

Indirect Inference, Nuisance Parameter, and Threshold Moving Average Models

Alain GUAY

Departement des Sciences Economiques, Université du Québec à Montréal and CREFE,
Montréal, P.Q. H3C 3P8, Canada (guay.alain@uqam.ca)

Olivier SCAILLET

HEC Genève and FAME (olivier.scaillet@hec.unige.ch)

We analyze the modifications that occur in indirect inference when a nuisance parameter is not identified under the null hypothesis. We develop a testing procedure adapted to this simulation-based estimation method, and detail its use for detecting the threshold effect in threshold moving average models with contemporaneous and lagged asymmetries. In contrast to existing threshold models, these models allow taking into account the presence of asymmetric effects of current and lagged random shocks. We use them to measure the persistence of shocks to U.S. output.

KEY WORDS: Indirect inference; Nuisance parameter; Threshold model.

1. INTRODUCTION

This article considers inference when a nuisance parameter is not identified under the null hypothesis in the context of simulation-based econometric methods. Simulation-based econometric methods are increasingly used in economics and finance for models which were previously thought to be too complex for a proper estimation. The reason is that they allow the estimation of models where the optimizing function takes no simple analytical form. In such models, the difficulty may, for example, arise from the presence of multidimensional integrals in the likelihood function or in the moment conditions. In this context, conventional econometric methods cannot be directly used since they require an explicit form for the optimizing function. Simulation-based econometric methods are designed to overcome such a numerical difficulty through an approach based on simulated data. The only requirement to implement these indirect methods is that the model can be simulated. They include the simulated method of moments (Duffie and Singleton 1993), indirect inference (Gouriéroux, Monfort, and Renault 1993), and the efficient method of moments (Gallant and Tauchen 1996).

In this article, we examine the problem occurring when a nuisance parameter is not identified under the null in the context of indirect estimation methods. This problem is well studied in the direct estimation context [see Andrews and Ploberger (1994), Hansen (1996), among others]. Indirect estimation methods are characterized by the use of an auxiliary model and simulation paths from the structural model. The purpose of our article is to analyze the problem of a nuisance parameter, which is not identified under the null in this setup. In the case of structural change tests with unknown breakpoint, Ghysels and Guay (2001a,b) show, for simulation-based estimation methods, that the asymptotic distribution of standard tests is free of nuisance parameters. In that setting, the parameter that appears under the alternative, but not under the null, is the time of structural change. However, the asymptotic distribution of standard tests generally depends upon unknown parameters. In this article, we analyze the modifications that occur in indirect methods in the general case. To

do so, we modify the strategy proposed by Hansen (1996) in the direct estimation context. In particular, we show how to deal with inference when additional uncertainty is introduced due to simulations. The general framework is developed for estimation by indirect inference. As shown by Gouriéroux and Monfort (1995a), indirect inference contains the simulated method of moments (SMM) and the efficient method of moments (EMM) as special cases. The results derived in this article are thus directly applicable to these methods.

The general setting is applied to threshold moving average (TMA) models in the context of an analysis of the persistence of shocks to output. By relaxing the hypothesis of symmetry, Beaudry and Koop (1993) and Elwood (1998) allow the disentanglement of effects of positive shocks versus negative shocks. Beaudry and Koop (1993) use an exogenous proxy to represent shocks, and provide evidence of asymmetric effects of innovations to the gross national product (GNP) [see Elwood (1998) for a critic of the specification used by Beaudry and Koop (1993) to identify positive and negative shocks]. Negative shocks to GNP seem to be less persistent than positive shocks. In contrast to Beaudry and Koop (1993), Elwood (1998) identifies directly positive and negative shocks by using an unobserved component model corresponding to a threshold moving average model. He finds no evidence of asymmetry in the persistence of shocks to output. However, his modeling excludes an a priori asymmetric effect of contemporaneous shocks, and imposes a threshold depending on the sign of shocks.

In this article, we consider a more flexible model to analyze the persistence of shocks to output. This model takes the form of the following threshold moving average model [TMA(1)]:

$$Y_t = \mu + d_0^+ \epsilon_t \mathbb{I}_{\epsilon_t > \gamma} + d_0^- \epsilon_t \mathbb{I}_{\epsilon_t \leq \gamma} \\ + \dots + d_l^+ \epsilon_{t-l} \mathbb{I}_{\epsilon_{t-l} > \gamma} + d_l^- \epsilon_{t-l} \mathbb{I}_{\epsilon_{t-l} \leq \gamma} \quad (1)$$

where $\mathbb{I}_A$ takes the value 1 if A is true and 0 otherwise. In model (1), successive shocks are not transmitted symmetrically through time, but their influence does depend on their exceedance of the threshold γ . The main feature of model (1) is thus its ability to capture asymmetric effects induced by the random shocks. As opposed to Elwood (1998), we do not impose a threshold value, and we introduce contemporaneous asymmetry. Such an asymmetry, if present, will affect the measurement of persistence of shocks.

Threshold moving average (TMA) models have already been considered in Tong (1990), but without any contemporaneous asymmetry [see also related work by Wecker (1981) and De Gooijer (1998)]. This extension parallels the construction proposed for the conditional variance by El Babsiri and Zakoian (2001) in a generalized autoregressive conditional heteroscedastic (GARCH) context. The contemporaneous asymmetry induces second moments, which differ depending on the threshold. The introduction of the contemporaneous asymmetry in model (1) prevents a direct approach by maximum likelihood. Therefore, we need to proceed by indirect methods to estimate the parameters of this type of TMA model.

In empirical work, the presence of contemporaneous and lagged asymmetries needs testing. The inference is not standard because a nuisance parameter (the threshold) is not identified under the null hypothesis. We apply the general results developed in the first part of the article to test for the presence of asymmetry in the moving average representation.

The article is organized as follows. In Section 2, we present the general framework and testing problem. The setting is sufficiently large to cover a broad category of dynamic models. In Section 3, we describe how to accommodate the indirect inference procedure of Gouriéroux, Monfort, and Renault (1993) in the presence of a nuisance parameter. The general framework and results are then applied to TMA models in Section 4. Properties, estimation, and tests for such models are therein detailed, the threshold being the nuisance parameter. Some Monte Carlo results are also proposed. An empirical application is delivered in Section 5, and consists of an analysis of the persistence of shocks to U.S. GNP growth rates. Section 6 contains some concluding remarks. All proofs are gathered in the Appendix. The usual notational conventions are used in the article. $\|\cdot\|$ denotes the Euclidean norm of a vector or a matrix, $\|\cdot\|_r$ denotes the L^r norm of a random vector, that is, $\|X\|_r = (E\|X\|^r)^{1/r}$, and the symbol $\Rightarrow$ denotes weak convergence as defined in Pollard (1994).

2. GENERAL FRAMEWORK AND TESTING PROBLEM

We consider a multivariate stationary process: $w_t = (y_t', x_t')$ where y_t is a G -dimensional vector and x_t is a K -dimensional vector. We assume that the true conditional pdf of w_t given $\underline{w}_{t-1} = \{w_{t-1}, w_{t-2}, \dots\}$ (w.r.t. some measure ν) only depends on $w_{t-1}, \dots, w_{t-k}$, for some k . This pdf is denoted $f_0(w_t|\underline{w}_{t-1})$, and can be written as $f_0(w_t|\underline{w}_{t-1}) = f_{0y}(y_t|x_t, \underline{w}_{t-1})f_{0x}(x_t|\underline{w}_{t-1})$, where f_{0y} and f_{0x} are pdf w.r.t. some measures μ_x and μ_y , respectively.

We are interested in modeling f_{0y} , and consider $M = \{f(y_t|x_t, \underline{w}_{t-1}; \theta, \gamma), \theta \in \Theta, \gamma \in \Gamma\}$, a parametric family of pdf, where Θ and Γ are compact and bounded subsets of $\mathbf{R}^q$ and $\mathbf{R}^m$, respectively. In short, M is a parametric model for the conditional distribution of the process y_t given the process x_t , where θ will be the parameter of interest and γ the nuisance parameter. The parameter θ is decomposed into $\theta = (\theta_1', \theta_2')'$, where $\theta_1 \in \mathbf{R}^{q_1}$, $\theta_2 \in \mathbf{R}^{q_2}$, and $q = q_1 + q_2$. For notational convenience, $f(y_t|x_t, \underline{w}_{t-1}; \theta, \gamma)$ is denoted hereafter as $f_t(\theta, \gamma)$.

Our testing problem can be described as follows. The null hypothesis is

$$H_0 : \{\theta_2 = 0\}$$

while the alternative hypothesis is

$$H_1 : \{\theta_2 \neq 0\}, \text{ and the model } M \text{ depends on the parameter } \gamma.$$

We let θ_0 denote a parameter vector in the null hypothesis. Under the null, we assume that $f_{0y} = f_t(\theta_0, \gamma)$ does not depend on the parameter γ . The parameter γ is not identified under the null, and has to be treated as a nuisance parameter. The testing procedure is therefore not standard as we treat γ as unknown. As usual, in this setting, we adopt a local-to-null reparameterization $\theta_2 = c/\sqrt{T}$, and the null hypothesis becomes $H_0 : \{c = 0\}$.

