A Smooth Permanent Surge Process

Andrés González
Stockholm School of Economics
Banco de la República de Colombia

Stockholm School of Economics
SSE/EFI Working Paper Series in Economics and Finance
No.572

November 2004

Abstract

In this paper we introduce the Smooth Permanent Surge [SPS] model. The model is an integrated non lineal moving average process with possibly unit roots in the moving average coefficients. The process nests the Stochastic Permanent Break [STOPBREAK] process by Engle and Smith (1999) and in a limiting case it converges to Threshold Integrated Moving Average [TIMA] models by Gonzalo and Martínez (2003). A test of SPS against STOPBREAK process is presented. Additionally, we introduce a new test for testing SPS process against the random walk. The small sample properties of these tests are investigated by Monte Carlo experiments. An application to the stock markets is presented.

Keywords: Linearity test, Monte Carlo testing, Smooth transitions, Moving Averages Models, Permanent Shock, Transitory Shocks.

JEL codes: C12,C15,C22, C51, C52.

Stockholm School of Economics, P.O. Box 6501, SE-113 83 Stockholm, Sweden

1 Introduction

Recently the literature of time series analysis has been developing models in which stochastic shocks can have transitory as well as permanent effects. These models close the gap between stationary autoregressive models, in which all shocks are transitory, and models like the random walk in which all shocks are permanent. Example of these models are: the stochastic unit root process by Granger and Swanson (1997), which is an AR(1) process with the autoregressive parameter varying stochastically around one; the autoregressive conditional root by Rahbek and Shephard (2002), in which the autoregressive parameter changes between one and stationarity following a deterministic function of the past observations; the stochastic permanent break model by Engle and Smith (1999) in which the permanence of a given shock is stochastic and depends on its magnitude, and finally, Gonzalo and Martínez (2003) introduce a threshold integrated moving average model in which large shocks are permanent whereas small ones are transitory.

In the present paper we introduce the Smooth Permanent Surge [SPS] model. The model is a generalization of the stochastic permanent break model by Engle and Smith (1999). The permanent effect of an innovation is stochastic and depends on a deterministic function of past shocks. In the SPS model, small shocks have transitory effects and large shocks may have permanent effects. The model can be seen as an alternative both to the stochastic break model and to the threshold integrated moving average model.

We present three tests in the smooth permanent surge framework. The first is a test for linearity in moving average models. This test follows Brännäs, De Gooijer, and Teräsvirta (1998). The second test is a test of SPS against a random walk and the third is a test against the stochastic permanent break model by Engle and Smith (1999). The performance of these tests in small samples is evaluated by Monte Carlo experiments. Finally, in other to compare our model with the stochastic permanent break model, we apply our method to the same data set and economic problem as the one used in Engle and Smith (1999). That is, we investigate whether stock prices of two companies that belong to the same market move together or not.

The outline of the paper is the following. In the second section the smooth permanent surge model is introduced and conditions for invertibility of the model are given. The third section describes the proposed tests and explains their implementation. Section 4 presents the results of the Monte Carlo investigation. The application to the stock prices is presented in section 5. Section 6 concludes.

2 Smooth Permanent Surge model

The Stochastic Permanent Break [STOPBREAK] process of Engle and Smith (1999) is defined through the following equations:

$$y_t = m_t + \epsilon_t, \quad t = 0, 1, \dots, T \tag{1}$$

where ϵ_t is a stationary martingale difference sequence with respect to $\mathcal{F}_{t-1}$ where $\{\mathcal{F}_t\}$ denotes an increasing σ-algebra adapted to y_t. Furthermore, m_t is a time-varying conditional mean given by

$$m_t = m_{t-1} + q_{t-1}\epsilon_{t-1} \tag{2}$$

where q_{t-1} is a function of ϵ bounded by zero and one.

In order to characterize the dynamic properties of the STOPBREAK process it is useful to measure the effect that a given innovation will have on future values of y_t. One such measure is the *permanent effect of an innovation* defined by Engle and Smith (1999) as follows:

$$\lambda_t \overset{d}{=} \lim_{k \to \infty} \frac{\partial f(y_t, k)}{\partial \epsilon} \tag{3}$$

where $f(y_t, k) \equiv \mathsf{E}(y_{t+k}|\mathcal{F}_t)$, $\epsilon_t = y_t - \mathsf{E}(y_t|\mathcal{F}_{t-1})$. The *permanent effect of an innovation* in the STOPBREAK model is,

$$\begin{aligned} \lambda_t =& q_t + \frac{\partial q}{\partial \epsilon} |_\epsilon \\ =& q_t(1 + \eta_{q,t}) \end{aligned} \tag{4}$$

where $\eta_{q,t} \equiv (\partial q / \partial \epsilon |_\epsilon)(\epsilon_t / q_t)$.

From (4) it is seen that in the STOPBREAK model the long-run effect of an innovation is random and varies over time. The sign and magnitude of the effect depend on the specific functional form of q_t. For instance, if q_t is positive and has positive first derivatives with respect to $|\epsilon|$, $\lambda_t > 0$ for all t, and consequently all shocks have permanent effects.

