Econometric Theory, **26**, 2010, 965–993.
doi:10.1017/S0266466609990430

SUP-TESTS FOR LINEARITY IN A GENERAL NONLINEAR AR(1) MODEL

CHRISTIAN FRANCQ
Université Lille III, GREMARS-EQUIPPE

LAJOS HORVATH
University of Utah

JEAN-MICHEL ZAKOÏAN
Université Lille III, GREMARS-EQUIPPE
and
CREST

We consider linearity testing in a general class of nonlinear time series models of order one, involving a nonnegative nuisance parameter that (a) is not identified under the null hypothesis and (b) gives the linear model when equal to zero. This paper studies the asymptotic distribution of the likelihood ratio test and asymptotically equivalent supremum tests. The asymptotic distribution is described as a functional of chi-square processes and is obtained without imposing a positive lower bound for the nuisance parameter. The finite-sample properties of the sup-tests are studied by simulations.

1. INTRODUCTION

Building nonlinear time series models is, in general, a difficult task that requires a large amount of care. As can be seen from recent studies comparing the forecast accuracy of linear autoregressive (AR) models and nonlinear models on real macroeconomic time series, a careful specification of the nonlinear models is required to produce forecasts that improve upon linear forecasts (see Stock and Watson, 1999; Teräsvirta, van Dijk, and Medeiros, 2004).

In general, nonlinear models (such as the threshold autoregressive [TAR], and the smooth transition autoregressive [STAR] regime-switching or bilinear models) contain the linear one as a particular case, but often some of the parameters are not identified when linearity holds. This is, e.g., the case of the threshold value in the TAR framework. This identifiability problem results in parameter inconsistency, and, if the series under consideration is close to being linear, the nonlinear model is bound to produce forecasts that are unreliable compared to linear ones. It is therefore essential to test first for linearity before fitting any particular nonlinear model.

This research was partially supported by NSF grant DMS 0604670 and grant RGC-HKUST 6428/OGH. The authors gratefully acknowledge the suggestions of the referees and the associate editor. Address correspondence to Jean-Michel Zakoïan, CREST, 15 boulevard Gabriel Péri, 92245 Malakoff Cedex, France; e-mail: zakoian@ensae.fr.

© Cambridge University Press 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

The aim of this paper is to consider linearity testing in a relatively general, first-order nonlinear framework. Given the unlimited number of nonlinear models it is not possible to nest all of them in a general class. Many of them, however, can be seen as particular cases of a nonlinear AR(1) model of the form

$$Y_t = \mu_0 + \{a_0 + b_0 H(\gamma_0, Y_{t-1})\} Y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{IID}(0, \sigma^2) \tag{1.1}$$

for some function H defined on $\Gamma \times \mathbb{R}$, for some set $\Gamma \subset \mathbb{R}$ containing zero, and such that $H(0, \cdot) = 0$. Clearly, the specification of the function H may include more than one parameter, but we only need to underline the parameter γ_0 controlling the nullity of the function H. Examples and precise assumptions will be given in the next sections. We are interested in testing the linearity hypothesis $b_0 = 0$. Problems of this nature, where a nuisance parameter γ_0 is present only under the alternative hypothesis, often occur in econometric models and have been considered by many authors. See, among others, Davies (1977, 1987), King and Shively (1993), Andrews and Ploberger (1995), Hansen (1996), and Stinchcombe and White (1998).

The contribution of this paper is to derive the asymptotic distribution of supremum tests, namely, the likelihood ratio (LR) test, and asymptotically equivalent sup-Wald and Lagrange multiplier (LM) tests, without bounding the nuisance parameter away from zero. The difficulty is that when γ_0 approaches zero the nonlinear term vanishes in (1.1) and the Fisher information matrix becomes singular. In the literature, this problem is typically circumvented by imposing a lower bound for the nuisance parameter. We avoid this restriction. To our knowledge, this is the first paper deriving the asymptotic distribution of a supremum test with a nuisance-parameter range implying a case of noninvertible information matrix.

The paper is organized in the following way: Section 2 discusses the model and gives stationarity conditions. Section 3 derives the asymptotic properties of the least squares estimator (LSE) of (a_0, b_0) under the null assumption of linearity, i.e., $b_0 = 0$. Section 4 defines the LR-, Wald-, and LM-type tests that are based on the LSE. The asymptotic null distribution is derived. Section 4 also presents a Monte Carlo study, in which the supremum tests enjoy good size and power properties. This study compares the powers of the sup-tests and of tests based on expansions of the function $H(\cdot, y)$, which are often used in practice. The Appendix provides proofs of the results given in the paper.

2. EXAMPLES AND STATIONARITY CONDITIONS

Before turning to the framework of this paper, leaving the function H unspecified, it is of interest to present special cases of (1.1) that have been popular in forecasting applications. See Tong (1990) and Teräsvirta et al. (2004) for a more complete discussion.

One example is the exponential autoregressive (EXPAR) model introduced by Haggan and Ozaki (1981), which, after reparameterization, is obtained for

$$H(\gamma_0, y) = 1 - e^{-\gamma_0 y^2}. \tag{2.1}$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

The parameter γ_0 is often referred to as the slope parameter. Model (1.1) includes other smooth transition models, such as the logistic smooth transition autoregressive (LSTAR) model, introduced in the time series literature by Luukkonen, Saikkonen, and Teräsvirta (1988). In this latter model, we have $H(\gamma_0, y) = H(\gamma_0, c, y) = (1 + e^{-\gamma_0(y-c)})^{-1} - \frac{1}{2}$ where c is a location coefficient allowing for asymmetries in the conditional mean of Y_t. When $c = 0$ the model is simply

$$H(\gamma_0, y) = \frac{1}{1 + e^{-\gamma_0 y}} - \frac{1}{2}. \tag{2.2}$$

Letting the slope parameter $\gamma_0 \to \infty$, we obtain the two-regime self-exciting threshold autoregressive (SETAR) model of Tong and Lim (1980). The SETAR model will not be covered by the results of this paper, however, because smoothness assumptions on the function H will be required.

The existence of strictly stationary solutions to (1.1) can be investigated using Markov chains theory. The following result is an immediate consequence of Tjøstheim (1990, Thm. 4.1 and Lem. 6.1).

THEOREM 2.1. *Suppose that ϵ_t has a positive density function over the real line. Then, if there exists $r, K > 0$ such that*

$$\sup_y |a_0 + b_0 H(\gamma_0, y)| < K, \qquad \sup_{|y|>r} |a_0 + b_0 H(\gamma_0, y)| < 1$$

there exists a strictly stationary and geometrically ergodic solution to model (1.1). Moreover, for any $k > 1$, if $\mathrm{E}|\epsilon_t|^k < \infty$ then $\mathrm{E}|Y_t|^k < \infty$.

For example the EXPAR model admits a strictly stationary solution whenever $|a_0 + b_0| < 1$ and $\gamma_0 \geq 0$. For other models, such as the LSTAR, $\gamma_0 > 0$ is not required for stationarity but is a natural constraint for interpretation and identifiability (see, e.g., Teräsvirta et al., 2004). For this reason we will take throughout a compact nuisance-parameter space of the form $\Gamma = [0, \overline{\gamma}]$. Now we turn to the least squares estimation.

3. ASYMPTOTIC PROPERTIES OF THE LSE OF μ_0, a_0, AND b_0 UNDER THE LINEAR MODEL

Let $Y_1, \ldots, Y_n$ be observations of a nonanticipative strictly stationary solution of (1.1). Recall that the function H is known a priori. Throughout we make the following assumption so that the standard AR(1) model is obtained for $\gamma_0 = 0$ but also for $b_0 = 0$.

A0. $H(0, \cdot) = 0$ and $H(\gamma, \cdot)$ is not identically zero, for any $\gamma > 0$.

Thus it is not restrictive to assume $\gamma_0 > 0$ and interpret γ_0 as a nuisance parameter, which is not present when $b_0 = 0$. Notice also that b_0 cannot be identified

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

when $\gamma_0 = 0$. For a given value γ of γ_0, the LSE of $\theta_0 = (\mu_0, a_0, b_0)'$ coincides with the Gaussian quasi-maximum likelihood estimator and is defined as any measurable solution of

$$\hat{\theta} := (\hat{\mu}_\gamma, \hat{a}_\gamma, \hat{b}_\gamma) = \arg\max_{\theta\in\Theta} L_n(\theta) = \arg\min_{\theta\in\Theta} Q_n(\theta),$$

where

$$L_n(\theta) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{n}{2\sigma^2}Q_n(\theta), \qquad Q_n(\theta) = n^{-1}\sum_{t=2}^{n}\epsilon_t^2(\theta)$$

with

$$\epsilon_t(\theta) = Y_t - \mu - \{a + bH(\gamma, Y_{t-1})\}Y_{t-1}.$$

Assuming $\gamma > 0$, the LSE of $(\mu_0, a_0, b_0)'$ is explicitly given, when $J_n(\gamma)$ is non-singular, by

$$\begin{pmatrix} \hat{\mu}_\gamma \\ \hat{a}_\gamma \\ \hat{b}_\gamma \end{pmatrix} := \begin{pmatrix} \hat{\delta}_\gamma \\ \hat{b}_\gamma \end{pmatrix} = J_n^{-1}(\gamma) \begin{pmatrix} n^{-1}\sum_{t=2}^{n} Y_t \\ n^{-1}\sum_{t=2}^{n} Y_t Y_{t-1} \\ n^{-1}\sum_{t=2}^{n} Y_t Y_{t-1} H(\gamma, Y_{t-1}) \end{pmatrix}, \tag{3.1}$$

where

$$J_n(\gamma) = \begin{pmatrix} 1 & U_{n,1,0} & U_{n,1,1}(\gamma) \\ U_{n,1,0} & U_{n,2,0} & U_{n,2,1}(\gamma) \\ U_{n,1,1}(\gamma) & U_{n,2,1}(\gamma) & U_{n,2,2}(\gamma) \end{pmatrix},$$

$$U_{n,i,j}(\gamma) = n^{-1}\sum_{t=2}^{n} Y_{t-1}^i H^j(\gamma, Y_{t-1}) \tag{3.2}$$

for $i = 1, 2$ and $j = 0, 1, 2$, with the convention $0^0 = 1$. The matrix $J_n(\gamma)$ is singular for $\gamma = 0$. But it can be shown that, under appropriate moment assumptions and Assumption A4, which appears later in this section, when $\gamma > 0$ the matrix $J_n(\gamma)$ is almost surely (a.s.) invertible, at least for large n. See Chesher (1984), Lee and Chesher (1986), and Rotnitzky, Cox, Bottai, and Robins, (2000) for cases where the information matrix is singular for any value of the nuisance parameter.

Under the constraint $b_0 = 0$ the restricted LSE for (μ_0, a_0) is simply

$$\begin{pmatrix} \tilde{\mu} \\ \tilde{a} \end{pmatrix} := \tilde{\delta} = \tilde{J}_n^{-1} \begin{pmatrix} n^{-1}\sum_{t=2}^{n} Y_t \\ n^{-1}\sum_{t=2}^{n} Y_t Y_{t-1} \end{pmatrix}, \qquad \tilde{J}_n = \begin{pmatrix} 1 & U_{n,1,0} \\ U_{n,1,0} & U_{n,2,0} \end{pmatrix}. \tag{3.3}$$

We will now derive asymptotic properties of the LSE under the linear model. We assume that H admits second-order partial derivatives with respect to γ, and we make the following assumptions on the first and second partial derivatives $H_1(\gamma, y) = \partial H(\gamma, y)/\partial\gamma$ and $H_2(\gamma, y) = \partial^2 H(\gamma, y)/\partial\gamma^2$, where $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, $\alpha_3 \geq 0$ and K are constants.

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

A1. $|H_1(\gamma, y)| \le K(|y|^{\alpha_1} + 1)$.

A2. $|H_2(\gamma, y)| \le K(|y|^{\alpha_2} + 1)$.

A3. $|H_2(\gamma, y) - H_2(\gamma', y)| \le K|\gamma - \gamma'|^{\alpha}(|y|^{\alpha_3} + 1)$, with some $1/2 < \alpha \le 1$.

In what follows we use the notation K as a generic constant whose value can change. Conditions A1 and A2 are needed for the existence of the limit process in Theorem 3.1. The proofs are based on Taylor expansions of $H(\cdot, y)$, and A3 is used to control the remainder terms.