3. INDIRECT INFERENCE WITH NUISANCE PARAMETER

We consider a situation where maximum likelihood estimation or estimation by a method of moments are not feasible, and an indirect procedure is called for. Indeed, maximum likelihood estimators are not always available, and one should then rely on auxiliary or instrumental models through indirect procedures. The terminology auxiliary model versus instrumental model can be used interchangeably [see Dhaene, Gouriéroux, and Scaillet (1998) for interpretation]. We modify here the results of Hansen (1996) to account for this indirect estimation, and derive the asymptotic distribution theory under $H_0 : \{c = 0\}$. In the following, we sketch briefly the indirect inference procedure, and refer the reader to Gouriéroux, Monfort and Renault (1993) and Gouriéroux and Monfort (1995a) for details [see also Smith (1993)]. As already pointed out, SMM and EMM can be viewed as particular cases of indirect inference.

Let us assume that we can draw freely some paths from the conditional model for a given value of the parameters. We denote by $y_t^n(\theta, \gamma)$, $t = 1, \dots, T$, $n = 1, \dots, N$, the components of the N drawn paths. To assist us in the estimation, we choose an instrumental criterion characterized by $Q_T(y_T|x_T, \underline{w}_{T-1}; \beta) = \sum_{t=1}^T q_t(y_t|x_t, \underline{w}_{t-1}; \beta)$, $\beta \in B$ with B a compact subset of $\mathbf{R}^p$ and $p \geq q$. This criterion may, for example, correspond to the opposite of a likelihood function. Further, let us introduce the M -estimators of β :

$$\hat{\beta}_T = \arg \min_{\beta \in B} Q_T(y_T|x_T, \underline{w}_{T-1}; \beta)$$

and

$$\hat{\beta}_T^n(\theta, \gamma) = \arg \min_{\beta \in B} Q_T(y_T^n(\theta, \gamma) | x_T, \underline{w}_{T-1}^n(\theta, \gamma); \beta, \gamma)$$

$$\hat{\beta}_T^N(\theta, \gamma) = \frac{1}{N} \sum_{n=1}^N \hat{\beta}_T^n(\theta, \gamma)$$

computed from the data and the simulated values, respectively. The binding function $b(\theta, \gamma)$, for a given value of γ , will be the limiting value of $\hat{\beta}_T^N(\theta, \gamma)$ (Gouriéroux and Monfort 1995a). Under the null, the binding function does not depend on the parameter γ , and is denoted $\beta_0 = b(\theta_0)$. The indirect estimator $\hat{\theta}_T^N(\gamma)$ is defined as

$$\hat{\theta}_T^N(\gamma) = \arg \min_{\theta \in \Theta} (\hat{\beta}_T - \hat{\beta}_T^N(\theta, \gamma))' \hat{\Omega}_T(\gamma) (\hat{\beta}_T - \hat{\beta}_T^N(\theta, \gamma))$$

where $\hat{\Omega}_T(\gamma)$ is a positive-definite matrix converging to a deterministic positive matrix $\Omega(\gamma)$ for a fixed γ . Optimal indirect estimators are obtained with a weighting matrix corresponding to an estimator of $\Omega^*(\gamma)$ defined below in assumption (p).

The scores associated with the instrumental model are denoted by

$$s_{q,i}(\beta) = \nabla_{\beta} q_i(\beta), \quad s_{q,i}^n(\beta, \theta, \gamma) = \nabla_{\beta} q_i^n(\beta, \theta, \gamma)$$

with $q_i(\beta) = q_i(y_i | x_i, \underline{w}_{i-1}; \beta)$, and $q_i^n(\beta, \theta, \gamma) = q_i(y_i^n(\theta, \gamma) | x_i, \underline{w}_{i-1}^n(\theta, \gamma); \beta, \gamma)$, $n = 1, \dots, N$. We require the following high-level assumptions.

Assumption 1.

(a) $\hat{\beta}_T$, $\hat{\beta}_T^N(\theta, \gamma)$, and $\hat{\Omega}_T(\gamma)$ do not depend on γ for all θ in the null hypothesis.

(b) $\hat{\beta}_T \xrightarrow{p} \beta_0$ under the null for some β_0 in the interior of B .

(c) $\sup_{\gamma \in \Gamma} \|\hat{\beta}_T^N(\theta, \gamma) - b(\theta, \gamma)\| \xrightarrow{p} 0$ under the null for some nonrandom function $b(\theta, \gamma)$ and for $\theta \in \Theta_0$ where Θ_0 is some neighborhood of θ_0 .

(d) $\hat{\beta}_T^N(\theta, \gamma)$ is continuously partially differentiable in θ for all $\theta \in \Theta_0$ and $\gamma \in \Gamma$ with probability 1 under the null.

(e) $b(\theta, \gamma)$ is continuously partially differentiable in θ for all $\theta \in \Theta_0$ and $\gamma \in \Gamma$.

(f) $\sup_{\gamma \in \Gamma} \|\nabla_{\theta} \hat{\beta}_T^N(\theta, \gamma) - \nabla_{\theta} b(\theta, \gamma)\| \xrightarrow{p} 0$ under the null for $\theta \in \Theta_0$.

(g) $\nabla_{\theta} b(\theta_0, \gamma)$ has full column rank for all $\gamma \in \Gamma$ and $\beta_0 = b(\theta_0, \gamma)$ only when $\theta = \theta_0$ for all $\gamma \in \Gamma$.

(h) $q_i(\beta)$ is twice continuously partially differentiable in β for all $\beta \in B_0$ with probability 1 under the null, where B_0 is some neighborhood of β_0 .

(i) $q_i^n(\beta, \theta, \gamma)$ is twice continuously partially differentiable in β for all $\beta \in B_0$, $\theta \in \Theta_0$ and $\gamma \in \Gamma$ with probability 1 under the null.

(j) $T^{-1} \sum_{i=1}^T \nabla_{\beta} s_{q,i}(\beta) \xrightarrow{p} J_q(\beta, \gamma)$ uniformly over $\gamma \in \Gamma$, and $\beta \in B_0$ under the null for some nonrandom function $J_q(\beta, \gamma)$ uniformly continuous in (β, γ) over $B_0 \times \Gamma$.

(k) $T^{-1} \sum_{i=1}^T \nabla_{\beta} s_{q,i}^n(b(\theta, \gamma), \theta, \gamma) \xrightarrow{p} J_q(b(\theta, \gamma), \gamma)$ uniformly over $\gamma \in \Gamma$, and $\theta \in \Theta_0$ under the null for $n = 1, \dots, N$, and for some nonrandom function $J_q(b(\theta, \gamma), \gamma)$ uniformly continuous in $(b(\theta, \gamma), \gamma)$ over $\Theta_0 \times \Gamma$.

(l) $J_q(\beta_0, \gamma)$ is uniformly positive definite over $\gamma \in \Gamma$.

(m) $T^{-1/2} \sum_{i=1}^T s_{q,i}(\beta_0) \implies S_q(\beta_0, \cdot)$ under the null as Gaussian processes indexed by $\gamma \in \Gamma$ with mean zero and covariance matrix $I_q(\beta_0, \cdot)$.

(n) $T^{-1/2} \sum_{i=1}^T s_{q,i}^n(\beta_0, \theta_0, \cdot) \implies S_q^n(\beta_0, \cdot)$ under the null as Gaussian processes indexed by $\gamma \in \Gamma$ with mean zero and covariance matrix $I_q(\beta_0, \cdot)$.

(o) $\hat{\Omega}_T^*(\gamma) \xrightarrow{p} \Omega^*(\beta_0, \gamma)$ uniformly for some nonrandom function over $\gamma \in \Gamma$.

(p) $\Omega^*(\beta_0, \gamma) = J_q(\beta_0, \gamma) I_q(\beta_0, \gamma)^{-1} J_q(\beta_0, \gamma)$ is uniformly positive definite over $\gamma \in \Gamma$, where $I_q(\beta_0, \gamma) = V_{as}[S_q(\beta_0, \gamma)]$.

Assumption (a) gives an indirect estimator, which does not depend on the nuisance parameter under the null. Assumptions (b)–(g) are classical in indirect inference, and are only slightly modified to handle the presence of a nuisance parameter. They concern the behavior of the estimators and their limits. In particular, the injectivity of the binding function and the rank condition on its derivative embodied by assumption (g) are global and local identifiability conditions for indirect inference (Dhaene, Gouriéroux, and Scaillet 1998). Note that we only require differentiability w.r.t. the parameter of interest θ , and *not* w.r.t. the nuisance parameter γ . Assumptions (h), (i), (k), and (l) are standard in the context of M estimation. In assumption (j), the probability limit of the score function for the data is indexed by the nuisance parameter γ since the derivative of the limit criterion of the instrumental depends on θ and γ [see Gouriéroux, Monfort, and Renault (1993)]. Assumptions (m) and (n) are relative to the weak convergence of the score functions. In particular, assumption (n) corresponds to the score of the M estimator for the simulated path n . Assumptions (o) and (p) on the weighting matrix lead to optimal indirect estimators.