Further understanding of the role of q_t in STOPBREAK models can be gained by writing (1) and (2) as an integrated nonlinear moving average model:

$$\Delta y_t = \epsilon_t - \varpi_{t-1}\epsilon_{t-1} \tag{5}$$

where $\varpi_{t-1} = 1 - q_{t-1}$. When $q_{t-1} = 1$ for all t, it follows that $\varpi_{t-1} = 0$, which implies that all shocks have permanent effects. On the contrary, when $q_{t-1} = 0$ for all t, $\varpi_{t-1} = 1$ and all innovations will have a transitory effect on y_t.

Different specifications for ϖ_{t-1} have been proposed in the literature. Engle and Smith

(1999) define $\varpi_{t-1} = 1 - q_{t-1}$ with

$$q_{t-1} = \epsilon_{t-1}^2/(\gamma + \epsilon_{t-1}^2). \tag{6}$$

In this specification, large positive and negative shocks have large (in absolute value) permanent effects while small shocks have small effects. The main drawback of this specification is that only zero shocks ($\epsilon_{t-1} = 0$) have transitory effects. To eliminate this drawback, Gonzalo and Martínez (2003) proposed the Shock-Exciting Threshold Integrated Moving Average [STIMA] model in which ϖ_{t-1} is an indicator function such that $\varpi_{t-1} = \theta_2$ for $|\epsilon_{t-1}| \leq \kappa$ and $\varpi_{t-1} = \theta_1$ otherwise. Hence, if $\theta_2 \leq 1$ small shocks have only transitory effects. The main disadvantage of this model is that due to the discontinuity of the likelihood function, statistical inference is nonstandard, and conducting inference is computationally expensive. Moreover, the STIMA model implies that shocks of either sign greater that $|\kappa|$ will have large permanent effects.

In this paper, ϖ_{t-1} is defined as

$$\varpi_{t-1} = \theta_1 + \theta_2 g(\epsilon; \gamma, \mathbf{c}) \tag{7}$$

where $g(\epsilon_t, \gamma, \mathbf{c})$ is the following logistic function [see Jansen and Teräsvirta (1996)]:

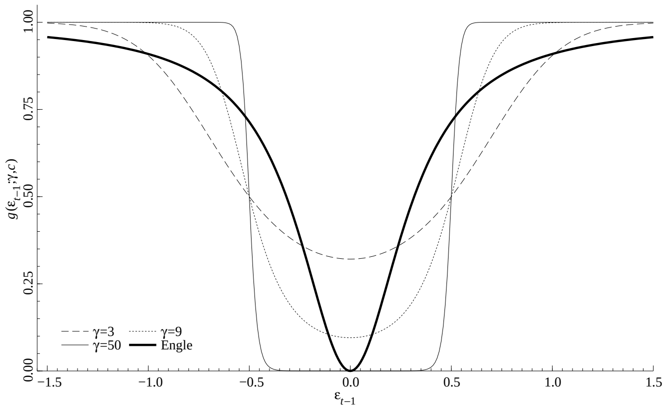
$$g(\epsilon_t; \gamma, \mathbf{c}) = (1 + \exp(-\gamma(\epsilon_t - c_1)(\epsilon_t - c_2)))^{-1} \tag{8}$$

with $\gamma > 0$ and $c_1 \leq c_2$. The model defined by (5), (7) and (8) is called Smooth Permanent Surge [SPS] model .

The definition of ϖ_{t-1} in the SPS model has similarities with both the STOPBREAK and the TIMA models. Figure 1 plots the transition function (8) for different values of γ. For comparison we also include (6). As can be seen, the transition function (8) has a U-shape form similar to the Logistic function (6). However, transition function (8) has a broader base. In fact, for relatively large values of γ, the transition function (8) practically takes value zero for all $\epsilon_{t-1} > c_1$ and $\epsilon_{t-1} < c_2$, $c_1 < c_2$.

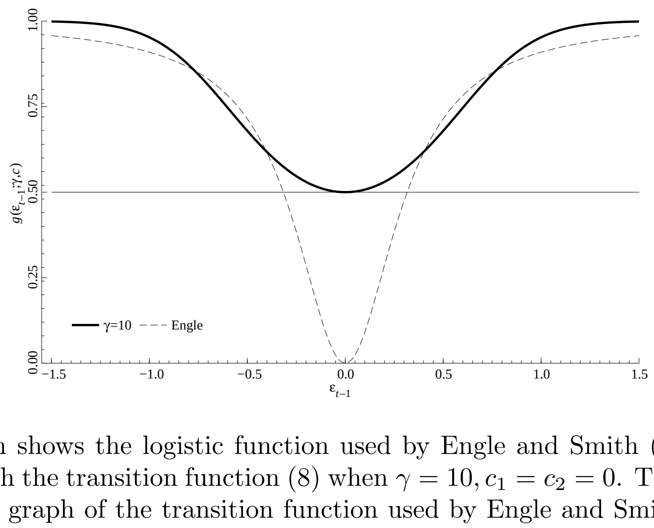
Under some parameter restrictions the behaviour of ϖ_{t-1} in the SPS model approximates the functional form of ϖ_{t-1} in the STOPBREAK one. For instance, when $c_1 = c_2 = 0$, in (8), the SPS model approximates the STOPBREAK model quite well. Figure 2 shows the transition function (8) with $c_1 = c_2 = 0$ together with the function used by Engle and Smith (1999). As can be seen, both transition functions are rather similar for $|\epsilon| > 0$ and attain their minimum at $\epsilon = 0$. A difference between them is that the minimum value in the Logistic function (6) equals zero whereas it equals 1/2 in the transition function (8). Consequently, if in addition to having $c_1 = c_2 = 0$ in (8) we have that $\theta_2 = -1$ and $\theta_2 = 2$ in (7) the SPS model is an approximation to the STOPBREAK model.