Elementary calculations show that Assumptions A1–A3 hold for the EXPAR model with $\alpha_1 = 2$, $\alpha_2 = 4$, and $\alpha_3 = 6$ and $\alpha = 1$. Also, the LSTAR model satisfies A1–A3 with $\alpha_1 = 1$, $\alpha_2 = 2$, and $\alpha_3 = 3$ and $\alpha = 1$. Similarly, for any constant c and any $\beta > 0$, the generalized EXPAR

$$H(\gamma, y) = 1 - e^{-\gamma|y-c|^{\beta}}$$

and the generalized LSTAR

$$H(\gamma, y) = \frac{1}{1 + e^{-\gamma|y-c|^{\beta}\operatorname{sign}(y)}} - \frac{1}{2}, \quad \text{where } \operatorname{sign}(y) = \begin{cases} 1 & y > 0 \\ 0 & y = 0, \\ -1 & y < 0 \end{cases}$$

satisfy A1–A3. Another example is the normal STAR model of Chan and Tong (1986), defined by $H(\gamma, y) = \Phi\{\gamma(y - c)\} - 1/2$, where $\Phi(\cdot)$ is the $\mathcal{N}(0, 1)$ cumulative distribution function. In all these models, other nuisance parameters c and/or β may be present but vanish under the linearity hypothesis $b_0 = 0$. The results obtained in what follows will hold for any fixed values of c and β. Our method can be extended to the case when the other nuisance parameters c and β are also estimated from the sample. For the sake of simplicity, we assume that the only unknown parameter in $H(\gamma, y)$ is γ.

We have

$$\begin{pmatrix} \hat{\mu}_\gamma - \mu_0 \\ \hat{a}_\gamma - a_0 \\ \hat{b}_\gamma - b_0 \end{pmatrix} = J_n^{-1}(\gamma) \begin{pmatrix} S_{n,0,0} \\ S_{n,1,0} \\ S_{n,1,1}(\gamma) \end{pmatrix}, \qquad \begin{pmatrix} \tilde{\mu} - \mu_0 \\ \tilde{a} - a_0 \end{pmatrix} = \tilde{J}_n^{-1} \begin{pmatrix} S_{n,0,0} \\ S_{n,1,0} \end{pmatrix}, \quad \textbf{(3.4)}$$

where

$$S_{n,i,j}(\gamma) = n^{-1} \sum_{t=2}^{n} \epsilon_t Y_{t-1}^i H^j(\gamma, Y_{t-1}).$$

We will also need to consider the sums

$$T_{n,i}(\gamma) = n^{-1} \sum_{t=2}^{n} \epsilon_t Y_{t-1} H_i(\gamma, Y_{t-1}), \qquad i = 1, 2.$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

Our first result establishes the weak convergence of the processes $\{S_{n,i,j}(\gamma), T_{n,i}(\gamma), \gamma > 0\}$. For any $\overline{\gamma} > 0$, the symbol $\overset{\mathcal{D}[0,\overline{\gamma}]}{\Longrightarrow}$ denotes the weak convergence in the Skorokhod space $\mathcal{D}[0, \overline{\gamma}]$. The existence of the variances of $S_{n,i,j}(\gamma)$ and $T_{n,i}(\gamma)$ requires $\mathrm{E}|Y_0|^\kappa < \infty$ with $\kappa = 2 + 2\max(\alpha_1, \alpha_2)$, i.e., $\mathrm{E}|\epsilon_0|^\kappa < \infty$ under H_0. For testing against EXPAR we need $\mathrm{E}\epsilon_0^{10} < \infty$ and $\mathrm{E}\epsilon_0^6 < \infty$ in the LSTAR model. However, the tightness condition, which is used in the proof of the following theorem, requires a stronger moment condition.

THEOREM 3.1. *Let* $b_0 = 0$ *and suppose that* $\mathrm{E}|Y_0|^\kappa < \infty$ *with* $\kappa = 2 + 2\max(\alpha_1, \alpha_2, \alpha_3)$. *Then, under Assumptions A0–A3, for any* $\overline{\gamma} > 0$,

$$\frac{\sqrt{n}}{\sigma}\left(S_{n,0,0}, S_{n,1,0}, S_{n,1,1}(\gamma), T_{n,1}(\gamma), T_{n,2}(\gamma)\right)$$

$$\overset{\mathcal{D}[0,\overline{\gamma}]}{\Longrightarrow} \left(W(1), \int_{\mathbb{R}} x\, dW(F(x)), \int_{\mathbb{R}} x H(\gamma, x)\, dW(F(x)),\right.$$

$$\left.\int_{\mathbb{R}} x H_1(\gamma, x)\, dW(F(x)), \int_{\mathbb{R}} x H_2(\gamma, x)\, dW(F(x))\right),$$

where W *is a standard Brownian motion and* $F(x) = P(Y_0 \le x)$.

For the next result we need the following assumption.

A4. For any constants K_1, K_2 and any $0 \le \gamma \le \overline{\gamma}$,

$$P[Y_0 = K_1 Y_0 H(\gamma, Y_0) + K_2] < 1 \quad \text{and} \quad P[Y_0 = K_1 Y_0 H_1(0, Y_0) + K_2] < 1.$$

We can now state the following result. By convention, for $\gamma = 0$ we set $\hat{\mu}_\gamma = \tilde{\mu}$, $\hat{a}_\gamma = \tilde{a}$, $\gamma \hat{b}_\gamma = 0$, and $\hat{b}_\gamma Y_{t-1} H(\gamma, \cdot) = 0$.

THEOREM 3.2. *Under the assumptions of Theorem 3.1 and condition A4 we have*

$$\sup_{0 \le \gamma \le \overline{\gamma}} |\hat{\mu}_\gamma - \mu_0| = O_P(n^{-1/2}), \qquad \sup_{0 \le \gamma \le \overline{\gamma}} |\hat{\mu}_\gamma - \tilde{\mu}| = O_P(n^{-1/2}),$$

$$\sup_{0 \le \gamma \le \overline{\gamma}} |\hat{a}_\gamma - a_0| = O_P(n^{-1/2}), \qquad \sup_{0 \le \gamma \le \overline{\gamma}} |\hat{a}_\gamma - \tilde{a}| = O_P(n^{-1/2}),$$

$$\sup_{0 \le \gamma \le \overline{\gamma}} \gamma\, |\hat{b}_\gamma| = O_P(n^{-1/2}).$$

Now we turn to asymptotic properties of the constrained and unconstrained LS estimators of σ^2, which are defined, respectively, by

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{t=2}^{n} (Y_t - \tilde{\mu} - \tilde{a} Y_{t-1})^2,$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

$$\hat{\sigma}_\gamma^2 = \frac{1}{n}\sum_{t=2}^{n}\{Y_t - \hat{\mu}_\gamma - \hat{a}_\gamma Y_{t-1} - \hat{b}_\gamma Y_{t-1} H(\gamma, Y_{t-1})\}^2. \quad (3.5)$$

THEOREM 3.3. *Under the assumptions of Theorem 3.2 we have*

$$\sup_{0\le\gamma\le\overline{\gamma}} |\hat{\sigma}_\gamma^2 - \sigma^2| = o_P(1).$$

4. LINEARITY TESTING

Given that model (1.1) involves four parameters, a natural idea would be to estimate the parameters $(\mu_0, a_0, b_0, \gamma_0)$ of the unconstrained model by quasi maximum likelihood estimation. The asymptotic properties of this estimator could be derived when no identifiability problem arises, i.e., $b_0\gamma_0 \neq 0$. The constraint $b_0 \neq 0$ is however an important restriction. When $b_0 = 0$ the parameter γ_0 is not identified, so that we do not know the behavior of the quasi maximum likelihood estimator when the data generating process is an AR(1). Consequently the test of

$$H_0 : b_0 = 0 \quad \text{against } H_1 : b_0 \neq 0$$

is not standard. We first consider a strategy based on setting an arbitrary value of γ. Then the testing problem can be easily solved by a standard test, using, e.g., the Wald, LM, or LR principle.

4.1. Setting an Arbitrary Value γ

Fixing an arbitrary value of γ for the nuisance parameter, a convenient form for the Wald-type, LM-type, and LR-type statistics is given by

$$\mathbf{W}_n(\gamma) = n\frac{\tilde{\sigma}^2 - \hat{\sigma}_\gamma^2}{\hat{\sigma}_\gamma^2}, \qquad \mathbf{LM}_n(\gamma) = n\frac{\tilde{\sigma}^2 - \hat{\sigma}_\gamma^2}{\tilde{\sigma}^2}, \qquad \mathbf{LR}_n(\gamma) = n\log\frac{\tilde{\sigma}^2}{\hat{\sigma}_\gamma^2}. \quad (4.1)$$

The form of these statistics is obtained under normal errors, but we do not make this assumption in the discussion that follows. The expression for the LR statistic is the standard one. For the Wald statistic, the standard expression is

$$\mathbf{W}_n(\gamma) = n\frac{\hat{b}_\gamma^2}{\hat{\sigma}_{\hat{b}_\gamma}^2}, \quad \text{where } \hat{\sigma}_{\hat{b}_\gamma}^2 = \hat{\sigma}_\gamma^2 J_n^{-1}(\gamma)(3,3),$$

where $J_n^{-1}(\gamma)(3,3)$ denotes the $(3,3)$ element of the matrix $J_n^{-1}(\gamma)$. The form given in (4.1) for $\mathbf{W}_n(\gamma)$, and similarly for $\mathbf{LM}_n(\gamma)$, relies on the linearity of the model when γ is fixed. See, e.g., Godfrey (1988) and Gouriéroux and Monfort (1995). Notice that $\mathbf{W}_n(0) = \mathbf{LM}_n(0) = \mathbf{LR}_n(0) = 0$ because $\tilde{\sigma}^2$ and $\hat{\sigma}_\gamma^2$, as defined in (3.5), are equal when $\gamma = 0$.

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

For every $\gamma > 0$, the three statistics $\mathbf{W}_n(\gamma)$, $\mathbf{LM}_n(\gamma)$, and $\mathbf{LR}_n(\gamma)$ are asymptotically χ^2_1 distributed under H_0. Note that the tests based on those statistics are in general consistent, even for alternatives such that $\gamma_0 \neq \gamma$. However this procedure may lack power for alternatives where γ_0 is far from γ. In other words, the test statistics are sensitive to γ, and so this coefficient cannot be selected in a completely arbitrary way if it is not known. On the other hand, when γ_0 is unknown, then its LS estimator $\hat{\gamma}$ can be found by minimizing $\hat{\sigma}^2_\gamma$ over $\Gamma = [0, \overline{\gamma}]$. A plug-in approach seems natural, but the asymptotic null distribution of $\mathbf{W}_n(\hat{\gamma})$, $\mathbf{LR}_n(\hat{\gamma})$, and $\mathbf{LM}_n(\hat{\gamma})$ is no longer χ^2_1.

4.2. Using Supremum Statistics

The sup-LR statistic is defined by

$$\mathbf{LR}_n = \sup_{\gamma \in \Gamma} \mathbf{LR}_n(\gamma) = n \log \frac{\tilde{\sigma}^2}{\hat{\sigma}^2}, \quad \text{where } \hat{\sigma}^2 = \inf_{\gamma \in \Gamma} \hat{\sigma}^2_\gamma = Q_n(\hat{\theta}).$$

The sup-Wald and the LM statistics can similarly be defined as

$$\mathbf{W}_n = \sup_{\gamma \in \Gamma} \mathbf{W}_n(\gamma) = n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}, \qquad \mathbf{LM}_n = \sup_{\gamma \in \Gamma} \mathbf{LM}_n(\gamma) = n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\tilde{\sigma}^2}.$$

Note that the sup-LR statistic is actually the conventional LR statistic, i.e., $\mathbf{LR}_n = \mathbf{LR}_n(\hat{\gamma})$ where $\hat{\gamma} = \arg\inf_{\gamma \in \Gamma} \hat{\sigma}^2_\gamma$ is the LS estimator of γ_0. In the next theorem we will obtain the asymptotic null distribution of the LR, LM, and Wald statistics. Given that $\mathbf{W}_n \geq \mathbf{W}_n(\gamma)$ for any n, the same inequality holds asymptotically for any γ, and the asymptotic distribution of $\mathbf{W}_n$ is expected to be different from the $\chi^2(1)$. Figure 1 reveals an important difference between the two distributions.

The sup-Wald statistic is also the conventional Wald statistic. This is less straightforward than it is for the LR statistic because the model is no longer linear when γ is not fixed, and so it is not obvious that a form equivalent to (4.1) holds for the standard Wald statistic. However we have

$$\mathbf{W}_n(\hat{\gamma}) = n \frac{\hat{b}^2_{\hat{\gamma}}}{\hat{\sigma}^2_{\tilde{b}_{\hat{\gamma}}}} = n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2_{\hat{\gamma}}}{\hat{\sigma}^2_{\hat{\gamma}}} = n \frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} = \mathbf{W}_n,$$

noting that $\hat{\sigma}^2 = \hat{\sigma}^2_{\hat{\gamma}}$. The same remark holds for the LM statistic.

The main result of this paper is the following theorem, providing the asymptotic null distribution of the supremum test statistics.