The Wald statistic $W_T(\gamma)$ for the null hypothesis $H_0 : \{c = 0\}$ is given by

$$W_T(\gamma) = T \hat{\theta}_T^N(\gamma)' R (R' \hat{V}_T(\gamma) R)^{-1} R' \hat{\theta}_T^N(\gamma)$$

where R is the selector matrix $R = (0 \quad Id_{q_2})'$, and

$$\hat{V}_T(\gamma) = \left(1 + \frac{1}{N}\right) \times [\nabla_{\theta} \hat{\beta}_T^N(\hat{\theta}_T^N(\gamma), \gamma)' \hat{\Omega}_T^*(\gamma) \nabla_{\theta} \hat{\beta}_T^N(\hat{\theta}_T^N(\gamma), \gamma)]^{-1}.$$

The score statistic is the gradient of the objective function w.r.t. θ evaluated at the constrained estimator $\hat{\theta}_T^N$ obtained by fixing $\theta_2 = 0$ in the simulations:

$$C_T(\gamma) = \nabla_{\theta} \hat{\beta}_T^N(\hat{\theta}_T^N, \gamma) \Omega_T^*(\gamma) (\hat{\beta}_T - \hat{\beta}_T^N(\hat{\theta}_T^N, \gamma)).$$

The test statistic is

$$LM_T(\gamma) = T C_T(\gamma)' \tilde{V}_T(\gamma) C_T(\gamma)$$

where $\tilde{V}_T(\gamma)$ is the variance-covariance matrix of $C_T(\gamma)$ [see Gouriéroux, Monfort, and Renault (1993) for the expression of this matrix].

Finally, the LR -type test is based on the difference of the optimal values of the objective function for the constrained and unconstrained estimators:

$$LR_T(\gamma) = \frac{TN}{1+N} [(\hat{\beta}_T - \hat{\beta}_T^N(\tilde{\theta}_T^N))' \hat{\Omega}_T^*(\gamma) (\hat{\beta}_T - \hat{\beta}_T^N(\tilde{\theta}_T^N))] - \frac{TN}{1+N} \times [(\hat{\beta}_T - \hat{\beta}_T^N(\hat{\theta}_T^N, \gamma))' \hat{\Omega}_T^*(\gamma) (\hat{\beta}_T - \hat{\beta}_T^N(\hat{\theta}_T^N, \gamma))].$$

The following theorem gives the asymptotic distribution of the concentrated indirect estimator and the $W_T(\gamma)$, $LM_T(\gamma)$, and $LR_T(\gamma)$ statistics.

Theorem 1. Under $H_0 : \{c = 0\}$ and Assumption 1,

$$\sqrt{T}(\hat{\theta}_T^N(\cdot) - \theta_0) \implies -\bar{K}_q(\beta_0, \theta_0, \cdot) \bar{S}_q^N(\beta_0, \cdot)$$

and

$$W_T(\cdot), LM_T(\cdot), LR_T(\cdot) \implies \bar{S}_q^N(\beta_0, \cdot)' \bar{K}_q(\beta_0, \theta_0, \cdot)' R \times [R' \bar{V}_q(\beta_0, \theta_0, \cdot) R]^{-1} R' \bar{K}_q(\beta_0, \theta_0, \cdot) \bar{S}_q^N(\beta_0, \cdot)$$

as processes indexed by $\gamma \in \Gamma$, with

$$\bar{S}_q^N(\beta_0, \gamma) = J_q(\beta_0, \gamma)^{-1} \left(S_q(\beta_0, \gamma) - N^{-1} \sum_{n=1}^N S_q^n(\beta_0, \gamma) \right) \\ \bar{K}_q(\cdot) = [\nabla_\theta b(\theta_0, \cdot)' \Omega^*(\beta_0, \cdot) \nabla_\theta b(\theta_0, \cdot)]^{-1} \times \nabla_\theta b(\theta_0, \cdot)' \Omega^*(\beta_0, \cdot)$$

and

$$\bar{V}_q(\cdot) = \left(1 + \frac{1}{N} \right) [\nabla_\theta b(\theta_0, \cdot)' J_q(\beta_0, \gamma) I_q(\beta_0, \gamma)^{-1} \times J_q(\beta_0, \gamma) \nabla_\theta b(\theta_0, \cdot)]^{-1}.$$

Proof. See Appendix.

The asymptotic distribution of $W_T(\gamma)$, $LM_T(\gamma)$, and $LR_T(\gamma)$ statistics is a chi-square distribution for a given γ . For $\gamma \in \Gamma$, one can then build statistics such as $g(W_T(\gamma))$, where g maps functionals on Γ to $\mathbf{R}$. Davis (1977, 1987) suggested using the supremum (Sup), and Andrews and Ploberger (1994) show that superior local power can be obtained by the average (Ave) of the statistics on Γ or by an exponential transformation. However, the asymptotic null distribution of these mappings depends in general on the true parameter value θ_0 , and critical values cannot be tabulated.

Hansen (1996) proposes a nice remedy, namely, a p -value transformation, based on a simple simulation technique in order to obtain empirical estimates of asymptotic p values. His methodology can be easily adapted to our framework. It consists of working conditionally on the sample and simulated paths and using iid $N(0, 1)$ draws to build conditional mean zero Gaussian processes with appropriate second-moment characteristics. Let $v_i, v_i^n, n = 1, \dots, N$, be $N + 1$ independent standard normal variables. We set $\hat{S}_{q,T} =$

$T^{-1/2} \sum_{i=1}^T \hat{s}_{q,i} v_i, \hat{S}_{q,T}^n(\gamma) = T^{-1/2} \sum_{i=1}^T \hat{s}_{q,i}^n(\gamma) v_i^n$, where $\hat{s}_{q,i}$ and $\hat{s}_{q,i}^n(\gamma)$ denote the estimates of the scores. The difference between direct and indirect estimation lies in the presence of the N estimates of the scores corresponding to the N simulated paths. The four steps described in Hansen (1996) can then be performed by using $\bar{S}_{q,T}^N(\gamma) = J_{q,T}(\gamma)^{-1} (\hat{S}_{q,T} - \frac{1}{N} \sum_{n=1}^N \hat{S}_{q,T}^n(\gamma))$ at the second step, and the finite sample statistic corresponding to the asymptotic distribution of the statistic tests (Theorem 1) at the third step to obtain the empirical p values. Since we operate conditionally on the sample and the simulated paths, all randomness appears in the iid variables v_i and $v_i^n, n = 1, \dots, N$. Therefore, as in Hansen (1996), we may conclude that $\hat{S}_{q,T}$ and $\hat{S}_{q,T}^n(\cdot)$ converge weakly in probability to $S_q(\beta_0, \cdot)$ and $S_q^n(\beta_0, \cdot)$, respectively, as Gaussian processes indexed by $\gamma \in \Gamma$ with mean zero and covariance matrix $I_q(\beta_0, \cdot)$. By the Glivenko–Cantelli theorem, the simulated p values converge in probability to the true one. It is important to note that the asymptotic covariance matrix $\bar{V}_q(\beta_0, \gamma)$ is scaled by a factor $(1 + 1/N)$, which contrasts with the direct estimation case. This scale factor depends on the number of simulated paths. In Section 4.2, a small sample study examines the impact of the number of simulated paths on the estimation and testing properties for our application.

4. THRESHOLD MOVING AVERAGE MODELS

In this section, we outline the threshold moving average model which will be used in the empirical section to measure asymmetry in the persistence of shocks to output. Basic properties of this model are presented before developing estimation and testing strategies.

Testing procedures involving a nuisance parameter have already been used to detect a threshold effect in threshold autoregressive models [see, e.g., Chan (1990), Hansen (1996)] or parameter instability [see, e.g., Andrews (1993), Andrews and Ploberger (1994), Ghysels, Guay, and Hall (1998), Ghysels and Guay (2001a,b), and Sowell (1996)]. In this article, we want to study the threshold model:

$$Y_t = \mu + D^+(L) \epsilon_t^+ + D^-(L) \epsilon_t^- \quad (2)$$

with $\epsilon_t^+ = \epsilon_t \mathbb{1}_{\epsilon_t > \gamma}$, $\epsilon_t^- = \epsilon_t \mathbb{1}_{\epsilon_t \leq \gamma}$, and $\epsilon_t \sim \text{iid } N(0, 1)$. The nuisance parameter γ corresponds to the threshold governing the asymmetric effects of the random shocks. The polynomials of order $L : D^+(L) = d_0^+ + d_1^+ L + \dots + d_L^+ L^L$ and $D^-(L) = d_0^- + d_1^- L + \dots + d_L^- L^L$ drive the moving average dynamics of the error term. The introduction of d_0^+ and d_0^- leads to the presence of a contemporaneous asymmetry [see Zakoian (1994) for ARCH asymmetries and El Babsiri and Zakoian (2001) for similar contemporaneity in the conditional variance]. The parameter of interest $\theta = (\mu, d_0^+, d_1^+, \dots, d_L^+, d_0^-, d_1^-, \dots, d_L^-)'$ is made of the mean and the coefficients of the moving average polynomials.