Figure 1: Transition functions for different γ values



The graph shows different transition functions as a function of ϵ_{t-1}. The thick line is the graph of the transition function (6) used by Engle and Smith (1999) with $\gamma = 0.1$. The other lines are the plots of the transition function (8) for different value of γ. $c_1 = -0.5$ and $c_2 = 0.5$.

Figure 2: Transition function used in the SPS together with the Logistic function used by Engle and Smith (1999)



The graph shows the logistic function used by Engle and Smith (1999) together with the transition function (8) when $\gamma = 10, c_1 = c_2 = 0$. The dashed line is the graph of the transition function used by Engle and Smith (1999) when $\gamma = 0.1$. The thick back line represents the transition function (8)

The SPS model can also approximate the STIMA model. In fact, the STIMA specification can be seen as a limiting case of an unrestricted SPS model. For instance, when $\gamma \to \infty$ and $c_1 = -c_2$ and $c_2 = \kappa > 0$ in (8), the SPS model is identical to a STIMA model. This feature of the SPS model makes it possible to describe nonlinearities that the STIMA model also captures. An advantage of the present model over STIMA is that the case $c_1 \neq -c_2$ is included. This feature allows asymmetries between the effects of large positive and negative innovations.

2.1 Invertibility

Since the SPS process is a moving average model, the estimation of parameters has to be carried out recursively. In fact, in order to estimate the model one has to be able to estimate the innovation process given the observed data and the generating formula. This is only possible if the model is invertible. Following the definition of invertibility by Granger and Andersen (1978), Engle and Smith (1999) established the invertibility conditions of STOPBREAK models in the following theorem:

Theorem 1 *The nonlinear moving average process in equation (5) is invertible if* $\mathsf{E}(|1-q_t(1+\eta_{qt})|F_{t-1}) \leq z_t < 1$, *where* $\eta_{qt} = (\epsilon_t/q_t)(\partial q/\partial \epsilon)\,|_{\epsilon_t}$ *and* $\{z_t\}$ *is a deterministic sequence defined such that* $\lim_{T\to\infty} \prod_{t=1}^{T} z_t = 0$

Proof. See Engle and Smith (1999). ■

Thus, for invertibility it is required that the average total effect of innovation has to be less than one. Applying Theorem 1 to the SPS model implies that the model is invertible if

$$\mathsf{E}\left[\left|\varpi_{t-1} + \frac{\partial \varpi_{t-1}}{\partial \epsilon_{t-1}}\epsilon_{t-1}\right| |\mathcal{F}_{t-1}\right] < 1. \tag{9}$$

The invertibility condition has the form,

$$\begin{aligned}\mathsf{E}\left[\left|\varpi_{t-1} + \frac{\partial \varpi_{t-1}}{\partial \epsilon_{t-1}}\epsilon_{t-1}\right| |\mathcal{F}_{t-1}\right] &\leq \mathsf{E}_{t-1}|\theta_1 + \theta_2 g(.)| \\ &+\mathsf{E}_{t-1}|\theta_2\gamma(1-g(.))g(.)\epsilon_{t-1}[(\epsilon_{t-1}-c_1)+(\epsilon_{t-1}-c_2)]|\end{aligned} \tag{10}$$

The first term on the right-hand side of (10) is less than one if $|\theta_1 + \theta_2| < 1$. The second term is not necessarily zero since its value depends on γ and consequently large values of γ might affect invertibility. Fortunately, when γ is large $g(.)$ tends to a step function taking values zero and one. Hence, for large γ the second term on the right-hand side of (10) is practically zero and equals zeros for $\gamma \to \infty$.

3 Inference in Smooth Permanent Surge models

This section presents the statistical properties of the SPS model defined by (5), (7) and (8). The random variable of interest is Δy_t and not y_t, which means that the inference is based on the stationary variable Δy_t.

3.1 Hypothesis testing

In this subsection we present three tests in the SPS framework. The first test is a test of linearity in (5). It is based on Brännäs, De Gooijer, and Teräsvirta (1998). The second test is a test of the random walk hypothesis. The final test is a test of the SPS model against the STOPBREAK alternative.

3.1.1 Testing linearity in the SPS model

Testing linearity in (5) is equivalent to testing the hypothesis $\varpi_{t-1} = \theta^*$ for all t. Given that transition function (8) is constant when $\gamma = 0$, the linearity test can be carried out by the null hypothesis $\mathrm{H}_0 : \gamma = 0$. However, the standard testing procedures are not valid because θ_2, c_1 and c_2 are not identified parameters under the null hypothesis. Brännäs, De Gooijer, and Teräsvirta (1998) circumvent this identification problem following Luukkonen, Saikkonen, and Teräsvirta (1988) and Granger and Teräsvirta (1993). They replace the transition function (8) in (7) with its first-order Taylor expansion around $\gamma = 0$. After doing that and merging terms it turns out that the null hypothesis $\mathrm{H}_0 : \gamma = 0$ in (5) is equivalent to $\mathrm{H}_0^1 : \tilde{\theta}_2 = \tilde{\theta}_3 = 0$ in the following auxiliary regression:

$$\Delta y_t = \tilde{\theta}_1 \epsilon_{t-1} + \tilde{\theta}_2 \epsilon_{t-1}^2 + \tilde{\theta}_3 \epsilon_{t-1}^3 + e_t^* \tag{11}$$

where $e_t^* = \epsilon_t + \epsilon_{t-1} R(\gamma, c; \epsilon_{t-1})$. R is the remainder in the Taylor expansion. Note that under H_0 $e_t^* = \epsilon_t$ so the asymptotic theory is not affected by this approximation.