THEOREM 4.1. *Suppose the conditions of Theorem 3.3, in particular the null hypothesis H_0, hold. Then, for any $\overline{\gamma} > 0$*

$$\mathbf{W}_n = \sup_{\gamma \in [0, \overline{\gamma}]} \mathbf{W}_n(\gamma) \Longrightarrow \mathbf{W} := \sup_{\gamma \in (0, \overline{\gamma}]} \mathbf{W}(\gamma),$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

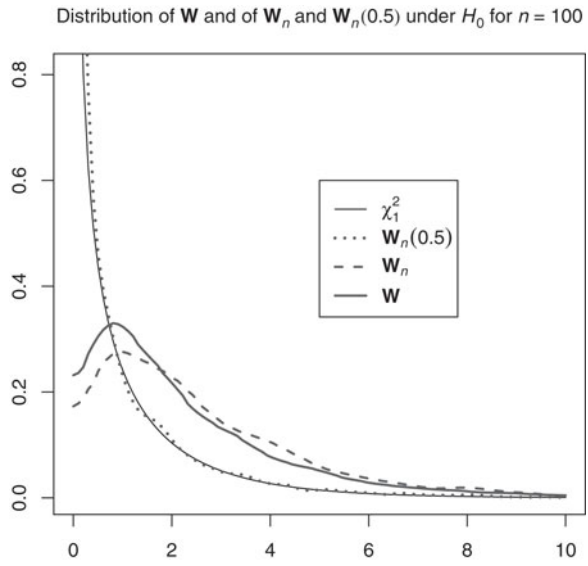


FIGURE 1. For the EXPAR model, kernel density estimator of the distribution of $\mathbf{W}_n(0.5)$ (dotted line) and limiting distribution of $\mathbf{W}_n(0.5)$ (the χ^2_1 distribution, thin full line), kernel density estimator of the distribution of $\mathbf{W}_n = \sup_{\gamma \in [0,100]} \mathbf{W}_n(\gamma)$ (dashed line), and kernel density estimator of the distribution of $\mathbf{W}$ (i.e., the limiting distribution of $\mathbf{W}_n$, thick full line). The density estimators are obtained by computing the statistics on $N = 5{,}000$ independent replications of $\mathcal{N}(0, 1)$ simulated samples of length $n = 100$ for the first two kernel density estimators and of length 500 for the last one.

where for $\gamma > 0$

$$\mathbf{W}(\gamma) = \frac{\{V(0)Z(\gamma) - V(\gamma)Z(0) - \Delta(\gamma)\sigma W(1)\}^2}{\sigma^2 D(\gamma)\mathrm{Var}(Y_0)},$$

with

$$Z(\gamma) = \int_{\mathbb{R}} x\{H(\gamma, x) + 1\}\, dW(F(x)),$$

$$V(\gamma) = \mathrm{Cov}(Y_0, Y_0\{H(\gamma, Y_0) + 1\}),$$

$$\Delta(\gamma) = \mathrm{E}Y_0^2 \mathrm{E}Y_0 H(\gamma, Y_0) - \mathrm{E}Y_0 \mathrm{E}Y_0^2 H(\gamma, Y_0),$$

$$D(\gamma) = \mathrm{Var}(Y_0)\mathrm{Var}(Y_0 H(\gamma, Y_0)) - \{\mathrm{Cov}(Y_0, Y_0 H(\gamma, Y_0))\}^2.$$

Moreover,

$$\sup_{\gamma \in [0,\overline{\gamma}]} \mathbf{LM}_n(\gamma) \Longrightarrow \sup_{\gamma \in (0,\overline{\gamma}]} \mathbf{W}(\gamma), \qquad \sup_{\gamma \in [0,\overline{\gamma}]} \mathbf{LR}_n(\gamma) \Longrightarrow \sup_{\gamma \in (0,\overline{\gamma}]} \mathbf{W}(\gamma).$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

Contrary to the standard situation (γ fixed) where the asymptotic distribution is a $\chi^2(1)$ whatever the model, the law of $\mathbf{W}$ depends on the model, through the function H.

Notice that $\mathbf{W}(\gamma)$ is not defined when $\gamma = 0$ because $D(0) = 0$. However, the limiting distribution of $\mathbf{W}(\gamma)$ when $\gamma \to 0$ is nondegenerate and is that of a $\chi^2(1)$. Lemma A.7 in the Appendix shows that we can define $\mathbf{W}(0)$ as

$$\lim_{\gamma \to 0} \mathbf{W}(\gamma) = \mathbf{W}(0),$$

where the limit exists with probability one. It is clear that the law of $\mathbf{W}(0)$ is not the limiting distribution of $\mathbf{W}_n(0)$, which is always equal to zero. In other words,

$$\lim_n \lim_{\gamma \to 0} \mathbf{W}_n(\gamma) = 0 \quad \text{a.s.}, \quad \text{but} \quad \lim_{\gamma \to 0} \lim_n \mathbf{W}_n(\gamma) \sim \chi^2(1).$$

It is important to notice that we do not require that γ be bounded away from zero. The supremum can be taken over all possible values of the nuisance parameter, instead of restricting γ to a compact subset excluding zero as is done when testing for structural change (see Andrews, 1993). Note that the framework of the present paper is quite different from that of Andrews (1993). When testing for a structural break, it seems necessary to bound the nuisance parameter away from zero. The reason is that the asymptotic distribution of the test statistic indexed by the nuisance parameter π, say, is a function of a Brownian bridge. Thus, when taking the supremum over the full range of values of π, the statistic diverges under the null hypothesis (see Andrews, 1993, Cor. 1). In our setup, the asymptotic distribution of the test statistics indexed by γ is a process whose supremum is well behaved for all possible values of the nuisance parameter belonging to a bounded set.[1]

This theorem can be adapted to deal, more generally, with statistics of the form $g(\{\mathbf{W}_n(\gamma), \gamma \in [0, \overline{\gamma}]\})$ for arbitrary functions g that are continuous with respect to the uniform metric (and likewise for $\mathbf{LM}_n(\cdot)$ and $\mathbf{LR}_n(\cdot)$). The use of a function g that differs from the sup function can depend on the alternatives of interest. See Andrews and Ploberger (1994) for discussion of different statistics of this form.

4.3. Model without Intercept

When the intercept is not present in model (1.1), i.e., when

$$Y_t = \{a_0 + b_0 H(\gamma_0, Y_{t-1})\} Y_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{IID}(0, \sigma^2), \tag{4.2}$$

the results are slightly different. The test statistics are still of the form (4.1) but with

$$\tilde{\sigma}^2 = \min_a \frac{1}{n} \sum_{t=2}^{n} (Y_t - aY_{t-1})^2,$$

$$\hat{\sigma}_\gamma^2 = \min_{a,b} \frac{1}{n} \sum_{t=2}^{n} \{Y_t - aY_{t-1} - bY_{t-1} H(\gamma, Y_{t-1})\}^2.$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

We give them without proof, keeping the previous notation with obvious adaptations.

THEOREM 4.2. *Suppose that* $H_0 : b_0 = 0$ *in Model (4.2). Let the assumptions of Theorem 3.1 be satisfied. Then, the results of Theorem 4.1 continue to hold with*

$$\mathbf{W}(\gamma) = \frac{\{V(0)Z(\gamma) - V(\gamma)Z(0)\}^2}{\sigma^2 D(\gamma)\mathrm{E}Y_0^2},$$

$V(\gamma), Z(\gamma)$ *as in Theorem 4.1, and* $D(\gamma) = \mathrm{E}Y_0^2\mathrm{E}Y_0^2H^2(\gamma, Y_0) - \{\mathrm{E}Y_0^2H(\gamma, Y_0)\}^2$.

It can be noted that the asymptotic distribution depends on constants and $\{Z(\gamma), \gamma \geq 0\}$, which is a zero mean Gaussian process with covariance kernel

$$K(\gamma, \gamma') = \mathrm{E}Z(\gamma)Z(\gamma') = \sigma^2\mathrm{E}[Y_0^2\{H(\gamma, Y_0) + 1\}\{H(\gamma', Y_0) + 1\}].$$

It is interesting to see that in general, unless $\mathrm{E}Y_0H(\gamma, Y_0) = 0$, the distribution of the process $\{\mathbf{W}(\gamma), \gamma > 0\}$ is not simply obtained from that of Theorem 4.1 with μ_0 replaced by 0.

4.4. Implementation

We now focus on the practical implementation of the tests of this paper. For simplicity, we present the results for model (4.2) without intercept. Some of the results of this section are not new but are given for the reader's convenience.

4.4.1. Computation of the Test Statistics. We focus on the LM statistic, which is very easy to compute. Following Godfrey (1988), the $\mathbf{LM}_n(\gamma)$ test can be implemented as follows:

1. fit an AR(1), compute the residuals $\tilde{\epsilon}_t$ and the residual sum of squares $\mathrm{RSS} = n\tilde{\sigma}^2$;
2. regress linearly $\tilde{\epsilon}_t$ on Y_{t-1} and $Y_{t-1}H(\gamma, Y_{t-1})$; compute the residual sum of squares RSS_γ and the uncentered determination coefficient R_γ^2 of the regression.

Noting that the residuals of the second regression are also the residuals of the regression of Y_t on Y_{t-1} and $Y_{t-1}H(\gamma, Y_{t-1})$, we have $\mathrm{RSS}_\gamma = n\hat{\sigma}_\gamma^2$, which gives $\mathbf{LM}_n(\gamma) = nR_\gamma^2 = n(\mathrm{RSS} - \mathrm{RSS}_\gamma)/\mathrm{RSS}$.

For the computation of the $\mathbf{LM}_n$ statistic we can replace step 2 by

2′. compute the residual sum of squares $\mathrm{RSS}_{\hat{\gamma}} = n\hat{\sigma}^2$ of the nonlinear regression model

$$\tilde{\epsilon}_t = cY_{t-1} + bY_{t-1}H(\gamma, Y_{t-1}) + \epsilon_t.$$

We have $\mathbf{LM}_n = n(\mathrm{RSS} - \mathrm{RSS}_{\hat{\gamma}})/\mathrm{RSS}$.

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

4.4.2. Computation of the Critical Values. In view of (A.19) in the Appendix, and following Hansen (1996), one can approximate the distribution of $\sup_{\gamma\in[0,\overline{\gamma}]}\mathbf{W}_n(\gamma)$ by that of

$$\sup_{\gamma\in[0,\overline{\gamma}]}\widehat{\mathbf{W}}_n(\gamma),\qquad \widehat{\mathbf{W}}_n(\gamma)=\frac{\{V_n(0)Z_n^\circ(\gamma)-V_n(\gamma)Z_n^\circ(0)\}^2}{D_n(\gamma)U_{n,2,0}},$$

where $V_n(0)=U_{n,2,0}$, $V_n(\gamma)=U_{n,2,1}(\gamma)+U_{n,2,0}$, where $U_{n,2,j}(\gamma)$ and $D_n(\gamma)$ are defined by (3.2) and (A.20) and where $\{Z_n^\circ(\gamma),\gamma\geq 0\}$ is, conditionally on the observation $Y_1,\ldots,Y_n$, a zero mean Gaussian process with covariance kernel

$$K_n(\gamma,\gamma')=\mathrm{E}Z_n^\circ(\gamma)Z_n^\circ(\gamma')=\frac{1}{n}\sum_{t=2}^{n}Y_{t-1}^2\{H(\gamma,Y_{t-1})+1\}\{H(\gamma',Y_{t-1})+1\}].$$

The conditional distribution of $\sup_{\gamma\in[0,\overline{\gamma}]}\widehat{\mathbf{W}_n(\gamma)}$ can be obtained by the following algorithm.

Algorithm 1. For $i=1,\ldots,N$:

(i) generate a $\mathcal{N}(0,1)$ sample $\epsilon_1^{(i)},\ldots,\epsilon_n^{(i)}$;

(ii) set $Z_n^{(i)}(\gamma)=n^{-1/2}\sum_{t=2}^{n}\epsilon_t^{(i)}Y_{t-1}\{H(\gamma,Y_{t-1})+1\}$;

(iii) set $\widehat{\mathbf{W}}_n^{(i)}(\gamma)=\left\{V_n(0)Z_n^{(i)}(\gamma)-V_n(\gamma)Z_n^{(i)}(0)\right\}^2D_n^{-1}(\gamma)U_{n,2,0}^{-1}$;

(iv) compute $\sup_{\gamma\in[0,\overline{\gamma}]}\widehat{\mathbf{W}}_n^{(i)}(\gamma)$.

Conditional on $Y_1,\ldots,Y_n$, the sequence $\sup_{\gamma\in[0,\overline{\gamma}]}\widehat{\mathbf{W}}_n^{(i)}(\gamma)$, $i=1,\ldots,N$ constitutes an independent and identically distributed sample of the random variable $\sup_{\gamma\in[0,\overline{\gamma}]}\widehat{\mathbf{W}}_n(\gamma)$. At the nominal level α, the common critical value c_α of the tests of rejection regions

$$\left\{\sup_{\gamma\in[0,\overline{\gamma}]}\mathbf{W}_n(\gamma)>c_\alpha\right\},\qquad \left\{\sup_{\gamma\in[0,\overline{\gamma}]}\mathbf{LM}_n(\gamma)>c_\alpha\right\}\quad\text{or}$$

$$\left\{\sup_{\gamma\in[0,\overline{\gamma}]}\mathbf{LR}_n(\gamma)>c_\alpha\right\}$$

will be defined as being the empirical $(1-\alpha)$-quantile of the artificial sample $\sup_{\gamma\in[0,\overline{\gamma}]}\widehat{\mathbf{W}}^{(i)}(\gamma)$, $i=1,\ldots,N$.