Wecker (1981) proposes asymmetric moving average models with a threshold depending on the error term, but the case with contemporaneous asymmetry is not considered, and the threshold is arbitrarily fixed to zero. Elwood (1998) employs the asymmetric moving average model proposed by Wecker to study the asymmetry in the persistence of shocks to output. In

contrast to Elwood, we allow for contemporaneous asymmetry, and we do not restrict the threshold to be equal to zero. In the following, we describe how to estimate asymmetric moving average models with contemporaneous asymmetry, and how to test the presence of asymmetry without a priori imposing the value of the threshold. De Gooijer (1998) also considers threshold moving average models with a contemporaneous effect. However, his threshold depends on the observed variable [self-exciting threshold moving average (SETMA) models].

Let us first examine the basic features of model (2) before moving to the estimation and testing procedures.

4.1 Basic Properties

In model (2), the conditional distribution of Y_t is a mixture of the distributions of ϵ_t^+ and ϵ_t^- . This will lead to skewed and leptokurtic (or platykurtic) conditional distributions. To quantify this effect, one needs to define an information set to yield truly iid errors. The required information set must include past values of ϵ_t . Indeed, the information set including only past values of the observed process is not sufficient to obtain iid errors. The required information set at time $t-1$, I_{t-1} is then defined as $I_{t-1} = \{\epsilon_{t-1}\}$. Let us introduce the strong innovation process U_t of Y_t , defined by $U_t = Y_t - E[Y_t | \epsilon_{t-1}] = d_0^+ \eta_t^+ + d_0^- \eta_t^-$, where $\eta_t^+ = \epsilon_t^+ - E[\epsilon_t^+]$, and $\eta_t^- = \epsilon_t^- - E[\epsilon_t^-]$ are iid centered processes, independent of the past, and where d_0^+ and d_0^- are the first coefficients of the polynomials $D^+(L)$ and $D^-(L)$.

Proposition 1. In model (2) when $\gamma = 0$, the conditional skewness of Y_t is given by

$$\frac{E[U_t^3 | \epsilon_{t-1}]}{E[U_t^2 | \epsilon_{t-1}]^{3/2}} = \frac{[(d_0^+)^3 - (d_0^-)^3](\pi + 2) + 3[(d_0^+)^2 d_0^- - d_0^+ (d_0^-)^2](\pi - 2)}{[(d_0^+)^2 + (d_0^-)^2](\pi - 1) + 2d_0^+ d_0^-}^{3/2}$$

and the conditional kurtosis of Y_t by

$$\frac{E[U_t^4 | \epsilon_{t-1}]}{E[U_t^2 | \epsilon_{t-1}]^2} = \frac{[(d_0^+)^4 + (d_0^-)^4](6\pi^2 - 10\pi - 3) + 4[(d_0^+)^3 d_0^- + d_0^+ (d_0^-)^3] \times (\pi + 3) + 6(d_0^+)^2 (d_0^-)^2 (2\pi - 3)}{[(d_0^+)^2 + (d_0^-)^2](\pi - 1) + 2d_0^+ d_0^-}^2.$$

Proposition 1 gives the conditional skewness and kurtosis of Y_t when the threshold γ is equal to zero. Expressions for any γ can also be computed thanks to the general forms given in the Appendix. Eventually, unconditional skewness and kurtosis can also be derived, but the computations involve tedious algebra. The conditional moments w.r.t. the information set $I_{t-1}^* = Y_{t-1}$ can be computed by simulations. The marginal and conditional skewness and kurtosis of Y based on the information set $I_{t-1}(I_{t-1} \supset I_{t-1}^*)$ will imply marginal and conditional skewness and kurtosis of Y based on the information set I_{t-1}^* .

The next proposition is related to the autocovariance structure of a TMA(l) model with $\gamma = 0$.

Proposition 2. In a threshold moving average model with l lags and $\gamma = 0$, the autocovariances $\tilde{\gamma}(u)$ are given by

$$\tilde{\gamma}(u) = \sum_{i=0}^{l-u} \left((d_{i+u}^+ d_i^+ + d_{i+u}^- d_i^-) \frac{\pi - 1}{2\pi} + (d_{i+u}^+ d_i^- + d_{i+u}^- d_i^+) \frac{1}{2\pi} \right), \quad u \leq l$$

$$= 0, \quad u > l.$$

From this proposition, it transpires that the autocorrelation function (ACF) $\{\rho(u) = \tilde{\gamma}(u)/\tilde{\gamma}(0); u = 1, 2, \dots\}$ of a series following a TMA(l) model is zero for lags $u > l$. Thus, it will be difficult to separate this model from an MA(l) model on the basis of the cutoff point of the sample ACF. Even though second-order features are not relevant to distinguish between TMA(l) and MA(l) models, possible skewness and excess kurtosis in the series will lead to favoring the former instead of the latter. The threshold testing procedure developed in the next section will also be of further help in making such a decision.

4.2 Estimation and Tests

The $q \times 1$ parameter $\theta = (\mu, d_0^+, d_1^+, \dots, d_l^+, d_0^-, d_1^-, \dots, d_l^-)'$ cannot be estimated by maximum likelihood. Indeed, due to the contemporaneous asymmetry, we cannot adopt a direct approach, and we need to rely on indirect estimation.

We consider here simple linear instrumental models in estimation by indirect inference. Indeed, Gouriéroux, Monfort, and Renault (1993) have shown in the case of an MA(1) process that the indirect inference estimator obtained with an AR(3) instrumental model has good finite sample properties, and outperforms the asymptotically most efficient estimator. Since a TMA model boils down to a standard MA model under the null, that is, the absence of a threshold effect, we also choose linear autoregressive processes as instrumental models. The first one is a pure linear AR(p) process with the following log-likelihood function:

$$-\frac{(T-p)}{2} \ln 2\pi - \frac{(T-p)}{2} \ln \sigma^2 - \sum_{t=p+1}^T \frac{(Y_t - \alpha - \Phi_1 Y_{t-1} - \dots - \Phi_p Y_{t-p})^2}{2\sigma^2}.$$

This type of instrumental model is particularly appealing since the PML estimator of the parameter $\beta = (\alpha, \phi_1, \dots, \phi_p, \sigma^2)$ of dimension larger or equal to q corresponds to the OLS estimator. Regularity conditions concerning the behavior of the binding and score functions are further met (see Theorem 2 below) since the instrumental model is a regression type model, and Y_t takes a linear form in θ (a pure MA model under the null).

The second instrumental model tries to capture nonlinearity in the observed process in order to obtain more precise estimates of the parameters of interest. To do so, we introduce the second and third polynomials for the lags of Y_t in the AR representation as proposed by Michaelides and Ng (2000).

The log-likelihood function is

$$-\frac{(T-p)}{2} \ln 2\pi - \frac{(T-p)}{2} \ln \sigma^2 - \sum_{t=p+1}^T \frac{(Y_t - \alpha - \Phi_1 Y_{t-1} - \Phi_2 Y_{t-1}^2 - \Phi_3 Y_{t-1}^3 - \dots - \Phi_{3 \times p} Y_{t-p}^3)^2}{2\sigma^2}.$$

In order to evaluate the performance of these instrumental models, we execute a simulation study. Two dimensions are of interest in the design of experiments: the order of the AR process and the number of simulated paths.

The model of interest for the simulation study is the following threshold moving average model of order 1:

$$Y_t = d_0^+ \epsilon_t \mathbb{I}_{\epsilon_t > 0} + d_0^- \epsilon_t \mathbb{I}_{\epsilon_t \leq 0} + d_1^+ \epsilon_{t-1} \mathbb{I}_{\epsilon_{t-1} > 0} + d_1^- \epsilon_{t-1} \mathbb{I}_{\epsilon_{t-1} \leq 0} \quad (3)$$

where the threshold is fixed to zero, and $d_0^+ = .5$, $d_0^- = 1$, $d_1^+ = .2$, and $d_1^- = .8$. In this representation, the negative shocks have a stronger impact than the positive ones. The sample size is equal to 200, and the number of replications is 500.

How can we know whether we should take into account a threshold effect in the dynamics of the shocks in model (3)? By writing $\theta_1 = (d_0^+, d_1^+)'$, $\theta_2 = (d_0^-, d_1^-)' - (d_0^+, d_1^+)',$ we see that the answer is given by testing the null hypothesis $H_0 : \{\theta_2 = 0\}$. The following theorem justifies the application of results in Section 3 to the model of interest and instrumental models under consideration.

Theorem 2. Under the model of interest and the instrumental models defined above, assumptions (a)–(p) hold. The asymptotic distribution of Wald, LM, and LR statistics is given under the null by Theorem 1, and under the alternative by

$$\sqrt{T} R'(\hat{\theta}_T^N(\cdot) - \theta_0) \implies (-R' \bar{K}_q(\beta_0, \theta_0, \cdot) \bar{S}_q^N(\beta_0, \cdot)) + c$$

and

$$\begin{aligned} W_T(\cdot), LM_T(\cdot), LR_T(\cdot) &\implies ((R' \bar{K}_q(\beta_0, \theta_0, \cdot) \bar{S}_q^N(\beta_0, \cdot)) + c)' \\ &\quad \times [R' \bar{V}_q(\beta_0, \theta_0, \cdot) R]^{-1} \\ &\quad \times ((R' \bar{K}_q(\beta_0, \theta_0, \cdot) \bar{S}_q^N(\beta_0, \cdot)) + c) \end{aligned}$$

as processes indexed by $\gamma \in \Gamma$.