The LM test statistic is a convenient statistic for testing H_0^1 since it only requires the estimation of a MA(1) process. The resulting LM-type test can be carried out in three steps as follows:

1. Estimate the MA(1) model
$$\Delta y_t = \epsilon_t + \tilde{\theta}_1 \epsilon_{t-1}$$
and compute the residuals $\hat{\epsilon}_t$, $t = 1, \ldots, T$, and the sum of squared residuals SSR_0.

2. Regress $\hat{\epsilon}_t$ on $(\frac{\partial \epsilon_t}{\partial \tilde{\theta}_1}, \frac{\partial \epsilon_t}{\partial \tilde{\theta}_2}, \frac{\partial \epsilon_t}{\partial \tilde{\theta}_3})|_{H_0^1}$ and compute the sum of squares residuals SSR_1. From (11) it is seen that the first derivatives of the residuals ϵ_t with respect to $\tilde{\theta}_j$, $j = 1, 2, 3$,

under the null hypothesis are of the form

$$\frac{\partial \epsilon_{t-1}}{\partial \tilde{\theta}_j} = \hat{\epsilon}_{t-1} + \hat{\tilde{\theta}}_1 \frac{\partial \epsilon_{t-1}}{\partial \tilde{\theta}_j}$$

and thus have to be computed recursively.

3. The test statistic is,

$$LM = T\frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)}{\mathrm{SSR}_0} \tag{12}$$

and has an asymptotic χ^2 distribution with two degrees of freedom under the linearity hypothesis and the assumption $\mathsf{E}\epsilon_t^6 < \infty$.

3.1.2 Testing SPS against random walk

The random walk hypothesis is an interesting one to test, because the behaviour of y_t in the SPS model resembles the behaviour of realizations of the random walk process, and distinguishing between the two is important in applications. In fact, the SPS model can be defined as in (5) which is a unit root process with a specific moving average component.

The random walk hypothesis in (5) implies $\mathrm{H}_0^0 : \varpi_{t-1} = 0$ for all t. This null hypothesis is then equivalent to testing $\mathrm{H}_0^1 : \theta_1 = \theta_2 = 0$ in (7). The testing problem is again a nonstandard one, because the parameters γ, c are not identified under the null hypothesis. To circumvent the identification problem we follow Davies (1977,1987) and first derive the LM test of $\theta_1 = \theta_2 = 0$ in (5) assuming γ and c known. Based in the results in Andrews and Ploberger (1994), the identification problem is solved by applying ExpLM or AveLM tests.

The LM statistic for any given (γ, c) has the form:

$$LM(\gamma, c) = \frac{1}{\hat{\sigma}^2}\hat{u}' X_1(\gamma, c)(X_1(\gamma, c)' X_1(\gamma, c))^{-1} X_1(\gamma, c)'\hat{u} \tag{13}$$

where $X_1(\gamma, c) = [\hat{u}_{-1} \vdots G(\hat{u}_{-1}, \gamma, c) \odot \hat{u}_{-1}]$, $\hat{u}$ are the vector of residuals under the null and and $\hat{u}_{-1}$ its first lag, respectively.

The computation of ExpLM and AveLM tests can be based on a dense grid over γ and c. The grid should include possibly large positive values of γ and values of c defined within the range of $\hat{u}$.

In this paper we do not derive the asymptotic distribution of the test because it is possible to use the small sample distribution. We follow Dufour (1995) and Dufour and Khalaf (2001) and approximate the small sample distribution using Monte Carlo testing techniques. The advantage of this approach is that the test is exact in the sense that it has size-corrected critical regions. The main requirement of the MC test is that the statistic can be simulated under the null hypothesis. Moreover, the test is provably exact when the null distribution is free from nuisance parameters. Statistic (13) has this property because when $\mathrm{H}_0 : \theta_1 = \theta_2 = 0$ is valid, we have $\Delta y_t = \epsilon_t$, and consequently all that is needed for simulating Δy_t is the distribution of ϵ_t. Hence, the null distribution of the ExpLM and AveLM tests only depends on the distribution of the errors.

Following Dufour (1995) the small sample distribution of ExpLM and AveLM test can be obtained by simulation as follows:

1. Compute ExpLM or AveLM test from the original sample and call the statistic S_0.

2. Generate the LM test (13) by replacing $\hat{u}/\hat{\sigma}$ in (13) with $\hat{u}_s = u_s$ where u_s is a draw from the assumed error distribution. Compute the test statistic S_j from the simulated sample. Notice that for simulating the LM-test statistic under H_0, no knowledge of any parameters is needed.