4.4.3. Cases where the Limiting Law Is Parameter-Free. We now describe a situation where the previous algorithm (i)–(iv) can be avoided and the critical values of the test can be obtained once and for all. Assume that

$$H(\gamma,y)=\mathfrak{h}(\gamma y^k) \tag{4.3}$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

for some integer k and some measurable function $\mathfrak{h}(\cdot)$. Note that the previous assumption is satisfied in the EXPAR case (2.1) with $k = 2$ and in the LSTAR case (2.2) with $k = 1$, when the location parameter $c = 0$.

Denote by $\sigma_{Y_0}^2$ the variance of Y_0. Let $\breve{V}(\gamma)$, $\breve{D}(\gamma)$, and $\breve{K}(\gamma, \gamma')$ be obtained by replacing Y_0 by $\sigma_{Y_0}^{-1} Y_0$ and σ^2 by one in the definition of $V(\gamma)$, $D(\gamma)$, and $K(\gamma, \gamma')$ given immediately after Theorem 4.2 and let the process $\{\breve{Z}(\gamma) = \sigma^{-1}\sigma_{Y_0}^{-1} Z(\gamma\, \sigma_{Y_0}^{-k}), \gamma \geq 0\}$. By (4.3) we conclude that

$$V(\gamma) = \sigma_{Y_0}^2 \breve{V}(\gamma\, \sigma_{Y_0}^k), \qquad D(\gamma) = \sigma_{Y_0}^4 \breve{D}(\gamma\, \sigma_{Y_0}^k),$$

$$K(\gamma, \gamma') = \sigma^2 \sigma_{Y_0}^2 \breve{K}(\gamma\, \sigma_{Y_0}^k, \gamma' \sigma_{Y_0}^k),$$

and $\{\breve{Z}(\gamma), \gamma \geq 0\}$ is a zero mean a.s. continuous Gaussian process with covariance kernel $\breve{K}(\gamma, \gamma')$. We thus have

$$\sup_{\gamma \in (0,\, \sigma_{Y_0}^{-k} \overline{\gamma}]} \mathbf{W}(\gamma) = \sup_{\gamma \in (0,\, \sigma_{Y_0}^{-k} \overline{\gamma}]} \breve{\mathbf{W}}(\gamma\, \sigma_{Y_0}^k) = \sup_{\gamma \in (0, \overline{\gamma}]} \breve{\mathbf{W}}(\gamma),$$

where

$$\breve{\mathbf{W}}(\gamma) = \frac{\{\breve{V}(0)\breve{Z}(\gamma) - \breve{V}(\gamma)\breve{Z}(0)\}^2}{\breve{D}(\gamma)\mathrm{E}(\sigma_{Y_0}^{-2} Y_0^2)}.$$

Note that when ϵ_t is Gaussian, the moments $\breve{V}(\gamma)$, $\breve{D}(\gamma)$, and $\mathrm{E}(\sigma_{Y_0}^{-2} Y_0^2)$, and also the distribution of the process $\breve{Z}(\cdot)$, do not depend on any unknown parameter. In particular the kernel is explicitly given by

$$\breve{K}(\gamma, \gamma') = \int y^2 \{H(\gamma, y) + 1\}\{H(\gamma', y) + 1\} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

We deduce that in the Gaussian case, i.e., when ϵ_t is Gaussian, the asymptotic distribution of

$$\sup_{\gamma \in [0,\, \hat{\sigma}_Y^{-k} \overline{\gamma}]} \mathbf{W}_n(\gamma), \quad \hat{\sigma}_Y^2 = \frac{1}{n} \sum_{t=1}^n Y_t^2 - \left(\frac{1}{n} \sum_{t=1}^n Y_t \right)^2$$

is parameter-free under H_0 (i.e., does not depend on a_0 and σ^2). In consequence, the distribution of $\sup_{\gamma \in [0,\, \hat{\sigma}_Y^{-k} \overline{\gamma}]} \mathbf{W}_n(\gamma)$ can be approximated by that of the Wald statistic $\breve{\mathbf{W}}_n$ obtained by replacing $Y_1, \ldots, Y_n$ by an $\mathcal{N}(0, 1)$ sample $\epsilon_1, \ldots, \epsilon_n$ in $\mathbf{W}_n = \sup_{\gamma \in (0, \overline{\gamma}]} \mathbf{W}_n(\gamma)$.

Based on empirical quantiles over $N = 100{,}000$ independent replications of $\breve{\mathbf{W}}_n$ with $n = 500$, Table 1 displays approximated critical values for the EXPAR model (2.1). Note that in this model the regressor $Y_{t-1} \exp\left(-\gamma\, Y_{t-1}^2\right) \to 0$ in probability as $\gamma \to \infty$. More precisely, straightforward computation shows that $\mathrm{var}\left\{Y_t \exp\left(-\gamma\, Y_t^2\right)\right\} = (1 + 4\gamma)^{-3/2} \simeq 0.01 \mathrm{var} Y_t$ when $\gamma = 5$ and $Y_t \sim \mathcal{N}(0, 1)$. Thus, for an EXPAR model with parameter $|a_0 + b_0| < 1$ (which, by Theorem 2.1,

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

TABLE 1. Asymptotic critical values of the sup-tests of the hypothesis H_0 : $b_0 = 0$ in the EXPAR model (2.1)–(4.2) with $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$

		$\overline{\gamma}$												
α	$\chi^2_1(1-\alpha)$	0.04	0.08	0.12	0.16	0.2	0.4	0.6	0.8	1	2	3	4	5
1%	6.6	6.8	6.9	6.9	7.1	7.2	7.5	7.7	7.9	8.0	8.4	8.6	8.8	8.8
5%	3.8	4.0	4.1	4.1	4.2	4.3	4.5	4.7	4.8	5.0	5.3	5.4	5.6	5.7
10%	2.7	2.8	2.9	2.9	3.0	3.1	3.3	3.4	3.6	3.7	4.0	4.1	4.3	4.3

guarantees the stationarity) and γ_0 such that $\gamma_0 \text{var}(Y_t) > 5$, the part of the variance that is explained by the nonlinear term is low. This heuristic argument and Monte Carlo experiments lead us to think that any choice of $\overline{\gamma}$ between 2 and 5 is suitable. Note that when $\overline{\gamma} \to 0$ the critical value of the sup-test tends to the critical value of a standard test, i.e., the $(1-\alpha)$-quantile of the χ^2_1 distribution. Approximated critical values of the sup-tests based on the LSTAR model (2.2) are given in Table 2. Note that, in view of the remark given after Theorem 4.2, Tables 1 and 2 are also valid for the model with intercept (1.1) when the intercept $\mu_0 = 0$.

4.5. Monte Carlo Experiments

For testing linearity against STAR, such as the LSTAR, the test (hereafter LST) proposed by Luukkonen et al. (1988) is the most commonly used (see Tong, 1990; Granger and Teräsvirta, 1993). When applied to the LSTAR model (2.2), the simplest version of the LST test, denoted by 1-LST, consists of testing $a_1 = 0$ in the auxiliary model

$$Y_t = a_0 Y_{t-1} + a_1 Y_{t-1}^2 + \epsilon_t. \tag{4.4}$$

This auxiliary model is obtained from the Taylor expansion $H(\gamma, y) = \gamma y/4 + o(\gamma)$ and a reparameterization of the model (see Luukkonen et al., 1988). The second-order Taylor expansion of $H(\gamma, y)$ is the same as the first one: $H(\gamma, y) = \gamma y/4 + o(\gamma^2)$. The third-order Taylor expansion $H(\gamma, y) = \gamma y/4 - \gamma^3 y^3/48 + o(\gamma^3)$ leads to the 2-LST version, which consists of testing $a_1 = a_2 = 0$ in the

TABLE 2. As Table 1, for the LSTAR model (2.2)–(4.2)

		$\overline{\gamma}$												
α	$\chi^2_1(1-\alpha)$	0.04	0.08	0.12	0.16	0.2	0.4	0.6	0.8	1	2	3	4	5
1%	6.6	6.6	6.6	6.6	6.6	6.6	6.7	6.7	6.8	6.9	7.1	7.2	7.3	7.3
5%	3.8	3.9	3.9	3.9	3.9	3.9	3.9	4.0	4.0	4.1	4.3	4.4	4.4	4.4
10%	2.7	2.7	2.7	2.7	2.7	2.7	2.8	2.8	2.9	2.9	3.1	3.1	3.2	3.2

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

auxiliary model

$$Y_t = a_0 Y_{t-1} + a_1 Y_{t-1}^2 + a_2 Y_{t-1}^4 + \epsilon_t. \quad \textbf{(4.5)}$$

These tests are extremely simple and easy to implement, their critical values being, respectively, the quantiles $\chi_1^2(1-\alpha)$ and $\chi_2^2(1-\alpha)$. The same approach can be used for the EXPAR model (2.1) and leads to the 1-LST test of $a_1 = 0$ in the auxiliary model $Y_t = a_0 Y_{t-1} + a_1 Y_{t-1}^3 + \epsilon_t$ and the 2-LST test of $a_1 = a_2 = 0$ in the auxiliary model $Y_t = a_0 Y_{t-1} + a_1 Y_{t-1}^3 + a_2 Y_{t-1}^5 + \epsilon_t$.

The sup-tests and LST tests of linearity have been applied to $N = 1{,}000$ independent simulations of size $n = 100$ of model (4.2) for different nonlinear terms $H(\gamma_0, Y_{t-1})$ and different values of the parameters a_0, b_0, and γ_0. Two versions of the sup-tests are considered, corresponding to $\overline{\gamma} = 2$ and $\overline{\gamma} = 5$. The critical values of these tests are taken from Tables 1 and 2. Table 3 displays the relative frequency of rejection of the null $H_0 : b_0 = 0$ at the nominal level $\alpha = 1\%$, 5%, and 10%. With the designs I and II, in which the null hypothesis holds, the relative rejection frequency over the $N = 1{,}000$ replications is almost always within the 0.05 significant limits, which are 0.3% and 1.7% for $\alpha = 1\%$, 3.6% and 6.4% for $\alpha = 5\%$, and 8.1% and 11.9% for $\alpha = 10\%$. The rare exceptions are displayed in bold type in Table 3. For the designs III and IV, the null hypothesis does not hold,

TABLE 3. Empirical size and power of tests of the linearity hypothesis H_0 at the nominal level α; the rejection frequencies are computed on $N = 1{,}000$ independent replications of simulations of length $n = 100$

		Sup-tests $\overline{\gamma} = 2$			Sup-tests $\overline{\gamma} = 5$			1-LST			2-LST		
Model	α	W	LM	LR	W	LM	LR	W	LM	LR	W	LM	LR
I	1%	1.3	1.0	1.1	1.3	0.8	1.1	1.0	0.9	1.0	1.2	0.8	1.0
	5%	5.7	4.6	5.1	5.4	4.4	4.7	4.9	4.5	4.8	5.5	4.2	4.8
	10%	10.6	9.4	9.8	10.1	9.4	10.1	9.2	8.5	9.1	10.9	9.9	10.3
II	1%	1.4	1.0	1.4	1.5	1.0	1.1	1.5	1.3	1.4	1.1	0.7	0.7
	5%	5.9	5.1	5.6	**6.7**	5.6	6.0	5.8	5.1	5.5	5.9	5.3	5.7
	10%	10.7	10.2	10.4	10.8	10.4	10.5	11.6	11.0	11.3	**12.4**	11.3	11.8
III	1%	6.3	4.4	5.3	5.7	4.3	4.9	1.4	1.0	1.1	5.3	3.9	4.4
	5%	19.0	16.8	17.6	16.9	14.8	15.9	7.0	6.4	6.7	14.5	12.6	13.5
	10%	27.5	25.8	26.5	27.2	25.9	26.5	13.2	12.8	13.0	22.5	21.0	21.9
IV	1%	24.9	20.9	22.6	26.3	22.4	24.2	18.2	16.5	17.2	19.8	14.8	17.8
	5%	48.1	45.8	47.0	50.5	48.2	49.5	40.4	38.9	39.6	40.2	37.0	39.1
	10%	61.9	60.7	61.6	64.2	62.9	63.6	54.5	53.4	53.9	53.2	50.8	51.7

Note: I: EXPAR model under H_0: Model (2.1)–(4.2) with $a_0 = b_0 = 0$, $\epsilon_t \sim \mathcal{N}(0,1)$.
II: LSTAR model under H_0: Model (2.2)–(4.2) with $a_0 = b_0 = 0$, $\epsilon_t \sim \mathcal{N}(0,1)$.
III: EXPAR model under H_1: Model (2.1)–(4.2) with $a_0 = b_0 = \gamma_0 = 0.4$, $\epsilon_t \sim \mathcal{N}(0,1)$.
IV: LSTAR model under H_1: Model (2.2)–(4.2) with $a_0 = b_0 = 0.4$, $\gamma_0 = 4$, $\epsilon_t \sim \mathcal{N}(0,1)$.

