begin{aligned}
&-2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_{t-1} \\
&-2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_{t-1}x_{t-1}'\sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t u_t \\
&+\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_t x_t'\sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}x_{t-1}x_{t-1}'\sqrt{T}(\hat{\beta}-\beta)\ .
\end{aligned}
$$

By (4.20), Assumption 4.5 and the LLN applied to $\frac{1}{T}\sum_{t=1}^{T}x_{t-1}u_t^2u_{t-1}$, which converges to

$$
\mathsf{E}\big[x_{t-1}u_t^2u_{t-1}\big] = \mathsf{E}\big[u_t^2\mathsf{E}(u_{t-1}|\mathcal{F}_{t-2})\mathsf{E}(x_{t-1}|\mathcal{F}_{t-2})\big] = 0\ , \tag{A.45}
$$

we deduce that R_T is an $o_p(1)$-variable. Therefore, we have the asymptotic equivalence:

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta)\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^2u_{t-1}^2-\mu_{2,2}(1|\theta)\right]+o_p(1)\ . \tag{A.46}
$$

Hence,

$$
\begin{pmatrix}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2-\mu_2(\theta)\right] \\
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^4-\mu_4(\theta)\right] \\
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta)\right]
\end{pmatrix}
=
\begin{pmatrix}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^2-\mu_2(\theta)\right] \\
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^4-\mu_4(\theta)\right] \\
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[u_t^2u_{t-1}^2-\mu_{2,2}(1|\theta)\right]
\end{pmatrix}
+o_p(1)
$$

and

$$
\sqrt{T}\left[\bar{g}_T(\hat{U}_T)-\mu(\theta)\right] = \sqrt{T}\left[\bar{g}_T(U_T)-\mu(\theta)\right]+o_p(1)\ , \tag{A.47}
$$

with $\bar{g}_T(U_T)$ defined as in equation (4.13). □

References

Andersen, T. G., Chung, H.-J. and Sørensen, B. E. (1999), ‘Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study’, *Journal of Econometrics* **91**, 61–87.

Andersen, T. G. and Sørensen, B. E. (1996), ‘GMM estimation of a stochastic volatility model: A Monte Carlo study’, *Journal of Business and Economic Statistics* **14**(3), 328–352.

Buse, A. (1992), ‘The bias of instrumental variables estimators’, *Econometrica* **60**, 173–180.

Carrasco, M. and Chen, X. (2002), ‘Mixing and moment properties of various GARCH and stochastic volatility models’, *Econometric Theory* **18**(1), 17–39.

Chao, J. and Swanson, N. R. (2000), Bias and MSE of the IV estimators under weak identification, Technical report, Department of Economics, University of Maryland.

Danielsson, J. (1994), ‘Stochastic volatility in asset prices: Estimation with simulated maximum likelihood’, *Journal of Econometrics* **61**, 375–400.

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

Davidson, J. (1994), *Stochastic Limit Theory: An Introduction for Econometricians*, second edn, Oxford University Press, Oxford, U.K.

Dufour, J.-M. and Taamouti, M. (2003), Point-optimal instruments and generalized Anderson-Rubin procedures for nonlinear models, Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M. and Valéry, P. (2005), Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models, Technical report, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) and Centre interuniversitaire de recherche en économie quantitative (CIREQ), Université de Montréal.

Engle, R. F. (1982), ‘Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation’, *Econometrica* **50**(4), 987–1008.

Fuller, W. A. (1976), *Introduction to Statistical Time Series*, John Wiley & Sons, New York.

Gallant, A. R., Hsieh, D. and Tauchen, G. (1997), ‘Estimation of stochastic volatility models with diagnostics’, *Journal of Econometrics* **81**(1), 159–192.

Ghysels, E., Harvey, A. C. and Renault, E. (1996), Stochastic volatility, *in* (Maddala and Rao 1996), pp. 119–183.

Gonçalves, S. and Kilian, S. (2004), ‘Bootstrapping autoregressions with conditional heteroskedasticity of unknown form’, *Journal of Econometrics* **123**, 89–120.

Gouriéroux, C. (1997), *ARCH Models and Financial Applications*, Springer Series in Statistics, Springer-Verlag, New York.

Gouriéroux, C. and Monfort, A. (1995*a*), *Simulation Based Econometric Methods*, CORE Lecture Series, CORE Foundation, Louvain-la-Neuve.

Gouriéroux, C. and Monfort, A. (1995*b*), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K.

Hansen, L. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica* **50**, 1029–1054.

Harvey, A., Ruiz, E. and Shephard, N. (1994), 'Multivariate stochastic variance models', *Review of Economic Studies* **61**, 247–264.

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

Kim, S., Shephard, N. and Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**, 361–393.

Kuersteiner, G. M. (2001), 'Optimal instrumental variables estimation for ARMA models', *Journal of Econometrics* **104**(2), 349–405.

Maddala, G. S. and Rao, C. R., eds (1996), *Handbook of Statistics 14: Statistical Methods in Finance*, North-Holland, Amsterdam.

Mahieu, R. J. and Schotman, P. C. (1998), 'An empirical application of stochastic volatility models', *Journal of Applied Econometrics* **13**, 333–360.

Melino, A. and Turnbull, S. (1990), 'Pricing foreign currency options with stochastic volatility', *Journal of Econometrics* **45**, 239–265.

Nelson, D. (1988), Times Series Behavior of Stock Market Volatility and Returns, PhD thesis, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Newey, W. K. and McFadden, D. (1994), Large sample estimation and hypothesis testing, *in* R. F. Engle and D. L. McFadden, eds, 'Handbook of Econometrics, Volume IV', North-Holland, Amsterdam, chapter 36, pp. 2111–2245.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**, 703–708.

Palm, F. C. (1996), GARCH models of volatility, *in* (Maddala and Rao 1996), pp. 209–240.

Rudin, W. (1976), *Principles of Mathematical Analysis, Third Edition*, McGraw-Hill, New York.

Ruiz, E. (1994), 'Quasi-maximum likelihood estimation of stochastic variance models', *Journal of Econometrics* **63**, 284–306.

Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility, *in* D. R. Cox, O. E. Barndorff-Nielsen and D. V. Hinkley, eds, 'Time Series Models in Econometrics, Finance and Other Fields', Chapman & Hall, London, U.K.

Taylor, S. (1986), *Modelling Financial Time Series*, John Wiley and Sons, New York.

White, H. and Domowitz, I. (1984), 'Nonlinear regression with dependent observations', *Econometrica* **52**, 143–161.

Wong, C. K. J. (2002*a*), Estimating Stochastic Volatility: New Method and Comparison, PhD thesis, Oxford, U.K.

Wong, C. K. J. (2002*b*), The MCMC method for an extended stochastic volatility model, Technical report, Oxford University, Oxford, U.K.