   The number of replications N is typically small but it has to be such that $\alpha(N+1)$ is an integer for a given nominal size α. For example, for $\alpha = 0.05$, $N = 19$ is enough for correcting the size. Greater values of N increase the power of the test.

3. Compute the Monte Carlo p-value (P_{MC}) as

$$\mathrm{P}_{MC} = \frac{N\hat{G}_N(x) + 1}{N + 1} \tag{14}$$

   where $\hat{G}_N(S_0) = \frac{1}{N}\sum_{j=1}^{N} I_{[0,\infty)}(S_j - S_0)$ and $I_A(z) = 1$ for $z \in A$ and 0 otherwise.

The random walk hypothesis can also be tested with the SupLM test. The computation of the SupLM test is difficult because (13) is a highly erratic function of (γ, c). Despite this feature of the objective function it can be computed using a suitable global optimization procedure such as *simulated annealing* [See Brooks and Morgan (1995) and Goffe, Ferrier, and Rogers (1994) for details]. The advantage of this algorithm compared to numerical optimization algorithms based on derivatives of the objective function, is that it escapes local optima. The results in González and Teräsvirta (2004) indicate, however, that the ExpLM and AveLM tests have higher power than the SupLM and that they require fewer computations.

3.1.3 Testing STOPBREAK hypothesis within SPS

Even though the STOPBREAK model is not nested in the SPS process, there is a parametrization within the SPS that resembles the characterization of permanent effects in the STOPBREAK process. In the latter model the permanent effect of an innovation λ_t is a random variable defined within [0,2). However, the authors point out the following: *the intuition suggests that the majority of the probability mass for λ_t would lie in the [0,1] interval.* This suggest that even though the STOPBREAK model is invertible for $\lambda_t < 2$ in practice unity serves as upper bound for λ_t.

Using the fact that λ_t is defined on the [0,1] interval and the fact that transition function (8) approximates the Logistic function by (6) when $c_1 = c_2 = 0$ one can write an approximate STOPBREAK process as follows:

$$\begin{aligned} \Delta y_t =& \epsilon_{t-1} - \varpi_{t-1}\epsilon_{t-1} \\ \varpi_{t-1} =& -1 + 2g(\epsilon_{t-1}, \gamma) \\ g(\epsilon_{t-1}, \gamma) =& \left(1 + \exp(-\gamma\epsilon_{t-1}^2)\right)^{-1}. \end{aligned} \quad (15)$$

In (15), the *permanent effect of an innovation* λ_t equals

$$\lambda_t = -1 + 2g(\epsilon_{t-1}, \gamma) + 2(1 - g(\epsilon_{t-1}, \gamma))g(\epsilon_{t-1}, \gamma)\gamma\epsilon_{t-1} \quad (16)$$

The permanent effect of an innovation in (15) lies in the interval [0,1]. In fact, when $\epsilon = 0$, $\lambda_t = 0$ and when $|\epsilon| \to \infty$ we have that $g(.) = 1$ and $\lambda_t = 1$. This means that zero shocks have transitory effects whereas large shocks have permanent effects.

The second term on the right-hand side of (16) depends on γ. This might imply that for any ϵ, large values of γ are associated with a large permanent effect. However, when $\gamma \to \infty$, this second term is always zero because the transition function then becomes a step function taking only values zero and one.

This characterization of the STOPBREAK process within the SPS model allows us to formulate the STOPBREAK null hypothesis as $H_0 : c_1 = c_2 = 0;\ \theta_1 = -1;\ \theta_2 = 2$ in (5) with (7). This hypothesis is not testable because the elements of the score for c_1 and c_2 under the null are the same. However, reformulating the transition function (8) as in (15) yields

$$g(\epsilon, \gamma, c) = (1 + \exp(-\gamma(\epsilon_{t-1} - c)^2)^{-1}. \quad (17)$$

Using this formulation it is possible to test the equivalent null hypothesis $H_0 : c = 0$, $\theta_1 = -1$, $\theta_2 = 2$.

Since only γ has to be estimated under the null hypothesis, the LM test is computationally convenient. The test can be obtained in three steps as follows:

1. Estimate (5) under the null hypothesis, compute the residuals $\hat{u}$ and the sum of square residuals SSR_0.

2. Estimate the auxiliary regression

$$\hat{u}_t = x_t'\mathrm{b} + \mathrm{error}$$

and compute the sum of squared residuals SSR_1. Where, $x_t = (\frac{\partial \epsilon_t}{\partial \theta_1}, \frac{\partial \epsilon_t}{\partial \theta_2}, \frac{\partial \epsilon_t}{\partial c}, \frac{\partial \epsilon_t}{\partial \gamma})'$.

3. Compute the value of the LM statistic

$$LM = T\frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)}{\mathrm{SSR}_0}$$

Since the parameter estimates in the SPS model are asymptotically normally distributed, the LM test statistic has an asymptotic χ^2 distribution with three degrees of freedom under the null hypothesis.

The first step in this procedure requires nonlinear estimation of γ. Consequently, the estimated residuals and $\partial\epsilon_t/\partial\gamma$ are not necessarily orthogonal, which can affect the size of test. To circumvent this problem Eitrheim and Teräsvirta (1996) ortogonalized the estimated residuals under H_0 with respect to $\partial\epsilon_t/\partial\gamma$ before computing SSR_1 in step two. The second step in the above algorithm can thus be replaced by the following two steps:

2a. Regress $\hat{u}_t$ on $\frac{\partial \epsilon_t}{\partial \gamma}$ and compute the residuals $\hat{u}_t^o$ and SSR_0.