Proof. See Appendix.

For the first instrumental model, we consider AR(4), AR(6), AR(8), and AR(10) processes. The results for the bias and the root mean square errors (RMSE's) are reported in Table 1.

Table 1. Instrumental Model I

Parameters	AR(4)		AR(6)		AR(8)		AR(10)	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
d_0^+	-.0874	.2689	-.0864	.2912	-.1040	.3432	-.0903	.3601
d_0^-	-.0964	.2523	-.1037	.2943	-.0871	.3171	-.0915	.3417
d_1^+	.1126	.3281	.0987	.3628	.0661	.4058	.0492	.4168
d_1^-	.0360	.2611	.0255	.3022	.0125	.3484	.0008	.3759

Table 2. Instrumental Model I With AR(4)

Parameters	N = 1		N = 2		N = 5		N = 10	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
d_0^+	-.0874	.2689	-.1089	.2466	-.1084	.2394	-.1237	.2479
d_0^-	-.0964	.2523	-.0640	.2145	-.0693	.1996	-.0666	.1930
d_1^+	.1126	.3281	.1159	.2855	.1086	.2569	.1016	.2496
d_1^-	.0360	.2611	.0406	.2210	.0686	.2164	.0651	.2087

We note that the estimates are not very precise. While the increase of the order of the AR process decreases the bias, the extra parameters included in the instrumental model increase the variance of the estimates. Indeed, the smallest RMSE is obtained with an AR(4) process.

Table 2 reports the results for various number of simulated paths with the instrumental model corresponding to an AR(4) process. The precision of the estimates is greatly improved when increasing the number of simulated paths. The bias decreases for the parameters corresponding to the negative shocks. It increases for the parameters corresponding to the positive shocks. The importance of the reduction gain for RMSE appears particularly important for $N = 2$ and $N = 5$. The difference between $N = 5$ and $N = 10$ is negligible.

The results for instrumental model II are gathered in Table 3. Introducing nonlinearity in the conditional mean decreases the bias for the parameters corresponding to the positive shocks, and slightly increases the bias for the parameters corresponding to the negative shocks. In all cases, the precision of the estimates is significantly improved. The AR(2) representation seems to give the best results in terms of bias and RMSE.

In conclusion, parsimonious representations give the best results in light of the root mean square error criterion. Accounting for nonlinearity by the introduction of polynomials in the autoregressive instrumental model improves the precision of the estimates.

We consider a third instrumental model which is a mix of the two previous instrumental models. The model is the following:

$$\begin{aligned} Y_t = & \Phi_1 Y_{t-1} + \Phi_2 Y_{t-1}^2 + \Phi_3 Y_{t-1}^3 + \Phi_4 Y_{t-2} + \Phi_5 Y_{t-2}^2 \\ & + \Phi_6 Y_{t-2}^3 + \Phi_7 Y_{t-3} + \Phi_8 Y_{t-4} + \epsilon_t. \end{aligned}$$

This representation includes the specifications which yields the best performance for the first and second instrumental models.

Table 4 reports the results for various number of simulated paths. For $N = 1$, we see that this instrumental model clearly

Table 3. Instrumental Model II

Parameters	AR(1)		AR(2)		AR(3)		AR(4)	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
d_0^+	-.0010	.2100	-.0134	.1953	.0087	.2142	.0180	.2073
d_0^-	-.1561	.2974	-.0925	.2366	-.1159	.2541	-.1201	.2534
d_1^+	.1373	.3938	.0104	.2238	.0147	.2468	.0286	.2360
d_1^-	.0503	.2592	.0336	.2697	.0121	.2667	.0160	.2461

Table 4. Instrumental Model III

Parameters	N = 1		N = 2		N = 5		N = 10	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
d_0^+	-.0324	.1561	-.0154	.1550	-.0105	.1440	-.0070	.1336
d_0^-	-.0431	.1930	-.0578	.1781	-.0646	.1816	-.0616	.1722
d_1^+	.0196	.2028	.0046	.1571	-.0069	.1479	-.0072	.1410
d_1^-	.0118	.2142	.0204	.2077	.0270	.1937	.0161	.1874

outperforms the two other instrumental models in terms of RMSE with a large reduction in bias. We observe the same pattern as that of the other instrumental models when the number of simulated paths increases. In conclusion, the bias is small with this instrumental model, and the precision of the estimates is acceptable for a sample of 200 observations.

Let us now perform a simulation study to evaluate the small sample properties of the adaptation of the Hansen (1996) strategy to indirect inference. In this assessment of the size and power of the testing procedure, we keep the model given by Equation (3). Under the null, it corresponds to an MA(1) process. The parameter d_0 is fixed to 1 and d_1 to .5, while sample sizes are equal to 200. Due to the large computational requirements of the simulation design, only 200 simulated samples are generated. The interval for the threshold was set to $[-.5, .5]$, with evaluation points spaced by .05.

Two mappings of the Wald and LM statistics are examined: supremum (Sup) and average (Ave). The asymptotic distribution of these statistics under the null is approximated with 500 replications.

Table 5 contains the results where N equals 1, 2, and 5. For the case of a single simulated path, the estimation is too imprecise. Consequently, the null hypothesis is never rejected. However, the results with $N = 2$ and $N = 5$ are more encouraging. The functions based on the Wald statistics reject the null too frequently. The functions based on the LM statistics give results close to the corresponding size. In particular, the average LM statistic seems to reject the null with accurate size.

To evaluate the power of the tests, we consider Equation (3) with $d_0^+ = .5$, $d_0^- = 1$, $d_1^+ = .2$, and $d_1^- = .8$, and the threshold fixed to zero. Table 6 shows the results. When $N = 1$, the tests have no power to detect the threshold effects in the MA part. For $N = 2$, and particularly for $N = 5$, the power is very good. For example, the average LM statistic rejects the null with a probability of .91 at a size corresponding to .05.

In conclusion, for a sample of 200 observations, the functions based on the LM statistics have good size for a number of simulated paths equal to or greater than 2. Moreover, the power for N equal to or greater than 2 is high and increasing in N .

Table 5. Sample Size of the Tests for 200 Observations

Tests	N = 1			N = 2			N = 5		
	.01	.05	.10	.01	.05	.10	.01	.05	.10
SupWald	.00	.00	.00	.06	.12	.20	.09	.17	.20
AveWald	.00	.00	.00	.05	.09	.16	.08	.14	.19
SupLM	.00	.00	.00	.01	.04	.07	.06	.11	.15
AveLM	.00	.00	.00	.01	.07	.10	.05	.08	.13

Table 6. Sample Power of the Tests for 200 Observations

Tests	N = 1			N = 2			N = 5		
	.01	.05	.10	.01	.05	.10	.01	.05	.10
SupWald	.00	.00	.00	.55	.74	.85	.86	.93	.96
AveWald	.00	.00	.00	.53	.75	.85	.86	.93	.95
SupLM	.00	.01	.05	.46	.61	.66	.85	.91	.94
AveLM	.01	.03	.05	.48	.64	.70	.84	.91	.94

5. APPLICATION TO THE PERSISTENCE OF SHOCKS

In this section, we apply the estimation and testing strategies developed in the previous lines to measure the persistence of shocks to the real U.S. GNP growth rate (seasonally adjusted) for the period corresponding to the first quarter of 1947 to the first quarter of 1991. We use this sample to compare our results to the results obtained by Elwood (1998).

To assess the ARMA specification for the GNP, we examine the autocorrelation and partial autocorrelation functions. The first two autocorrelations and the first partial correlation are significant. Thus, an AR(1) or an MA(2) process could possibly fit the series. For the AR(1) process, the residuals are still correlated (the p value for the Breusch–Godfrey serial correlation LM test with two lags is equal to .04). In contrast, the residuals for the MA(2) process are not correlated, whether we look at the Breusch–Godfrey serial correlation LM test (p value equal to .46 with two lags) or the Q statistic (Box–Pierce statistic). Also, the sum of squared residuals is smaller for the MA(2) process compared to that of the AR(1) process. Therefore, the MA(2) model is adopted.

We run a RESET test to detect nonlinearity for the MA(2) process. The null hypothesis of linearity is highly rejected with a p value equal to .000024. This is confirmed by the skewness (−.15) and excess kurtosis (3.45) of the series. Besides, no ARCH effects are detected on residuals. Hence, a threshold moving average model seems to be an interesting alternative to the traditional moving average model.

To implement the estimation of the TMA model, we need an instrumental model. We choose instrumental model III, including six lags for the first-order polynomial and three lags for the second- and third-order polynomials. Under this model of interest and this instrumental model, Theorem 2 can be directly evoked to justify the estimation and testing strategy. To assess the performance of this instrumental model, we have compared the estimation of the MA(2) model by maximum likelihood and by indirect inference. The estimates obtained with both estimation methods were very close. This instrumental model is thus adopted for the estimation of the TMA model. The interval for the threshold is set to $[-1, 1]$ with evaluation points spaced by .01.