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

and, as expected, the sup-tests are more powerful than the LST tests. In Table 3 the highest rejection frequencies are underlined. One can see that the rejection frequencies of the LM, LR, and Wald tests are systematically in increasing order, both under the null and under the alternative, which is a well-known (see, e.g., Godfrey, 1988) consequence of the forms of the test statistics. In summary, all the tests well control the error of the first kind; the sup-tests are more powerful than the LST tests and are not too sensitive to the choice of $\overline{\gamma}$. Note however that the LST tests remain very attractive for their simplicity and their relative good performance.

NOTE

1. We thank a referee for pointing out to us the difference between the two kinds of testing problems.

REFERENCES

Adler, R.J. (1990) *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes.* Institute of Mathematical Statistics Lecture Notes, Monograph Series, 12. Hayward.

Andrews, D.W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Andrews, D.W.K. & W. Ploberger (1994) Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62, 1383–1414.

Andrews, D.W.K. & W. Ploberger (1995) Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. *Annals of Statistics* 23, 1609–1629.

Billingsley, P. (1961) The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Chan, K.S. & H. Tong (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 178–190.

Chesher, A. (1984) Testing for neglected heterogeneity. *Econometrica* 52, 865–872.

Davies, R.B. (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.

Davies, R.B. (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74, 33–43.

Godfrey, L.G. (1988) *Misspecification Tests in Econometrics.* Cambridge University Press.

Gouriéroux, C. & A. Monfort (1995) *Statistics and Econometric Models*, vol. II. Cambridge University Press.

Granger, C.W.J. & T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships.* Oxford University Press.

Haggan, V. & T. Ozaki (1981) Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* 68, 189–196.

Hansen, B. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

King, M. & T.S. Shively (1993) Locally optimal testing when a nuisance parameter is present only under the alternative. *Review of Economics and Statistics* 75, 1–7.

Lee, L.-F. & A. Chesher (1986) Specification testing when score test statistics are identically zero. *Journal of Econometrics* 31, 121–149.

Luukkonen, R., P. Saikkonen, & T. Teräsvirta (1988) Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–499.

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

Pötscher, B.M. & I.R. Prucha (1989) A uniform law of large numbers for dependent and heterogeneous data processes. *Econometrica* 57, 675–683.

Rotnitzky, A., D. Cox, M. Bottai, & J. Robins (2000) Likelihood-based inference with singular information matrix. *Bernoulli* 6, 243–284.

Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.

Stock, J. & M.W. Watson (1999) A comparison of linear and nonlinear univariate models for forecasting macroeconomic times series. In R.F. Engle & H. White (eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, pp. 1–44, Oxford University Press.

Teräsvirta, T., D. van Dijk, & M.C. Medeiros (2004) Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting* 21, 755–774.

Tjøstheim, D. (1990) Non-linear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.

Tong, H. (1990) *Non-linear Time Series. A Dynamical System Approach*. Oxford University Press.

Tong, H. & K.S. Lim (1980) Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B* 42, 245–292.

APPENDIX

Proof of Theorem 3.1. The first step is the convergence of the finite-dimensional distributions. Note that the sequences of variables involved in the $S_{n,i}(\gamma)$ and $T_{n,i}(\gamma)$ are square-integrable stationary martingale differences. The conclusion follows from the central limit theorem of Billingsley (1961).

It remains to show that the sequences are tight. We have, by the independence between ϵ_t and Y_{t-1}, for some γ_1 between γ and γ',

$$\begin{aligned}\frac{n}{n-1}\mathrm{E}\left(\frac{\sqrt{n}}{\sigma}\{S_{n,1,1}(\gamma)-S_{n,1,1}(\gamma')\}\right)^2 &= \mathrm{E}\left[Y_0^2\{H(\gamma,Y_0)-H(\gamma',Y_0)\}^2\right]\\ &= \mathrm{E}\left[Y_0^2H_1^2(\gamma_1,Y_0)\right](\gamma-\gamma')^2\\ &\le K\mathrm{E}\left[Y_0^2(|Y_0|^{\alpha_1}+1)^2\right](\gamma-\gamma')^2 \le K(\gamma-\gamma')^2,\end{aligned}$$

where the last inequalities follow from Assumption A1 and the existence of $\mathrm{E}(Y_0^{2+2\alpha_1})$ under $b_0=0$.

For any $\gamma' \in (0,\gamma)$, forgetting the asymptotically irrelevant factor $n/(n-1)$, we similarly have

$$\begin{aligned}\mathrm{E}\left(\frac{\sqrt{n}}{\sigma}\{T_{n,1}(\gamma)-T_{n,1}(\gamma')\}\right)^2 &= \mathrm{E}\left[Y_0^2\{H_1(\gamma,Y_0)-H_1(\gamma',Y_0)\}^2\right]\\ &= \mathrm{E}\left[Y_0^2H_2^2(\gamma_1,Y_0)\right](\gamma-\gamma')^2\\ &\le K\mathrm{E}\left[Y_0^2(|Y_0|^{\alpha_2}+1)^2\right](\gamma-\gamma')^2 \le K(\gamma-\gamma')^2,\\ \mathrm{E}\left(\frac{\sqrt{n}}{\sigma}\{T_{n,2}(\gamma)-T_{n,2}(\gamma')\}\right)^2 &= \mathrm{E}\left[Y_0^2\{H_2(\gamma,Y_0)-H_2(\gamma',Y_0)\}^2\right]\\ &\le K\mathrm{E}\left[Y_0^2(|Y_0|^{\alpha_3}+1)^2\right](\gamma-\gamma')^{2\alpha} \le K(\gamma-\gamma')^2.\end{aligned}$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

The tightness follows from Theorem 12.3 of Billingsley (1968, p. 95). To complete the proof let us show that the limiting Gaussian process has the form given by the theorem.

The processes $\sqrt{n}T_{n,i}(\gamma)$ and $\sqrt{n}S_{n,i,j}(\gamma)$, $i,j=0,1$, are in the form $n^{-1/2}\sum_{t=2}^{n}\epsilon_t \times \ell(\gamma, Y_{t-1})$ with some function ℓ. Because ϵ_t and Y_{t-1} are independent, we get that for any γ, γ'

$$\mathrm{E}\left\{n^{-1/2}\sum_{t=2}^{n}\epsilon_t\ell_1(\gamma,Y_{t-1})n^{-1/2}\sum_{s=2}^{n}\epsilon_s\ell_2(\gamma',Y_{s-1})\right\}=\mathrm{E}\left(\epsilon_0^2\right)\mathrm{E}\{\ell_1(\gamma,Y_0)\ell_2(\gamma',Y_0)\}$$

$$=\sigma^2\int_{\mathbb{R}}\ell_1(\gamma,y)\ell_2(\gamma',y)\,dF(y).$$

Thus we computed the covariance structure of the limiting Gaussian process. So it is enough to show that the stochastic integrals in the limit have the same covariance structure. It is easy to see that

$$\mathrm{E}\int_{\mathbb{R}}\ell_1(\gamma,y)\,dW(F(y))\int_{\mathbb{R}}\ell_2(\gamma,x)\,dW(F(x))=\int_{\mathbb{R}}\ell_1(\gamma,y)\ell_2(\gamma',y)\,dF(y),$$

and therefore the representation is established. ■

Before proving Theorem 3.2, we establish the following lemmas. Proofs are given for the reader's convenience but can be found elsewhere in a much more general framework (see, e.g., Pötscher and Prucha, 1989).

LEMMA A.1. *Let $(Z_t(\gamma))$ denote a strictly stationary and ergodic real-valued process with $\mathrm{E}|Z_0(\gamma)|<\infty$ for any $\gamma\in\Gamma$ where Γ is a real compact set. Assume that*

$$|Z_t(\gamma)-Z_t(\gamma')|\le A_t|\gamma-\gamma'|^{\alpha}, \tag{A.1}$$

where $\alpha>0$ and (A_t) is a strictly stationary and ergodic process with $\mathrm{E}A_0<\infty$. Then

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}Z_t(\gamma)-\mathrm{E}Z_0(\gamma)\right|\to 0\quad a.s.$$

Proof. The compact set Γ is covered by m balls $B(\gamma_i,\delta)$ of center γ_i, $i=1,\dots,m$, and radius $\delta>0$. We have

$$\sup_{\gamma\in\Gamma}\left|\frac{1}{n}\sum_{t=1}^{n}Z_t(\gamma)-\mathrm{E}Z_0(\gamma)\right|\le c_{1n}+c_{2n}+c_{3n},$$

where

$$c_{1n}=\max_{i=1,\dots,m}\sup_{\gamma\in B(\gamma_i,\delta)}\left|\frac{1}{n}\sum_{t=1}^{n}Z_t(\gamma)-Z_t(\gamma_i)\right|,$$

$$c_{2n}=\max_{i=1,\dots,m}\left|\frac{1}{n}\sum_{t=1}^{n}Z_t(\gamma_i)-\mathrm{E}Z_0(\gamma_i)\right|,$$

$$c_{3n}=\max_{i=1,\dots,m}\sup_{\gamma\in B(\gamma_i,\delta)}|\mathrm{E}Z_0(\gamma_i)-\mathrm{E}Z_0(\gamma)|.$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

By (A.1) and the ergodic theorem

$$c_{1n} \le \delta^\alpha \frac{1}{n}\sum_{t=1}^{n} A_t \to \delta^\alpha \mathrm{E}A_0 \quad \text{and} \quad c_{2n} \to 0 \quad \text{a.s.}$$

Condition (A.1) entails that $\left|\mathrm{E}Z_0(\gamma) - \mathrm{E}Z_0(\gamma')\right| \le \mathrm{E}A_0 \left|\gamma - \gamma'\right|^\alpha$. Thus $c_{3n} \le \delta^\alpha \mathrm{E}A_0$, and the conclusion follows. ■

LEMMA A.2. *Under the assumptions of Theorem 3.2 we have* $|S_{n,i,0}| = O_P(n^{-1/2})$ *for* $i = 0, 1$ *and*

$$\sup_{\gamma\in[0,\overline{\gamma}]} |S_{n,1,1}(\gamma)| = O_P(n^{-1/2}), \qquad \sup_{\gamma\in[0,\overline{\gamma}]} |T_{n,j}(\gamma)| = O_P(n^{-1/2}) \quad \textit{for } j = 1, 2. \tag{A.2}$$

Proof. This is a direct consequence of Theorem 3.1. ■

LEMMA A.3. *Under the assumptions of Theorem 3.2 we have*

$$\sup_{\gamma\in[0,\overline{\gamma}]} |U_{n,i,j}(\gamma) - \mathrm{E}Y_0^i H^j(\gamma, Y_0)| \to 0 \quad a.s., \qquad i = 1, 2, \qquad j = 0, 1, 2. \tag{A.3}$$

Moreover, for any $\delta > 0$ *there exists* $\epsilon > 0$ *such that*

$$\overline{\lim_{n\to\infty}} \sup_{\gamma\in(0,\epsilon]} \left|\frac{U_{n,i,j}(\gamma)}{\gamma^j} - \mathrm{E}Y_0^i H_1^j(0, Y_0)\right| < \delta \quad a.s. \tag{A.4}$$

Proof. First note that the expectations in (A.3) exist by Assumption A1 and

$$Y_0^2 H^i(\gamma, Y_0) \le K\gamma^i Y_0^2(|Y_0|^{\alpha_1} + 1)^i, \tag{A.5}$$

which is integrable because $\mathrm{E}|Y_0|^{2+2\alpha_1} < \infty$. The convergence in (A.3) follows, using Lemma A.1 and Assumptions A1–A3. The following expansions around $\gamma = 0$ hold. For some $\nu_1, \nu_2 \in (0, \gamma)$,

$$U_{n,i,1}(\gamma) = \frac{\gamma}{n}\sum_{t=2}^{n} Y_{t-1}^i H_1(0, Y_{t-1}) + \frac{\gamma^2}{2n}\sum_{t=2}^{n} Y_{t-1}^i H_2(\nu_1, Y_{t-1}),$$

$$\begin{aligned} U_{n,i,2}(\gamma) &= \frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^i H^2(\gamma, Y_{t-1}) \\ &= \frac{\gamma^2}{n}\sum_{t=2}^{n} Y_{t-1}^i \{H_1^2(\nu_2, Y_{t-1}) + H(\nu_2, Y_{t-1}) H_2(\nu_2, Y_{t-1})\}. \end{aligned} \tag{A.6}$$