2b. Estimate the auxiliary regression

$$\hat{u}_t^o = x_t'\mathrm{b} + \mathrm{error}$$

and compute the sum of squared residuals SSR_1.

3.2 Estimating SPS models

Invertibility is required for the estimation of the SPS model, and the parameter vector $\varphi = (\theta_1, \theta_2, \gamma, c_1, c_2, \sigma^2)'$ can be estimated by maximum likelihood. The invertibility condition ensures that the likelihood function is well-defined since ϵ_t can be obtained recursively from any initial condition. For practical purposes, we assume that $\epsilon_0 = 0$.

The log likelihood function for an SPS model (5) is,

$$L(y, \varphi_1, \sigma^2) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{T}\epsilon_t^2 \tag{18}$$

where $\epsilon_t = \Delta y_t + \varpi_{t-1}\epsilon_{t-1}$ and ϖ_{t-1} is defined in (7) and (8).

The score vector is given by

$$\frac{\partial L}{\partial \beta} = -\frac{1}{\sigma^2}\sum_{t=1}^{T}\epsilon_t w_t$$
$$\frac{\partial L}{\partial \sigma^2} = \frac{1}{\sigma^3}\sum_{t=1}^{T}(\epsilon_t^2 - \sigma^2)$$

where $\beta = (\theta_1, \theta_2, \gamma, c)'$ and

$$w_t = b_{t-1}\epsilon_{t-1} + \left(\theta_2 \frac{\partial g_{t-1}(\gamma, \mathbf{c})}{\partial \epsilon}\epsilon_{t-1} + \varpi_{t-1}\right) w_{t-1}$$

and $b_{t-1} = \left(1, g_{t-1}(\gamma, \mathbf{c}), \theta_2 \frac{\partial g_{t-1}(\gamma,\mathbf{c})}{\partial \gamma}, \theta_2 \frac{\partial g_{t-1}(\gamma,\mathbf{c})}{\partial c_1}, \theta_2 \frac{\partial g_{t-1}(\gamma,\mathbf{c})}{\partial c_2}\right)'$.

Furthermore, $g_{t-1}(\gamma, \mathbf{c}) = g(\epsilon_{t-1}; \gamma, \mathbf{c})$ and

$$\frac{\partial g_{t-1}}{\partial \gamma} = [1 - g_{t-1}(\gamma, \mathbf{c})]\, g_{t-1}(\gamma, \mathbf{c})(\epsilon_{t-1} - c_1)(\epsilon_{t-1} - c_2)$$
$$\frac{\partial g_{t-1}}{\partial c_1} = -\,[1 - g(\epsilon_{t-1})]\, g_{t-1}(\gamma, \mathbf{c})\gamma(\epsilon_{t-1} - c_2)$$
$$\frac{\partial g_{t-1}}{\partial c_2} = -\,[1 - g_{t-1}(\gamma, \mathbf{c})]\, g_{t-1}(\gamma, \mathbf{c})\gamma(\epsilon_{t-1} - c_1)$$
$$\frac{\partial g_{t-1}}{\partial \epsilon} = [1 - g_{t-1}(\gamma, \mathbf{c})]\, g_{t-1}(\gamma, \mathbf{c})\gamma\left[(\epsilon_{t-1} - c_1) + (\epsilon_{t-1} - c_2)\right].$$

The recursion for computing the likelihood and the score can be started from zero. This starting-value should not have any effect on the results.

The following theorem establishes the consistency and asymptotic normality of the maximum likelihood estimator of φ.

Theorem 2 *Suppose that y_t is generated by an invertible SPS model, where $\{\epsilon_t, \mathcal{F}_t\}$ is a strictly stationary α-mixing martingale sequence. Then the maximum likelihood estimator $\hat{\varphi} \equiv \mathsf{argmax}_{\Psi} L_T(y^T, \varphi)$ is consistent under the following assumptions: (i) $\mathsf{E}|\epsilon_t|^{2p} \leq M < \infty$ for some $p > 1$, (ii) the parameter space Ψ is a compact subset of $\mathbb{R}^5$, and (iii) $\varphi_0 = \mathsf{argmax}_{\Psi} \mathsf{E} L_T(y^T, \varphi)$ is unique. Moreover, if in addition of (ii) and (iii) the α-mixing coefficients on ϵ_t are of size $p/(1-p)$ and $\mathsf{E}|\epsilon_t|^{4p} \leq M < \infty$, then*

$$\sqrt{T}(\hat{\varphi} - \varphi_0) \xrightarrow{d} N(0, V_0^{-1}) \tag{19}$$

Proof. The proof of the theorem closely follows Engle and Smith (1999) and can be found in the Appendix A. ■

Even though the parameter estimates are asymptotically normal this result must be applied with caution. In particular, one must be aware of the identification problem involved in testing several null hypotheses. For instance, the standard t-test for θ_2 and γ creates a situation in which the model contains nuisance parameters not identified under the null hypothesis. Consequently, the standard asymptotic distribution of this test is not applicable. For this reason, we recommend that the final estimated model be considered together with the results of the linearity tests.