Table 7 shows the results for the supremum and the average of the Wald and LM statistics. The MA(2) model is highly

Table 7. Tests for Threshold Moving Average Effect in U.S. GNP

	SupWald	AveWald	SupLM	AveLM
p values	.000	.000	.039	.032

Table 8. Results of the Estimation

	Constant	d_0^+	d_1^+	d_2^+	d_0^-	d_1^-	d_2^-
Estimates	.0025	.0118	.0033	.0058	.0047	-.0004	-.0014
Standard errors	.0029	.0007	.0012	.0025	.0012	.0050	.0048

rejected against the threshold moving average model for all statistics. The minimum of the global specification test [see Gouriéroux, Monfort, and Renault (1993)] is obtained for a threshold equal to a recession value of $-.85$. The minimum of the specification test allows judging the fit of the model. This statistic is equal to 8.56, which has to be compared with the critical value of a chi-square with six degrees of freedom (12.59). The asymmetric moving average model is thus not rejected. The results of the estimation are given in Table 8. We see that the contemporaneous asymmetry is important, and therefore justifies our modeling strategy. A shock with a value greater than $-.85$ has an impact almost three times larger than a shock smaller than $-.85$. We can now examine the persistence of a shock depending on its value. The persistence of shocks above and below the threshold is measured by the sum of the moving average coefficients indexed by $+$ and $-$, respectively. A shock greater than the threshold has a persistence equal to .0209 compared to .0019 for a shock smaller than the threshold. This result corroborates the result of Beaudry and Koop (1993) obtained with an alternative estimation strategy. Obviously, we get different results from Elwood (1998) by considering contemporaneous asymmetry and without restricting the threshold value to be zero.

6. CONCLUSION

In this article, we have analyzed how to adapt the testing procedure in indirect inference when a nuisance parameter is not identified under the null hypothesis. This testing procedure has been illustrated on threshold moving average models to measure the persistence of shocks to output. A more advanced study of various TMA models and a comparison with competing nonlinear models such as TAR and regime-switching models is left to future research. Indeed, this requires further developments that are beyond the scope of this article, namely, the design of testing procedures for nonnested models in a dynamic setting when a nuisance parameter is involved.

ACKNOWLEDGMENTS

We benefited from suggestions of J. Wooldridge, an associate editor, and a referee which greatly helped us to improve the article. We would like to thank Éric Ghysels and Éric Renault for comments on an earlier draft. The first author gratefully acknowledges financial support of Fonds pour la Formation de Chercheurs et l'Aide à la recherche (FCAR). The second author gratefully acknowledges financial support by the grant "Actions de Recherche Concertée" 93/98-162 from the Ministry of Scientific Research (Belgian French Speaking Community).

APPENDIX

Proof of Theorem 1

The expansions of the M estimators $\hat{\beta}_T$ and $\hat{\beta}_T^N(\theta, \gamma)$ follow from standard asymptotic arguments, and are given by

$$\begin{aligned}\sqrt{T}(\hat{\beta}_T - \beta_0) &= -J_{q,T}(\beta_0, \gamma)^{-1}T^{-1/2} \\ &\quad \times \sum_{t=1}^T s_{q,t}(\beta_0, \gamma) + o_p(1), \\ \sqrt{T}(\hat{\beta}_T^N(\theta, \gamma) - b(\theta, \gamma)) &= -J_{q,T}(b(\theta, \gamma), \theta, \gamma)^{-1}T^{-1/2} \\ &\quad \times \sum_{t=1}^T s_{q,t}^n(b(\theta, \gamma), \theta, \gamma) + o_p(1)\end{aligned}$$

where $J_{q,T}(\beta_0, \gamma) = T^{-1} \sum_{t=1}^T \nabla_{\beta'} s_{q,t}(\beta_0, \gamma)$ and $J_{q,T}^n(b(\theta, \gamma), \theta, \gamma) = \sum_{t=1}^T \nabla_{\beta'} s_{q,t}^n(b(\theta, \gamma), \theta, \gamma)$.

From the first-order condition of the indirect estimator criterion and the above expansions [see Gouriéroux, Monfort, and Renault (1993)], we may deduce the asymptotic behavior of the indirect estimator. Indeed, as $\beta_0 = b(\theta_0, \gamma)$ under the null for all $\gamma \in \Gamma$, we have

$$\begin{aligned}\nabla_{\theta} \hat{\beta}_T^N(\theta_0, \gamma)' \hat{\Omega}_T^*(\gamma) ((\hat{\beta}_T - \beta_0) - (\hat{\beta}_T^N(\theta_0, \gamma) - b(\theta_0, \gamma))) \\ - \nabla_{\theta} \hat{\beta}_T^N(\theta_0, \gamma)' \hat{\Omega}_T^*(\gamma) \nabla_{\theta'} \hat{\beta}_T^N(\theta_0, \gamma) (\hat{\theta}_T^N(\gamma) - \theta_0) = o_p(T^{-1/2})\end{aligned}$$

and obtain

$$\begin{aligned}\sqrt{T}(\hat{\theta}_T^N(\gamma) - \theta_0) \\ = -[\nabla_{\theta} \hat{\beta}_T^N(\theta_0, \gamma)' \hat{\Omega}_T^*(\gamma) \nabla_{\theta'} \hat{\beta}_T^N(\theta_0, \gamma)]^{-1} \\ \times \nabla_{\theta} \hat{\beta}_T^N(\theta_0, \gamma)' \Omega_T^*(\gamma) J_{q,T}(\beta_0, \gamma)^{-1} T^{-1/2} \sum_{t=1}^T s_{q,t}(\beta_0, \gamma) \\ + [\nabla_{\theta} \hat{\beta}_T^N(\theta_0, \gamma)' \hat{\Omega}_T^*(\gamma) \nabla_{\theta'} \hat{\beta}_T^N(\theta_0, \gamma)]^{-1} \nabla_{\theta} \hat{\beta}_T^N(\theta_0, \gamma)' \Omega_T^*(\gamma) \\ \times N^{-1} \sum_{n=1}^N \left(J_{q,T}^n(\beta_0, \theta, \gamma)^{-1} T^{-1/2} \sum_{t=1}^T s_{q,t}^n(\beta_0, \theta, \gamma) \right) + o_p(1).\end{aligned}$$

Hence, by Assumption 1 and the weak convergence of the score functions,

$$\begin{aligned}\sqrt{T}(\hat{\theta}_T^N(\cdot) - \theta_0) \implies -[\nabla_{\theta} b(\theta_0, \cdot)' \Omega^*(\beta_0, \cdot) \nabla_{\theta'} b(\theta_0, \cdot)]^{-1} \\ \times \nabla_{\theta} b(\theta_0, \cdot)' \Omega^*(\beta_0, \cdot) J_q(\beta_0, \cdot)^{-1} \\ \times \left(S_q(\beta_0, \cdot) - N^{-1} \sum_{n=1}^N S_q^n(\beta_0, \cdot) \right)\end{aligned}$$

as a process indexed by $\gamma \in \Gamma$. The asymptotic distribution of $W_T(\cdot)$ can be easily deduced from the asymptotic distribution of $\hat{\theta}_T^N(\cdot)$. The asymptotic equivalence among $W_T(\cdot)$, $LM_T(\cdot)$, and $LR_T(\cdot)$ can be shown through modifications of the proofs in Appendix 4 of Gouriéroux, Monfort, and Renault (1993) to account for the presence of the nuisance parameter γ .

Proof of Proposition 1

The moments of the truncated standard normal variable ϵ_t^+ can be calculated with the following recursive formulas:

$$\begin{aligned} E[(\epsilon_t^+)] &= \varphi(\gamma) \\ E[(\epsilon_t^+)^2] &= \gamma\varphi(\gamma) + \Phi(-\gamma) \\ E[(\epsilon_t^+)^i] &= \int_{\gamma}^{\infty} \epsilon^i \varphi(\epsilon) d\epsilon \\ &= - \int_{\gamma}^{\infty} \epsilon^{(i-1)} d\varphi(\epsilon) \\ &= [-\epsilon^{i-1} \varphi(\epsilon)]_{\gamma}^{\infty} + (i-1) \int_{\gamma}^{\infty} \epsilon^{i-2} \varphi(\epsilon) d\epsilon \\ &= \gamma^{i-1} \varphi(\gamma) + (i-1) E[(\epsilon_t^+)^{i-2}], \quad i \geq 3 \end{aligned}$$

where φ and Φ denote the pdf and cdf of a standard Gaussian variable, respectively. Similarly, we may establish that

$$\begin{aligned} E[(\epsilon_t^-)] &= -\varphi(\gamma) \\ E[(\epsilon_t^-)^2] &= -\gamma\varphi(\gamma) + \Phi(\gamma) \\ E[(\epsilon_t^-)^i] &= -\gamma^{i-1} \varphi(\gamma) + (i-1) E[(\epsilon_t^-)^{i-2}], \quad i \geq 3. \end{aligned}$$