Hence

$$\begin{aligned} \sup_{\gamma\in(0,\epsilon]} \left|\frac{U_{n,i,1}(\gamma)}{\gamma} - \mathrm{E}Y_0^i H_1(0, Y_0)\right| \le{}& \left|\frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^i H_1(0, Y_{t-1}) - \mathrm{E}Y_0^i H_1(0, Y_0)\right| \\ &+ \sup_{\gamma\in[0,\epsilon]} \left|\frac{\gamma}{2n}\sum_{t=2}^{n} Y_{t-1}^i H_2(\nu_1, Y_{t-1})\right|. \end{aligned}$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

By the ergodic theorem

$$\frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^{i} H_1(0, Y_{t-1}) \to \mathrm{E} Y_0^i H_1(0, Y_0) \quad \text{a.s.}$$

Moreover, by A2 and because $\mathrm{E}|\epsilon_t|^{2+\alpha_2} < \infty$,

$$\left|\frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^{i} H_2(\nu_1, Y_{t-1})\right|$$

$$\leq \frac{K}{n}\sum_{t=2}^{n} Y_{t-1}^{i}(|Y_{t-1}|^{\alpha_2}+1) \to K\mathrm{E} Y_0^i(|Y_{t-1}|^{\alpha_2}+1) < \infty, \quad \text{a.s.}$$

It follows that a.s.

$$\overline{\lim_{n\to\infty}} \sup_{\gamma\in(0,\epsilon]} \left|\frac{U_{n,i,1}(\gamma)}{\gamma} - \mathrm{E} Y_0^i H_1(0, Y_0)\right| \leq \epsilon K. \tag{A.7}$$

Now in view of (A.6), by A1–A2 and another Taylor expansion of $H(\cdot, Y_{t-1})$ around zero we have, for some $\nu_3 \in (0, \nu_2)$,

$$\begin{aligned}
\sup_{\gamma\in(0,\epsilon]} \left|\frac{U_{n,i,2}(\gamma)}{\gamma^2} - \mathrm{E} Y_0^i H_1^2(0, Y_0)\right| &\leq \left|\frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^{i} H_1^2(0, Y_{t-1}) - \mathrm{E} Y_0^i H_1^2(0, Y_0)\right| \\
&\quad + \sup_{\gamma\in(0,\epsilon]} \left|\frac{\nu_2}{n}\sum_{t=2}^{n} Y_{t-1}^{i} 2H_1(\nu_3, Y_{t-1}) H_2(\nu_3, Y_{t-1})\right| \\
&\quad + \sup_{\gamma\in(0,\epsilon]} \left|\frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^{i} H(\nu_2, Y_{t-1}) H_2(\nu_1, Y_{t-1})\right| \\
&\leq \left|\frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^{i} H_1^2(0, Y_{t-1}) - \mathrm{E} Y_0^i H_1^2(0, Y_0)\right| \\
&\quad + \epsilon K \left|\frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^{i}(|Y_{t-1}|^{\alpha_1}+1)(|Y_{t-1}|^{\alpha_2}+1)\right|.
\end{aligned}$$

Thus, by arguments already given, a.s.

$$\overline{\lim_{n\to\infty}} \sup_{\gamma\in(0,\epsilon]} \left|\frac{U_{n,i,2}(\gamma)}{\gamma^2} - \mathrm{E} Y_0^i H_1^2(0, Y_0)\right| \leq \epsilon K.$$

This, together with (A.7), shows that (A.4) holds for ϵ sufficiently small. The proof of Lemma A.3 is complete. ■

Recall that

$$\begin{aligned}
D_n(\gamma) &= \det\{J_n(\gamma)\} \\
&= \{U_{n,2,2}(\gamma) - U_{n,1,1}^2(\gamma)\}\{U_{n,2,0} - U_{n,1,0}^2\} - \{U_{n,2,1}(\gamma) - U_{n,1,0}U_{n,1,1}(\gamma)\}^2
\end{aligned} \tag{A.8}$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

and let $D(\gamma) = \det\{J(\gamma)\}$ where

$$J(\gamma) = \begin{pmatrix} 1 & U_{1,0} & U_{1,1}(\gamma) \\ U_{1,0} & U_{2,0} & U_{2,1}(\gamma) \\ U_{1,1}(\gamma) & U_{2,1}(\gamma) & U_{2,2}(\gamma) \end{pmatrix}, \quad U_{i,j}(\gamma) = \mathrm{E} Y_0^i H^j(\gamma, Y_0).$$

Let also $A = \det\{J^{(1)}\}$ where

$$J^{(1)} = \begin{pmatrix} 1 & U_{1,0} & U_{1,1}^{(1)} \\ U_{1,0} & U_{2,0} & U_{2,1}^{(1)} \\ U_{1,1}^{(1)} & U_{2,1}^{(1)} & U_{2,2}^{(1)} \end{pmatrix}, \quad U_{i,j}^{(1)} = \mathrm{E} Y_0^i H_1^j(0, Y_0).$$

LEMMA A.4. *Under the assumptions of Theorem 3.2 we have*

$$\sup_{\gamma \in [0, \overline{\gamma}]} |D_n(\gamma) - D(\gamma)| \to 0 \quad a.s. \tag{A.9}$$

and $D(\gamma) > 0$ *for any* $\gamma > 0$. *Moreover, for any* $\delta > 0$ *there exists* $\epsilon > 0$ *such that*

$$\overline{\lim_{n \to \infty}} \sup_{\gamma \in (0, \epsilon]} \left| \frac{D_n(\gamma)}{\gamma^2} - A \right| < \delta \quad a.s., \tag{A.10}$$

where $A > 0$.

Proof. The convergence in (A.9) follows from Lemma A.3 and from (A.5) for the existence of the expectations. We have

$$\begin{aligned} &\left| \frac{D_n(\gamma)}{\gamma^2} - A \right| \\ &\quad \le \mathrm{Var}(Y_0) \left| \frac{U_{n,2,2}(\gamma) - U_{n,1,1}^2(\gamma)}{\gamma^2} - \mathrm{Var}\{Y_0 H_1(0, Y_0)\} \right| \\ &\qquad + \left| \frac{U_{n,2,1}(\gamma) - U_{n,1,0} U_{n,1,1}(\gamma)}{\gamma} - \mathrm{Cov}\{Y_0, Y_0 H_1(0, Y_0)\} \right|^2 \\ &\qquad + \frac{|U_{n,2,2}(\gamma) - U_{n,1,1}^2(\gamma)|}{\gamma^2} |U_{n,2,0} - U_{n,1,0}^2 - \mathrm{Var}(Y_0)| \\ &\qquad + 2 \frac{|U_{n,2,1}(\gamma) - U_{n,1,0} U_{n,1,1}(\gamma)|}{\gamma} \\ &\qquad\quad \times \left| \frac{U_{n,2,1}(\gamma) - U_{n,1,0} U_{n,1,1}(\gamma)}{\gamma} - \mathrm{Cov}\{Y_0, Y_0 H_1(0, Y_0)\} \right|. \end{aligned}$$

Note that by (A.4), for ϵ sufficiently small

$$\overline{\lim_{n \to \infty}} \sup_{\gamma \in (0, \epsilon]} \left| \frac{U_{n,i,j}(\gamma)}{\gamma^j} \right| \le K \quad \text{a.s.} \tag{A.11}$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

Then (A.10) straightforwardly follows from Lemma A.3. Note that

$$D(\gamma) = \mathrm{Var}\{Y_0 H(\gamma, Y_0)\}\mathrm{Var}(Y_0) - \mathrm{Cov}^2\{Y_0 H(\gamma, Y_0), Y_0\},$$

$$A = \mathrm{Var}\{Y_0 H_1(0, Y_0)\}\mathrm{Var}(Y_0) - \mathrm{Cov}^2\{Y_0 H_1(0, Y_0), Y_0\}.$$

The nonnegativity of $D(\gamma)$ and A follows from the Cauchy–Schwarz inequality. Moreover, $D(\gamma) = 0$ and $A = 0$ are precluded by Assumption A4. ■

LEMMA A.5. *Under the assumptions of Theorem 3.2 we have for any* $\delta > 0$

$$\lim_{\epsilon \to 0} \overline{\lim_{n \to \infty}} P\left[n^{-1/2} \sup_{\gamma \in (0,\epsilon]} \left| \frac{1}{\gamma} \sum_{t=2}^{n} \epsilon_t Y_{t-1} H(\gamma, Y_{t-1}) - \sum_{t=2}^{n} \epsilon_t Y_{t-1} H_1(0, Y_{t-1}) \right| > \delta \right] = 0.$$

Proof. By the mean-value theorem we have for some $\nu \in (0, \gamma)$

$$\sum_{t=2}^{n} \epsilon_t Y_{t-1} H(\gamma, Y_{t-1}) = \gamma \sum_{t=2}^{n} \epsilon_t Y_{t-1} H_1(0, Y_{t-1}) + \frac{\gamma^2}{2} \sum_{t=2}^{n} \epsilon_t Y_{t-1} H_2(\nu, Y_{t-1}).$$

Thus

$$n^{-1/2} \sup_{\gamma \in (0,\epsilon]} \left| \frac{1}{\gamma} \sum_{t=2}^{n} \epsilon_t Y_{t-1} H(\gamma, Y_{t-1}) - \sum_{t=2}^{n} \epsilon_t Y_{t-1} H_1(0, Y_{t-1}) \right| \le \frac{\epsilon}{2} \sup_{\gamma \in [0,\epsilon]} \sqrt{n} \left| T_{n,2}(\gamma) \right|.$$

In view of Lemma A.2, by definition of a variable bounded in probability, the conclusion follows. ■

LEMMA A.6. *Under the assumptions of Theorem 3.2*

$$\sup_{\gamma \in (0,\overline{\gamma}]} \left| \frac{\sqrt{n} S_{n,1,1}(\gamma)}{\gamma} \right| = O_P(1).$$

Proof. For any $\delta > 0$ we have

$$P\left(\sup_{\gamma \in (0,\overline{\gamma}]} \left| \frac{\sqrt{n} S_{n,1,1}(\gamma)}{\gamma} \right| > \delta \right) \le c_{1,n}(\delta, \epsilon) + c_{2,n}(\delta, \epsilon) + c_{3,n}(\delta),$$

where ϵ is an arbitrary point of the interval $(0, \overline{\gamma}]$ and

$$c_{1,n}(\delta, \epsilon) = P\left(\sup_{\gamma \in (\epsilon,\overline{\gamma}]} \left| \frac{\sqrt{n} S_{n,1,1}(\gamma)}{\gamma} \right| > \frac{\delta}{3} \right),$$

$$c_{2,n}(\delta, \epsilon) = P\left(n^{-1/2} \sup_{\gamma \in (0,\epsilon]} \left| \frac{1}{\gamma} \sum_{t=2}^{n} \epsilon_t Y_{t-1} H(\gamma, Y_{t-1}) - \sum_{t=2}^{n} \epsilon_t Y_{t-1} H_1(0, Y_{t-1}) \right| > \frac{\delta}{3} \right),$$

$$c_{3,n}(\delta) = P\left(\left| \frac{1}{\sqrt{n}} \sum_{t=2}^{n} \epsilon_t Y_{t-1} H_1(0, Y_{t-1}) \right| > \frac{\delta}{3} \right).$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

Theorem 3.1 and the continuous mapping theorem show that

$$\sup_{\gamma\in(\epsilon,\overline{\gamma}]}\left|\frac{\sqrt{n}S_{n,1,1}(\gamma)}{\gamma}\right| \Rightarrow \sup_{\gamma\in(\epsilon,\overline{\gamma}]}\left|\frac{\sigma}{\gamma}\int_{\mathbb{R}} yH(\gamma,y)\,dW(F(y))\right| \quad \text{as } n\to\infty.$$

Because the limit is finite with probability one we get for any $\epsilon > 0$

$$\lim_{\delta\to\infty}\overline{\lim_{n\to\infty}} c_{1,n}(\delta,\epsilon) = 0.$$

By Lemma A.5 we have for all δ

$$\lim_{\epsilon\to 0}\overline{\lim_{n\to\infty}} c_{2,n}(\delta,\epsilon) = 0.$$

The central limit theorem of Billingsley (1961), for square-integrable stationary martingale differences, yields that, as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\epsilon_t Y_{t-1}H_1(0,Y_{t-1}) \Rightarrow \sigma\int_{-\infty}^{\infty} yH_1(0,y)dW(F(y)).$$