The second factor that might affect the usefulness of the asymptotic distribution theory is that for very large values of γ the final estimated Hessian may be ill-conditioned. In this situation it is only possible to obtain standard errors for θ_1 and θ_2 by using the corresponding block of the final estimated Hessian.

4 Small-sample properties of tests

In this section we investigate the empirical size and power of our tests by simulation. The section is divided in three subsections. The first one is devoted to the linearity test, the second subsection contains results on the small-sample properties of the test of the random walk hypothesis. The final subsection is concerned with the test of the SPS process against the STOPBREAK model. All results are based on 5000 Monte Carlo replications. The data for each experiment is generated using the following SPS model:

$$\begin{aligned}\Delta y_t =& \epsilon_t + \varpi_{t-1}\epsilon_{t-1} \\ \varpi_{t-1} =& \theta_1 + \theta_2 g(\epsilon_{t-1}, \gamma, c) \\ g(\epsilon_{t-1}, \gamma, c) =& \Big(1 + \exp(-\gamma(\epsilon_{t-1} - c_1)(\epsilon_{t-1} - c_2))\Big)^{-1}.\end{aligned} \tag{20}$$

4.1 Small-sample properties of the linearity test

In order to investigate the size of the linearity test the data is generated using equation (20) with $\theta_2 = \gamma = c_1 = c_2 = 0$. Since only θ_1 is likely to affect the size of the test we consider the test for $\theta_1 = 0.1, 0.2, \ldots, 0.9$. The sample size is 100. The errors ϵ_t are independently normally distributed with mean zero and variance one. The results for the nominal size 0.05 are presented in Table 1.

It is seen that the size properties of the test are generally very good in the sense that the empirical size is close to the nominal size. The size deteriorates somewhat for values of θ_1 close to one. A likely explanation is that for θ_1 near one the null model is close to being noninvertible.

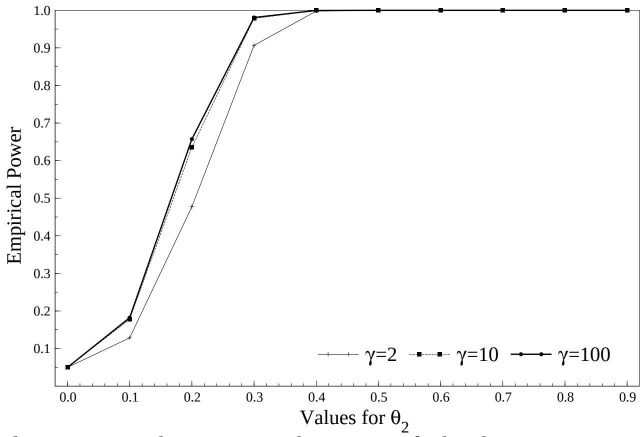
In order to investigate the power of the test we generate data from equation (20) with $\theta_1 = 0.5$ and different values for θ_2 and γ. We set $\theta_1 = 0.5$ because the empirical size was very close to its nominal size at this value of θ_1. Additionally, we use 10 different values for θ_2 and

Table 1: Empirical size of the linearity test

θ_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Size	0.048	0.049	0.049	0.056	0.050	0.054	0.060	0.060	0.085

The table contains the empirical size of the linearity test at 5% nominal level. The sample size is $n = 100$.

Figure 3: Empirical power of the linearity test at nominal significance level 0.05 as a function of θ_2 and γ.



The graph contains the empirical power of the linearity test for different values of γ and θ_2. Each line corresponds to one value of γ. The sample size equals 100 and $\theta_1 = 0.5$.

for each value of θ_2 we compute the power of the test for: γ, $\gamma = 2, 10$ and 100. Figure 3 shows the empirical power of the test as a function of θ_2. Each curve in the graph corresponds to one value of γ. As can be seen, the power increases with θ_2. Moreover, for each θ_2 the power is higher for larger values of γ. By comparing the estimated powers for $\gamma = 10$ and $\gamma = 100$ in Figure 3, it is seen that both curves are close to each other for all θ_2. This might suggest that the increment in power associated with γ decreases with the value of γ.

4.2 Small sample properties of the test of SPS against random walk

In order to obtain the empirical size of the test we simulate data from (20) with $\theta_1 = \theta_2 = 0$. ϵ_t are again drawn from a standard normal distribution. We set the nominal size to equal 0.05 and consequently only use 19 Monte Carlo replications for computing the Monte Carlo p-value. The sample size is 50. In order to compute ExpLM and AveLM test we conduct a grid search on γ and c. The grid for γ includes 100 evenly space values within 2 and 1000 and the grid for c includes 50 different values for c_1 and c_2. Since c_1 and c_2 are exchangeable in the likelihood, we only consider values of c_2 such that $c_2 \geq c_1$. Thus, in computing the ExpLM and AveLM test we evaluate the LM statistic (13) $(100 \times (50+1) \times 50)/2$ times for both the original sample and the simulated samples. The result is that the empirical size of both tests, the ExpLM and

AveLM, equals 0.0466 which is very close to the nominal size.