The moments of the centered processes $\eta_t^+ = \epsilon_t^+ - E[\epsilon_t^+]$ and $\eta_t^- = \epsilon_t^- - E[\epsilon_t^-]$ are thus equal to

$$\begin{aligned} E[(\eta_t^+)^2] &= \gamma\varphi(\gamma) - \varphi(\gamma)^2 + \Phi(-\gamma) \\ E[(\eta_t^+)^3] &= (\gamma^2 + 2)\varphi(\gamma) - 3\gamma\varphi(\gamma)^2 \\ &\quad - 3\varphi(\gamma)\Phi(-\gamma) + 2\varphi(\gamma)^3 \\ E[(\eta_t^+)^4] &= (\gamma^3 + 3\gamma)\varphi(\gamma) - [4(\gamma^2 + 2) - 6\Phi(-\gamma)]\varphi(\gamma)^2 \\ &\quad + 6\gamma\varphi(\gamma)^3 - 3\varphi(\gamma)^4 + 3\Phi(-\gamma) \end{aligned}$$

and

$$\begin{aligned} E[(\eta_t^-)^2] &= -\gamma\varphi(\gamma) - \varphi(\gamma)^2 + \Phi(\gamma) \\ E[(\eta_t^-)^3] &= -(\gamma^2 + 2)\varphi(\gamma) - 3\gamma\varphi(\gamma)^2 \\ &\quad + 3\varphi(\gamma)\Phi(\gamma) - 2\varphi(\gamma)^3 \\ E[(\eta_t^-)^4] &= -(\gamma^3 + 3\gamma)\varphi(\gamma) - [4(\gamma^2 + 2) - 6\Phi(\gamma)]\varphi(\gamma)^2 \\ &\quad - 6\gamma\varphi(\gamma)^3 - 3\varphi(\gamma)^4 + 3\Phi(\gamma). \end{aligned}$$

Expectations of products of powers of η_t^+ and η_t^- can similarly be computed:

$$\begin{aligned} E[\eta_t^+ \eta_t^-] &= \varphi(\gamma)^2 \\ E[(\eta_t^+)^2 \eta_t^-] &= -2\varphi(\gamma)^3 + \gamma\varphi(\gamma)^2 + \varphi(\gamma)\Phi(-\gamma) \\ E[\eta_t^+ (\eta_t^-)^2] &= 2\varphi(\gamma)^3 + \gamma\varphi(\gamma)^2 - \varphi(\gamma)\Phi(\gamma) \\ E[(\eta_t^+)^2 (\eta_t^-)^2] &= \varphi(\gamma)^2 (\Phi(\gamma) + \Phi(-\gamma)) - 3\varphi(\gamma)^2 \\ E[(\eta_t^+)^3 \eta_t^-] &= 3\varphi(\gamma)^4 + \varphi(\gamma)^2 \\ &\quad \times (\gamma^2 + 2 - 3(\gamma\varphi(\gamma) + \Phi(-\gamma))) \end{aligned}$$

and

$$E[\eta_t^+ (\eta_t^-)^3] = 3\varphi(\gamma)^4 + \varphi(\gamma)^2 (\gamma^2 + 2 + 3(\gamma\varphi(\gamma) - \Phi(\gamma))).$$

The conditional skewness and kurtosis of Y_t can be computed for any threshold γ by using the moments of $U_t =$

$Y_t - E[Y_t | \epsilon_{t-1}]$ and the aforementioned expressions. The conditional skewness of Y_t is equal to

$$\frac{E[U_t^3 | \epsilon_{t-1}]}{E[U_t^2 | \epsilon_{t-1}]^{3/2}} = \frac{\sum_{i=0}^3 \binom{3}{i} (f_0^+)^i (f_0^-)^{3-i} c(i, 3-i)}{[\sum_{i=0}^2 \binom{2}{i} (f_0^+)^i (f_0^-)^{2-i} c(i, 2-i)]^{3/2}}$$

where $c(i, j) = E[(\eta_t^+)^i (\eta_t^-)^j]$, $\forall i, j$. The conditional kurtosis of Y_t is given by

$$\frac{E[U_t^4 | \epsilon_{t-1}]}{E[U_t^2 | \epsilon_{t-1}]^2} = \frac{\sum_{i=0}^4 \binom{4}{i} (f_0^+)^i (f_0^-)^{4-i} c(i, 4-i)}{[\sum_{i=0}^2 \binom{2}{i} (f_0^+)^i (f_0^-)^{2-i} c(i, 2-i)]^2}.$$

When we set $\gamma = 0$, we get from $\varphi(0) = 1/\sqrt{2\pi}$, $\Phi(0) = 1/2$,

$$\begin{aligned} E[(\eta_t^+)^2] &= E[(\eta_t^-)^2] = \frac{\pi - 1}{2\pi} \\ E[(\eta_t^+)^3] &= -E[(\eta_t^-)^3] = \frac{\pi + 2}{(2\pi)^{3/2}} \\ E[(\eta_t^+)^4] &= E[(\eta_t^-)^4] = \frac{6\pi^2 - 10\pi - 3}{4\pi^2} \\ E[\eta_t^+ \eta_t^-] &= \frac{1}{2\pi} \\ E[(\eta_t^+)^2 \eta_t^-] &= -E[\eta_t^+ (\eta_t^-)^2] = \frac{\pi - 2}{(2\pi)^{3/2}} \\ E[(\eta_t^+)^2 (\eta_t^-)^2] &= \frac{2\pi - 3}{4\pi^2} \\ E[(\eta_t^+)^3 \eta_t^-] &= -E[\eta_t^+ (\eta_t^-)^3] = \frac{\pi + 3}{4\pi^2} \end{aligned}$$

which leads to the results of Proposition 1.

Proof of Proposition 2

From the definitions of the autocovariance $\tilde{\gamma}(u)$ and the centered processes η_t^+ , η_t^- , we get

$$\tilde{\gamma}(u) = E \left[\left(\sum_{i=0}^l d_i^+ \eta_{t-i}^+ + d_i^- \eta_{t-i}^- \right) \left(\sum_{i=0}^l d_i^+ \eta_{t-u-i}^+ + d_i^- \eta_{t-u-i}^- \right) \right].$$

Therefrom, we deduce, thanks to the iid properties of η_t^+ , η_t^- , that $\tilde{\gamma}(u) = 0$ for $u > l$, and

$$\begin{aligned} \tilde{\gamma}(u) &= \sum_{i=0}^{l-u} [d_{i+u}^+ d_i^+ E[(\eta_{t-i}^+)^2] + d_{i+u}^- d_i^- E[(\eta_{t-i}^-)^2] \\ &\quad + (d_{i+u}^+ d_i^- + d_{i+u}^- d_i^+) E[\eta_{t-i}^+ \eta_{t-i}^-]] \end{aligned}$$

for $u \leq l$. This leads to the stated result since $E[(\eta_{t-i}^+)^2] = E[(\eta_{t-i}^-)^2] = (\pi - 1)/(2\pi)$, and $E[\eta_{t-i}^+ \eta_{t-i}^-] = -E[\epsilon_{t-i}^+] E[\epsilon_{t-i}^-] = 1/(2\pi)$ when $\gamma = 0$.

Proof of Theorem 2

We focus on Assumptions (c), (f), (k), and (n). Other assumptions are straightforward to check.

First, we show that $S_{q,T} = T^{-1/2} \sum_{t=1}^T s_{q,t} \Rightarrow S_q$ [assumption (n)]. By Theorem 21.9 of Davidson (1994), uniform convergence over $\gamma \in \Gamma$ is obtained if and only if pointwise convergence holds and $S_{q,T}$ is stochastically equicontinuous. For each $\gamma \in \Gamma$, the score of instrumental models I–III takes the generic form

$$\sum_{t=1}^T s_{q,t}(\gamma) = \sum_{t=1}^T x_t(\gamma) u_t(\gamma)$$

where $u_t(\gamma)$ is the error term, and the column vector $x_t(\gamma)$ contains lags of observed values or polynomials of these lags for the regression at time t . The pointwise multivariate central limit theorem then applies immediately to the score since $Y_t(\gamma)$ is a q -dependent process for each γ . By Andrews (1992), the existence of a Lipschitz-type condition is sufficient to obtain stochastic equicontinuity [see also thm. 21.10 in Davidson (1994)]. We can rely here on the lines in Hansen (1996). The score is mixing at a rate infinitely fast since $Y_t(\gamma)$ is q -dependent. By establishing that the score fulfills Assumption 2 in Hansen, the uniform convergence will follow. So, we have to show

$$\|x_t(\gamma)u_t(\gamma) - x_t(\gamma^*)u_t(\gamma^*)\|_2 \leq B_t|\gamma - \gamma^*|$$

where B_t is $O_p(1)$, not depending on γ for all $\gamma, \gamma^* \in \Gamma$.