Thus we have

$$\lim_{\delta\to\infty}\overline{\lim_{n\to\infty}} c_{3,n}(\delta) = 0.$$

For all $\tau > 0$ let us choose δ_τ such that $\overline{\lim}_{n\to\infty} c_{3,n}(\delta_\tau) < \tau$, let $\epsilon_\tau > 0$ such that $\overline{\lim}_{n\to\infty} c_{2,n}(\delta_\tau,\epsilon_\tau) < \tau$, and let $\delta^*_\tau > 0$ such that $\overline{\lim}_{n\to\infty} c_{1,n}(\delta^*_\tau,\epsilon_\tau) < \tau$. Because $c_{1,n}(\cdot,\epsilon)$, $c_{2,n}(\cdot,\epsilon)$, and $c_{3,n}(\cdot)$ are decreasing functions, we have

$$\overline{\lim_{n\to\infty}} P\left(\sup_{\gamma\in(0,\overline{\gamma}]}\left|\frac{\sqrt{n}S_{n,1,1}(\gamma)}{\gamma}\right| > \delta\right) \le \overline{\lim_{n\to\infty}} c_{1,n}(\delta,\epsilon_\tau) + c_{2,n}(\delta,\epsilon_\tau) + c_{3,n}(\delta) < 3\tau$$

for all $\delta \ge \max\{\delta_\tau,\delta^*_\tau\}$. The conclusion follows. ■

Proof of Theorem 3.2. Denote by $J^*_n(\gamma) = \left(J^*_{n,i,j}(\gamma)\right)$ the matrix of cofactors of $J_n(\gamma)$. By (3.4) we have

$$\sqrt{n}|\hat{a}_\gamma - a_0| \le \frac{1}{|D_n(\gamma)|}\left\{|J^*_{n,2,1}(\gamma)||\sqrt{n}S_{n,0,0}| + |J^*_{n,2,2}(\gamma)||\sqrt{n}S_{n,1,0}(\gamma)|\right.$$
$$\left. + |J^*_{n,2,3}(\gamma)||\sqrt{n}S_{n,1,1}(\gamma)|\right\}. \tag{A.12}$$

Note that because D is a continuous function, we have $\inf_{\gamma\in[\epsilon,\overline{\gamma}]} D(\gamma) > 0$ for any $\epsilon > 0$. Thus, in view of

$$\inf_{\gamma\in[\epsilon,\overline{\gamma}]}|D_n(\gamma)| \ge \inf_{\gamma\in[\epsilon,\overline{\gamma}]} D(\gamma) - \sup_{\gamma\in[\epsilon,\overline{\gamma}]}|D_n(\gamma) - D(\gamma)|$$

we have

$$\sup_{\gamma\in[\epsilon,\overline{\gamma}]}\frac{1}{|D_n(\gamma)|} = O_P(1)$$

by Lemma A.4. Moreover, by Lemma A.3, $\sup_{\gamma\in[\epsilon,\overline{\gamma}]}|U_{n,i,j}(\gamma)| = O_P(1)$ for $i = 1,2$ and $j = 0,1,2$. Thus $\sup_{\gamma\in[\epsilon,\overline{\gamma}]}|J^*_{n,i,j}(\gamma)| = O_P(1)$. Then it follows from (A.12) and

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

Lemma A.2 that $\sup_{\gamma\in[\epsilon,\overline{\gamma}]}|\hat{a}_\gamma - a_0| = O_P(n^{-1/2})$ for all $\epsilon > 0$. Now we have

$$\inf_{\gamma\in(0,\epsilon]}\gamma^{-2}|D_n(\gamma)| \geq A - \sup_{\gamma\in(0,\epsilon]}|\gamma^{-2}D_n(\gamma) - A| \geq \frac{A}{2}$$

a.s. for n sufficiently large and ϵ sufficiently small, by Lemma A.4. Thus, for such n and ϵ we have a.s.

$$\begin{aligned}\sup_{\gamma\in(0,\epsilon]}\sqrt{n}|\hat{a}_\gamma - a_0| \leq \frac{2}{A}\Bigg\{&\sqrt{n}|S_{n,0,0}|\sup_{\gamma\in(0,\epsilon]}\frac{|J^*_{n,2,1}(\gamma)|}{\gamma^2}\\ &+\sup_{\gamma\in(0,\epsilon]}\frac{|J^*_{n,2,2}(\gamma)|}{\gamma^2}\sup_{\gamma\in(0,\epsilon]}|\sqrt{n}S_{n,1,0}|\\ &+\sup_{\gamma\in(0,\epsilon]}\frac{|J^*_{n,2,3}(\gamma)|}{\gamma}\sup_{\gamma\in(0,\epsilon]}\frac{|\sqrt{n}S_{n,1,1}(\gamma)|}{\gamma}\Bigg\}.\end{aligned}$$

It follows from (A.11) and Lemmas A.2 and A.6 that $\sup_{\gamma\in(0,\epsilon]}|\hat{a}_\gamma - a_0| = O_P(n^{-1/2})$ for ϵ small enough. Thus we have shown that $\sup_{\gamma\in(0,\overline{\gamma}]}|\hat{a}_\gamma - a_0| = O_P(n^{-1/2})$ for all $\overline{\gamma} > 0$. We prove that $\sup_{\gamma\in(0,\overline{\gamma}]}|\hat{\mu}_\gamma - \mu_0| = O_P(n^{-1/2})$ for all $\overline{\gamma} > 0$ by the same arguments.

Turning to $\hat{b}_\gamma$ we have by (3.4), with $b_0 = 0$, for n sufficiently large and ϵ sufficiently small

$$\begin{aligned}\sup_{\gamma\in(0,\epsilon]}\sqrt{n}|\gamma\,\hat{b}_\gamma| \leq \frac{2}{A}\Bigg\{&\sqrt{n}|S_{n,0,0}|\sup_{\gamma\in(0,\epsilon]}\frac{|J^*_{n,3,1}(\gamma)|}{\gamma}\\ &+\sup_{\gamma\in(0,\epsilon]}\frac{|J^*_{n,3,2}(\gamma)|}{\gamma}\sup_{\gamma\in(0,\epsilon]}|\sqrt{n}S_{n,1,0}|\\ &+|J^*_{n,3,3}|\sup_{\gamma\in(0,\epsilon]}\frac{|\sqrt{n}S_{n,1,1}(\gamma)|}{\gamma}\Bigg\}.\end{aligned}$$

By the arguments already given we thus find that $\sup_{\gamma\in[0,\overline{\gamma}]}|\gamma\,\hat{b}_\gamma| = O_P(n^{-1/2})$.

Finally, $\tilde{a} - a_0 = \{S_{n,1,0} - S_{n,0,0}U_{n,1,0}\}/\{U_{n,2,0} - U^2_{n,1,0}\} = O_P(n^{-1/2})$, which allows us to conclude that

$$\sup_{\gamma\in(0,\overline{\gamma}]}|\hat{a}_\gamma - \tilde{a}| \leq \sup_{\gamma\in(0,\overline{\gamma}]}|\hat{a}_\gamma - a_0| + |\tilde{a} - a_0| = O_P(n^{-1/2}).$$

The same arguments allow us to prove that $\sup_{\gamma\in(0,\overline{\gamma}]}|\hat{\mu}_\gamma - \tilde{\mu}| = O_P(n^{-1/2})$. The proof of Theorem 3.2 is now complete. ■

Proof of Theorem 3.3. We have

$$\begin{aligned}\hat{\sigma}^2_\gamma &= \frac{1}{n}\sum_{t=2}^{n}\{\epsilon_t + (\mu_0 - \hat{\mu}_\gamma) + (a_0 - \hat{a}_\gamma)Y_{t-1} - \hat{b}_\gamma Y_{t-1}H(\gamma, Y_{t-1})\}^2\\ &= \frac{1}{n}\sum_{t=2}^{n}\epsilon^2_t + (a_0 - \hat{a}_\gamma)^2\frac{1}{n}\sum_{t=2}^{n}Y^2_{t-1}\end{aligned}$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

$$+\hat{b}_\gamma^2 \frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^2 H^2(\gamma, Y_{t-1}) + \frac{2}{n}(a_0 - \hat{a}_\gamma)\sum_{t=2}^{n} \epsilon_t Y_{t-1}$$

$$-\hat{b}_\gamma \frac{2}{n}\sum_{t=2}^{n} \epsilon_t Y_{t-1} H(\gamma, Y_{t-1})$$

$$-(a_0 - \hat{a}_\gamma)\hat{b}_\gamma \frac{2}{n}\sum_{t=2}^{n} Y_{t-1}^2 H(\gamma, Y_{t-1}) + \frac{n-1}{n}(\mu_0 - \hat{\mu}_\gamma)^2$$

$$+(\mu_0 - \hat{\mu}_\gamma)\left\{\frac{2}{n}\sum_{t=2}^{n} \epsilon_t + \frac{2}{n}(a_0 - \hat{a}_\gamma)\sum_{t=2}^{n} Y_{t-1} - \hat{b}_\gamma \frac{2}{n}\sum_{t=2}^{n} Y_{t-1} H(\gamma, Y_{t-1})\right\}$$

$$= \frac{1}{n}\sum_{t=2}^{n} \epsilon_t^2 + (a_0 - \hat{a}_\gamma)^2 U_{n,2,0} + \{\gamma\,\hat{b}_\gamma\}^2 \frac{U_{n,2,2}(\gamma)}{\gamma^2} + 2(a_0 - \hat{a}_\gamma) S_{n,1,0}$$

$$-2\sqrt{n}\gamma\,\hat{b}_\gamma \frac{\sqrt{n} S_{n,1,1}(\gamma)}{n\gamma} - 2(a_0 - \hat{a}_\gamma)\gamma\,\hat{b}_\gamma \frac{U_{n,2,1}(\gamma)}{\gamma} + \frac{n-1}{n}(\mu_0 - \hat{\mu}_\gamma)^2$$

$$+2(\mu_0 - \hat{\mu}_\gamma)\left\{S_{n,0,0} + (a_0 - \hat{a}_\gamma) U_{n,1,0} - \gamma\,\hat{b}_\gamma \frac{U_{n,1,1}(\gamma)}{\gamma}\right\}.$$

The conclusion follows from the weak law of large numbers, Lemmas A.2, A.3, and A.6, and Theorem 3.2. ■

LEMMA A.7. *If the conditions of Theorem 3.3 are satisfied, then*

$$\lim_{\gamma \downarrow 0} \mathbf{W}(\gamma) = \mathbf{W}(0) \quad a.s.,$$

where

$$\mathbf{W}(0) = \frac{\left\{V(0)\int_{\mathbb{R}} y H_1(0, y)\, dW(F(y)) - \mathrm{E}Y_0^2 H_1(0, Y_0)\int_{\mathbb{R}} y\, dW(F(y)) - \Delta_1 \sigma W(1)\right\}^2}{A\,\mathrm{Var}(Y_0)},$$

$\Delta_1 = \mathrm{E}Y_0^2 \mathrm{E}Y_0 H_1(0, Y_0) - \mathrm{E}Y_0 \mathrm{E}Y_0^2 H_1(0, Y_0)$*, and A is defined before Lemma A.4.*

Proof. The dominated convergence theorem and Assumptions A0 and A1 entail that

$$\frac{D(\gamma)}{\gamma^2} \to A \qquad \text{as } \gamma \to 0.$$

Also, the same arguments show that when $\gamma \to 0$

$$\frac{V(\gamma) - V(0)}{\gamma} = \mathrm{E}Y_0^2 \frac{H(\gamma, Y_0)}{\gamma} \to \mathrm{E}Y_0^2 H_1(0, Y_0),$$

$$\frac{\Delta(\gamma)}{\gamma} = \mathrm{E}Y_0^2 \mathrm{E}Y_0 \frac{H(\gamma, Y_0)}{\gamma} - \mathrm{E}Y_0 \mathrm{E}Y_0^2 \frac{H(\gamma, Y_0)}{\gamma} \to \Delta_1.$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

Next we show that

$$\frac{Z(\gamma)-Z(0)}{\gamma}=\frac{\sigma}{\gamma}\int_{\mathbb{R}} yH(\gamma,y)\,dW(F(y))\to\sigma\int_{\mathbb{R}} yH_1(0,y)\,dW(F(y)) \tag{A.13}$$

a.s. as $\gamma\downarrow 0$. We define the following process on $[0,\overline{\gamma}]$:

$$u(\gamma)=\int_{\mathbb{R}} y\frac{1}{\gamma}H(\gamma,y)\,dW(F(y))\qquad \text{if } 0<\gamma\le\overline{\gamma}$$

and

$$u(0)=\int_{\mathbb{R}} yH_1(0,y)\,dW(F(y)).$$

The assertion in (A.13) is established if we show that $u(\gamma)$ is a.s. continuous on $[0,\overline{\gamma}]$. It follows from the definition that $u(\gamma)$ is a centered Gaussian process. Also,

$$\mathrm{E}\{u(\gamma)-u(0)\}^2=\int_{\mathbb{R}}\left\{y\left(\frac{1}{\gamma}H(\gamma,y)-H_1(0,y)\right)\right\}^2 dF(y)$$

and

$$\frac{1}{\gamma}H(\gamma,y)-H_1(0,y)=H_1(\gamma,y)+\frac{\gamma}{2}H_2(\gamma^*,y)-H_1(0,y)=\gamma\,H_2(\gamma^{**},y)+\frac{\gamma}{2}H_2(\gamma^*,y)$$

for some γ^* and γ^{**} between zero and γ. Thus, conditions A0 and A2 yield

$$\mathrm{E}\{u(\gamma)-u(0)\}^2\le C\gamma^2$$

for some constant C. On the other hand, for all $0<\gamma\le\gamma'\le\overline{\gamma}$ we have

$$\begin{aligned}\mathrm{E}\{u(\gamma')-u(\gamma)\}^2&=\int_{\mathbb{R}}\left\{y\left(\frac{1}{\gamma}H(\gamma,y)-\frac{1}{\gamma'}H(\gamma',y)\right)\right\}^2 dF(y)\\&\le 2\int_{\mathbb{R}}\left(y\frac{H(\gamma,y)-H(\gamma',y)}{\gamma}\right)^2 dF(y)\\&\quad+2\int_{\mathbb{R}}\left(y\frac{H(\gamma',y)}{\gamma'}\frac{\gamma'-\gamma}{\gamma}\right)^2 dF(y)\\&\le C\left(\frac{\gamma-\gamma'}{\gamma}\right)^2\end{aligned}$$

for some constant C. Thus we have

$$d(\gamma,\gamma'):=\left[\mathrm{E}\{u(\gamma)-u(\gamma')\}^2\right]^{1/2}\le\begin{cases}C\gamma' & 0=\gamma<\gamma'\le\overline{\gamma}\\ C\frac{\gamma'-\gamma}{\gamma} & 0<\gamma<\gamma'\le\overline{\gamma}.\end{cases}$$

Following Adler (1990) we provide an upper bound for the minimal number of d-balls of radius ϵ that cover $[0,\overline{\gamma}]$. According to Theorem 1.1 in Adler (1990, p. 4), it is enough to show that the number of d-balls is bounded by a polynomial function of $1/\epsilon$. The center

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

of the first ball is at zero, and it covers the interval $[0, \epsilon/C]$. Then the centers of the other balls are $\epsilon/C + i\epsilon^2/C^2$, $1 \le i \le [\overline{\gamma}C/\epsilon^2] + 1$. Indeed, when $\gamma' = \epsilon/C + i\epsilon^2/C^2$ and $\gamma = \epsilon/C + (i-1)\epsilon^2/C^2$ then $d(\gamma, \gamma') \le C(\gamma' - \gamma)/\gamma = \epsilon^2/(C\gamma) \le \epsilon$. Because the number of covering cloud d-balls is bounded by a multiple of $1/\epsilon^2$, the almost sure continuity of $u(\gamma)$ is established. Thus (A.13) holds. The conclusion then follows from

$$\mathbf{W}(\gamma) = \frac{\left\{ V(0)\frac{Z(\gamma)-Z(0)}{\gamma} - \frac{V(\gamma)-V(0)}{\gamma} Z(0) - \frac{\Delta(\gamma)}{\gamma}\sigma W(1) \right\}^2}{\sigma^2 \frac{D(\gamma)}{\gamma^2} \mathrm{E} Y_0^2}. \qquad \blacksquare$$

Proof of Theorem 4.1. By (3.1)–(3.3) we have

$$\hat{\delta}_\gamma - \tilde{\delta} = -\tilde{J}_n^{-1} J_n^{12}(\gamma)\hat{b}_\gamma \quad \text{where } J_n(\gamma) = \begin{pmatrix} \tilde{J}_n & J_n^{12}(\gamma) \\ J_n^{21}(\gamma) & U_{n,2,2}(\gamma) \end{pmatrix}.$$

Considering $\hat{\sigma}_\gamma^2$ and $\tilde{\sigma}^2$ as values of the same function at the points $(\hat{\delta}_\gamma, \hat{b}_\gamma)'$ and $(\tilde{\delta}, 0)'$, respectively, a second-order Taylor expansion gives, for any $\gamma > 0$,

$$\begin{aligned}
\tilde{\sigma}^2 &= \hat{\sigma}_\gamma^2 + \frac{1}{2}(\hat{\delta}_\gamma - \tilde{\delta}, \hat{b}_\gamma) 2 J_n(\gamma)(\hat{\delta}_\gamma - \tilde{\delta}, \hat{b}_\gamma)' \\
&= \hat{\sigma}_\gamma^2 + (\hat{\delta}_\gamma - \tilde{\delta})' \tilde{J}_n (\hat{\delta}_\gamma - \tilde{\delta}) + U_{n,2,2}(\gamma)\hat{b}_\gamma^2 + 2 J_n^{21}(\gamma)(\hat{\delta}_\gamma - \tilde{\delta})\hat{b}_\gamma \\
&= \hat{\sigma}_\gamma^2 + \hat{b}_\gamma^2 \left\{ U_{n,2,2}(\gamma) - J_n^{21}(\gamma)\tilde{J}_n^{-1} J_n^{12}(\gamma) \right\} \\
&= \hat{\sigma}_\gamma^2 + \hat{b}_\gamma^2 \{ J_n^{-1}(\gamma)(3,3) \}^{-1}
\end{aligned}$$

when $\mathrm{E}\epsilon_t^2 < \infty$. Thus we have

$$\begin{aligned}
\mathbf{W}_n(\gamma) &= n \frac{\hat{b}_\gamma^2}{\hat{\sigma}_\gamma^2 J_n^{-1}(\gamma)(3,3)} \\
&= \frac{\left\{ n^{1/2} S_{n,1,1}(\gamma) - J_n^{21}(\gamma)\tilde{J}_n^{-1} n^{1/2} (S_{n,0,0}, S_{n,1,0})' \right\}^2}{\hat{\sigma}_\gamma^2 \{ U_{n,2,2}(\gamma) - J_n^{21}(\gamma)\tilde{J}_n^{-1} J_n^{12}(\gamma) \}} \\
&= \frac{\left\{ n^{1/2} Z_n(\gamma) - (V_{n,1}(\gamma), V_{n,2}(\gamma))\tilde{J}_n^{-1} n^{1/2} (S_{n,0,0}, S_{n,1,0})' \right\}^2}{\hat{\sigma}_\gamma^2 \{ U_{n,2,2}(\gamma) - J_n^{21}(\gamma)\tilde{J}_n^{-1} J_n^{12}(\gamma) \}}, \\
&= \frac{\left\{ n^{1/2} Z_n(\gamma) V_n(0) - V_n(\gamma) n^{1/2} Z_n(0) - \Delta_n(\gamma) n^{1/2} S_{n,0,0} \right\}^2}{\hat{\sigma}_\gamma^2 (U_{n,2,0} - U_{n,1,0}^2) D_n(\gamma)},
\end{aligned} \tag{A.14}$$

where for $i = 1, 2$

$$V_{n,i}(\gamma) = \frac{1}{n}\sum_{t=2}^{n} Y_{t-1}^i \{ H(\gamma, Y_{t-1}) + 1 \}, \qquad V_n(\gamma) = V_{n,2}(\gamma) - V_{n,1}(\gamma) V_{n,1}(0),$$

$$\Delta_n(\gamma) = U_{n,2,0} U_{n,1,1}(\gamma) - U_{n,1,0} U_{n,2,1}(\gamma),$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

$$D_n(\gamma) = \det\{J_n(\gamma)\} = \{U_{n,2,2}(\gamma) - U_{n,1,1}^2(\gamma)\}\{U_{n,2,0} - U_{n,1,0}^2\} - \{U_{n,2,1}(\gamma) - U_{n,1,0}U_{n,1,1}(\gamma)\}^2,$$

$$Z_n(\gamma) = n^{-1}\sum_{t=2}^{n} \epsilon_t Y_{t-1}\{H(\gamma, Y_{t-1}) + 1\}.$$

A straightforward adaptation of Theorem 3.1 shows that for any $\overline{\gamma} > 0$

$$\left(\frac{\sqrt{n}}{\sigma} Z_n(\gamma), \frac{\sqrt{n}}{\sigma} S_{n,0,0}, \mathbf{W}_n^*\right) \overset{\mathcal{D}[0,\overline{\gamma}]\times\mathbb{R}\times\mathbb{R}}{\Longrightarrow} (Z(\gamma), W(1), \mathbf{W}(0)),$$

where

$$\mathbf{W}_n^* = \left\{ \mathrm{Var}(Y_0) n^{-1/2} \sum_{t=2}^{n} \epsilon_t Y_{t-1} H_1(0, Y_{t-1}) - \mathrm{Cov}(Y_0, Y_0 H_1(0, Y_0)) n^{-1/2} \sum_{t=2}^{n} \epsilon_t Y_{t-1} - \Delta_1 n^{-1/2} \sum_{t=2}^{n} \epsilon_t \right\}^2 \{\sigma^2 A \mathrm{Var}(Y_0)\}^{-1}$$

and $\mathbf{W}(0)$ is defined in Lemma A.7. Note that the same Brownian motion is used in the definitions of $Z(\gamma)$ and $\mathbf{W}(0)$. Thus we get that for all $0 < \underline{\gamma} < \overline{\gamma}$

$$\left(\mathbf{W}_n(\gamma), \mathbf{W}_n^*\right) \overset{\mathcal{D}[\underline{\gamma},\overline{\gamma}]\times\mathbb{R}}{\Longrightarrow} (\mathbf{W}(\gamma), \mathbf{W}(0)). \tag{A.15}$$

In view of (A.15) we have

$$\left(\sup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \mathbf{W}_n(\gamma), \mathbf{W}_n^*\right) \Longrightarrow \left(\sup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \mathbf{W}(\gamma), \mathbf{W}(0)\right). \tag{A.16}$$

By (A.14), Lemmas A.3 and A.5, and Theorems 3.2 and 3.3 we have for all $\delta > 0$

$$\lim_{\epsilon\to 0} \overline{\lim_{n\to\infty}} P\left(\sup_{\gamma\in[0,\epsilon]} \left|\mathbf{W}_n(\gamma) - \mathbf{W}_n^*\right| > \delta\right) = 0. \tag{A.17}$$

Also by Lemma A.7 we have for all $\delta > 0$

$$\lim_{\epsilon\to 0} P\left(\sup_{\gamma\in[0,\epsilon]} |\mathbf{W}(\gamma) - \mathbf{W}(0)| > \delta\right) = 0. \tag{A.18}$$

For any fixed $x \in \mathbb{R}$ and all $\delta > 0$ and $0 < \epsilon < \overline{\gamma}$, we have

$$P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}_n(\gamma) \le x\right) \le P\left(\sup_{\gamma\in[\epsilon,\overline{\gamma}]} \mathbf{W}_n(\gamma) \le x, \mathbf{W}_n^* \le x + \delta\right) + P\left(\sup_{\gamma\in[0,\epsilon]} \left|\mathbf{W}_n(\gamma) - \mathbf{W}_n^*\right| > \delta\right).$$

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press

Using (A.16) and (A.17), we obtain for all $\tau > 0$

$$\overline{\lim_{n\to\infty}} P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}_n(\gamma) \le x\right) \le P\left(\sup_{\gamma\in[\epsilon,\overline{\gamma}]} \mathbf{W}(\gamma) \le x, \mathbf{W}(0) \le x+\delta\right) + \tau$$

$$\le P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}(\gamma) \le x+\delta\right)$$

$$+ P\left(\sup_{\gamma\in[0,\epsilon]} |\mathbf{W}(\gamma) - \mathbf{W}(0)| > \delta\right) + \tau$$

for ϵ small enough. Using (A.18), we obtain

$$\overline{\lim_{n\to\infty}} P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}_n(\gamma) \le x\right) \le P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}(\gamma) \le x+\delta\right) + 2\tau.$$

Because the inequality holds for all δ and τ, we have

$$\overline{\lim_{n\to\infty}} P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}_n(\gamma) \le x\right) \le P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}(\gamma) \le x\right).$$

Using similar arguments, we obtain

$$\liminf_{n\to\infty} P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}_n(\gamma) \le x\right) \ge P\left(\sup_{\gamma\in[0,\overline{\gamma}]} \mathbf{W}(\gamma) < x\right).$$

Thus the proof is complete. ■

Proof of Theorem 4.2. The arguments of the proof of Theorem 4.1 can be used, replacing (A.14) by

$$\mathbf{W}_n(\gamma) = \frac{\left\{U_{n,2,0} n^{1/2} S_{n,1,1}(\gamma) - U_{n,2,1}(\gamma) n^{1/2} S_{n,1,0}\right\}^2}{\hat{\sigma}_\gamma^2 D_n(\gamma) U_{n,2,0}}$$

$$= \frac{\left\{V_{n,2}(0) n^{1/2} Z_n(\gamma) - V_{n,2}(\gamma) n^{1/2} Z_n(0)\right\}^2}{\hat{\sigma}_\gamma^2 D_n(\gamma) U_{n,2,0}}, \tag{A.19}$$

where

$$D_n(\gamma) = U_{n,2,0} U_{n,2,2}(\gamma) - U_{n,2,1}^2(\gamma). \tag{A.20}$$

■

https://doi.org/10.1017/S0266466609990430 Published online by Cambridge University Press