By the triangular inequality,

$$\begin{aligned} \|x_t(\gamma)u_t(\gamma) - x_t(\gamma^*)u_t(\gamma^*)\|_2 \\ \leq \|x_t(\gamma)u_t(\gamma)\|_2 + \|x_t(\gamma^*)u_t(\gamma^*)\|_2, \end{aligned}$$

and by Hölder's inequality,

$$\begin{aligned} \|x_t(\gamma)u_t(\gamma)\|_2 + \|x_t(\gamma^*)u_t(\gamma^*)\|_2 \\ \leq \|x_t(\gamma)\|_4 \|u_t(\gamma)\|_4 + \|x_t(\gamma^*)\|_4 \|u_t(\gamma^*)\|_4. \end{aligned}$$

$\|x_t(\gamma)\|_4$ is bounded for all $\gamma \in \Gamma$. Hence, $\|u_t(\gamma)\|_4 \leq \|x_t(\gamma)\|_4 \leq M_t$, and

$$\|x_t(\gamma)u_t(\gamma) - x_t(\gamma^*)u_t(\gamma^*)\|_2 \leq 2M_t^2.$$

Since the support of Γ is a finite subset of the real line, we can always find a $B_t < \infty$ so that $2M_t^2 \leq B_t|\gamma - \gamma^*|$ for all $\gamma, \gamma^* \in \Gamma$ (see Hansen (1996), p. 426). This is the desired result.

Now, we are interested in assumption (k). $\frac{1}{T} \sum_{t=1}^T \nabla_{\beta'} s_{q,t}(\gamma)$ is equal to $\frac{1}{T} \sum_{t=1}^T x_t(\gamma)x_t(\gamma)'$. This expression satisfies the assumptions of Lemma 1 of Hansen. Uniform almost sure convergence is a consequence of this lemma and the continuity of the support of γ .

The uniform convergence of $\hat{\beta}_T^N(\theta, \gamma)$ [assumption (c)] follows here from assumptions (k) and (n).

Next, we have to show the uniform convergence of $\nabla_{\theta} \hat{\beta}_T^N(\theta, \gamma)$ for all $\gamma \in \Gamma$. We have the following equality:

$$\begin{aligned} \nabla_{\theta} \hat{\beta}_T^N(\theta, \gamma) &= [X'(\theta, \gamma)X(\theta, \gamma)]^{-1} \\ &\times [\nabla_{\theta}(X(\theta, \gamma)'Y(\theta, \gamma)) - [X'(\theta, \gamma)X(\theta, \gamma)]^{-1} \\ &\times (\nabla_{\theta}(X(\theta, \gamma)'X(\theta, \gamma)))(X(\theta, \gamma)'Y(\theta, \gamma))] \end{aligned}$$

where $\hat{\beta}_T^N(\theta, \gamma) = [X'(\theta, \gamma)X(\theta, \gamma)]^{-1}(X(\theta, \gamma)'Y(\theta, \gamma))$ with obvious notation. Again, by applying Lemma 1 of Hansen to $\frac{1}{T} \nabla_{\theta}(X(\theta, \gamma)'X(\theta, \gamma))$, $\frac{1}{T} \nabla_{\theta}(X(\theta, \gamma)'Y(\theta, \gamma))$, $\frac{1}{T}(X'(\theta, \gamma)X(\theta, \gamma))$, and $\frac{1}{T}(X(\theta, \gamma)'Y(\theta, \gamma))$, we get uniform almost sure convergence for these expressions. Hence, $\nabla_{\theta} \hat{\beta}_T^N(\theta, \gamma)$ converges uniformly almost surely.

Finally, we derive the asymptotic distribution of the indirect inference estimator under the alternative. By a mean value expansion of the first-order condition of the indirect estimator criterion under the alternative

$$\begin{aligned} \nabla_{\theta} \hat{\beta}_T^N(\hat{\theta}, \gamma)' \hat{\Omega}_T^*(\gamma) (\hat{\beta}_T - (\hat{\beta}_T^N(\hat{\theta}, \gamma))) \\ = \nabla_{\theta} \hat{\beta}_T^N(\hat{\theta}, \gamma)' \hat{\Omega}_T^*(\gamma) (\hat{\beta}_T - (\hat{\beta}_T^N(\theta^*, \gamma))) \\ - \nabla_{\theta} \hat{\beta}_T^N(\hat{\theta}, \gamma)' \hat{\Omega}_T^*(\gamma) \nabla_{\theta'} \hat{\beta}_T^N(\tilde{\theta}, \gamma) \\ \times (\hat{\theta}_T^N(\gamma) - \theta^*) + o_p(T^{-1/2}) \end{aligned}$$

where $\tilde{\theta}' = [\tilde{\theta}^{(1)} \dots \tilde{\theta}^{(q)}]$ and $\tilde{\theta}^{(k)} = \lambda^{(k)} \theta_T^{*(k)} + (1 - \lambda^{(k)}) \hat{\theta}^{(k)}$ for some $0 \leq \lambda^{(k)} \leq 1$ and $k = 1, \dots, q$ and θ^* is the indirect pseudotrue value under the alternative. The expression above yields

$$\begin{aligned} (\hat{\theta}_T^N(\gamma) - \theta^*) &= [\nabla_{\theta} \hat{\beta}_T^N(\hat{\theta}, \gamma)' \hat{\Omega}_T^*(\gamma) \nabla_{\theta'} \hat{\beta}_T^N(\tilde{\theta}, \gamma)]^{-1} \\ &\times \nabla_{\theta} \hat{\beta}_T^N(\hat{\theta}, \gamma)' \hat{\Omega}_T^*(\gamma) \\ &\times (\hat{\beta}_T - (\hat{\beta}_T^N(\theta^*, \gamma))) + o_p(T^{-1/2}). \quad (\text{A.1}) \end{aligned}$$

Under the alternative with the appropriate selection matrix R defined in Section 3, we have

$$R'\theta^* = R'\theta_0 + \frac{c}{\sqrt{T}}.$$

By multiplying Equation (A.1) by the selection matrix R , and by using assumptions (a)–(p), as well as results provided in the proof of Theorem 1, we obtained the desired result:

$$\sqrt{T}R'(\hat{\theta}_T^N(\cdot) - \theta_0) \Rightarrow (-R'\bar{K}_q(\beta_0, \theta_0, \cdot)\bar{S}_q^N(\beta_0, \cdot)) + c$$

with

$$\bar{K}_q(\cdot) = [\nabla_{\theta} b(\theta_0, \cdot)' \Omega^*(\beta_0, \cdot) \nabla_{\theta'} b(\theta_0, \cdot)]^{-1} \nabla_{\theta} b(\theta_0, \cdot)' \Omega^*(\beta_0, \cdot).$$

[Received September 1999. Revised August 2001.]

REFERENCES

- Andrews, D. (1992), "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.
- (1993), "Tests for Parameter Instability and Structural Change With Unknown Change Point," *Econometrica*, 61, 821–856.
- Andrews, D., and Ploberger, W. (1994), "Optimal Tests When a Nuisance Parameter is Only Present Under the Alternative," *Econometrica*, 62, 1383–1414.
- Beaudry, P., and Koop, G. (1993), "Do Recessions Permanently Change Output?," *Journal of Monetary Economics*, 31, 149–163.
- Chan, K. (1990), "Testing for Threshold Autoregression," *Annals of Statistics*, 18, 1886–1894.
- Davidson, J. (1994), *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford, England: Oxford University Press.
- Davis, R. (1977), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247–254.

- (1987), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 74, 33–43.
- De Gooijer, J. (1998), "On Threshold Moving-Average Models," *Journal of Time Series Analysis*, 19, 1–18.
- Dhaene, G., Gouriéroux, C., and Scaillet, O. (1998), "Indirect Encompassing and Instrumental Models," *Econometrica*, 66, 673–688.
- El Babsiri, M., and Zakoïan, J. M. (2001), "Contemporaneous Asymmetry in GARCH Processes," *Journal of Econometrics*, 101, 257–294.
- Elwood, S. (1998), "Is the Persistence of Shocks to Output Asymmetric?," *Journal of Monetary Economics*, 41, 411–426.
- Gallant, R., and Tauchen, G. (1996), "Which Moments to Match?," *Econometric Theory*, 12, 657–681.
- Ghysels, E., Guay, A., and Hall, A. (1998), "Predictive Tests for Structural Change With Unknown Breakpoint," *Journal of Econometrics*, 82, 209–233.
- Ghysels, E., and Guay, A. (2001a), "Structural Change Tests for Simulated Method of Moments," mimeo.
- (2001b), "Indirect Inference, Efficient Method of Moments and Structural Change," mimeo.
- Gouriéroux, C., and Monfort, A. (1995a), *Simulation Based Econometric Methods*, CORE Lecture Series and Oxford University Press.
- (1995b), "Testing, Encompassing and Simulating Dynamic Econometric Models," *Econometric Theory*, 11, 195–228.
- Gouriéroux, C., Monfort, A., and Renault, E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.
- Hansen, B. (1996), "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis," *Econometrica*, 64, 413–430.
- Michaelides, A., and Ng, S. (2000), "Estimating the Rational Expectations Model of Speculative Storage: A Monte Carlo Comparison of Three Simulation Estimators," *Journal of Econometrics*, 96, 231–266.
- Pollard, D. (1994), *Convergence of Stochastic Processes*, New York: Springer-Verlag.
- Smith, A. (1993), "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, 63–84.
- Sowell, F. (1996), "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework," *Econometrica*, 64, 1085–1107.
- Wecker, W. (1981), "Asymmetric Time Series," *Journal of the American Statistical Association*, 76, 16–21.
- Zakoïan, J.-M. (1994), "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control*, 18, 931–955.