For computing the power of the test we simulate data from 18 different variants of model (20). In particular, we let θ_1 and θ_2 take values 0.1 and 0.3. Additionally, we also consider two values for γ; $\gamma = 2$ and 10. Since the power of the Monte Carlo tests depends on the number of Monte Carlo samples, N, we use $N = 19$ and $N = 59$ in computing the MC-p-value. The results are summarized in Table 2. They indicate that both the ExpLM and the AveLM

Table 2: Empirical power of the test of random walk hypothesis against SPS at the significance level 0.05

N=19									
γ	θ_1	θ_2	ExpLM	AveLM	γ	θ_1	θ_2	ExpLM	AveLM
		0.1	0.126	0.126			0.1	0.128	0.128
2	0.1	0.3	0.301	0.301	10	0.1	0.3	0.322	0.322
		0.5	0.542	0.542			0.5	0.575	0.575
		0.1	0.429	0.429			0.1	0.432	0.432
2	0.3	0.3	0.626	0.626	10	0.3	0.3	0.636	0.636
		0.5	0.774	0.774			0.5	0.786	0.786
		0.1	0.729	0.729			0.1	0.730	0.730
2	0.5	0.3	0.817	0.817	10	0.5	0.3	0.819	0.819
		0.5	0.877	0.877			0.5	0.881	0.881
N=59									
γ	θ_1	θ_2	ExpLM	AveLM	γ	θ_1	θ_2	ExpLM	AveLM
		0.1	0.137	0.137			0.1	0.141	0.141
2	0.1	0.3	0.325	0.325	10	0.1	0.3	0.343	0.343
		0.5	0.597	0.597			0.5	0.627	0.627
		0.1	0.475	0.475			0.1	0.478	0.478
2	0.3	0.3	0.678	0.678	10	0.3	0.3	0.689	0.689
		0.5	0.833	0.833			0.5	0.844	0.844
		0.1	0.789	0.789			0.1	0.788	0.788
2	0.5	0.3	0.873	0.873	10	0.5	0.3	0.875	0.875
		0.5	0.921	0.921			0.5	0.924	0.924

The table contains the empirical power of the test of SPS against random walk. The power is computed for the 5% nominal level. The sample size is $n = 50$. Two different number of Monte Carlo samples were used in computing the Monte Carlo p-values, $N = 19$ and $N = 59$.

tests have equal power in small samples. As expected, the power of the test increases with the number of Monte Carlo samples used to compute P_{MC}. The power differences when N is varied are not large. For instance, when $\gamma = 2$, $\theta_1 = 0.1$ and $\theta_2 = 0.3$ the power of the test with $N = 19$ equals 0.30 whereas for $N = 59$ it is 0.32. The power seems to be independent of γ and positively related with θ_1 and θ_2. Finally, considering θ_1 and θ_2 the results indicate that the power of the test depends more on θ_1 than on θ_2. For instance, with $\gamma = 2$, $\theta_2 = 0.1$ an increment of θ_1 from 0.1 to 0.3 increases the power in 2.46%, whereas the same exercise for $\theta_1 = 0.1$ with θ_2 changing from 0.1 to 0.3 only increases the power in 1.37%.

These results indicate that when the null hypothesis is rejected it is useful to test whether θ_2 equals zero. This null hypothesis can be tested using the linearity test proposed in section

3.1 or as suggested in González and Teräsvirta (2004).

4.3 Small sample properties of the test of SPS against the STOPBREAK model

To compute the size of the test against the STOPBREAK we generate data from (20) with $\theta_1 = -1$ and $\theta_2 = 2$, $c_1 = c_2 = 0$. Since the only free parameter under H_0 is γ we use three different values for it: $\gamma = 2, 10$ and 30. We also consider two sample sizes $n = 100$ and $n = 200$. Table 3 with four columns contains the results for the Monte Carlo experiment. The first and

Table 3: Empirical size of the test STOPBREAK against SPS at significance level 0.05

Sample Size	γ	Empirical size	Failures
	2	0.0454	5
100	5	0.0918	238
	30	-	>1000
Sample Size	γ	Empirical size	Failures
	2	0.0528	0
200	5	0.0814	59
	30	-	>1000

second columns of the table indicate the sample size and the value of γ, respectively. The third column contains the estimated size of the test and the last column indicates the number of discarded draws in the Monte Carlo experiment. These draws are samples for which the test was not available because the covariance matrix of the score was not invertible. The size of the test is good for small values of γ but deteriorates when γ increases. Moreover, the empirical size of the test cannot be computed for $\gamma = 30$. There are two explanations for this outcome. First, when $\gamma \to \infty$ the Hessian becomes noninvertible because the transition function under the null equals one for all ϵ_t different from zero. Second, large values of γ in (15) push the model under the null hypothesis towards the boundary of the invertibility condition, because the transition function in (9) always equals one.

In summary, the results of the Monte Carlo experiment suggest that the test is only available for small values of γ because the model is not identified under the null hypothesis and because it may not be invertible at large values of γ. The identification problem is not only present in the STOPBREAK approximation to SPS. It is also present in the original version of the STOPBREAK model. In fact, when γ in the (6) is close to zero, the logistic function takes value one for practically all $\epsilon_t \neq 0$ and the STOPBREAK model collapses to a random walk. The results of this Monte Carlo experiment support "second-order logistic" function (8) as an alternative parametrization to the simple logistic transition function used by Engle and Smith (1999).

Given the size results we only consider the power of the test for $\gamma = 2$ and we generate data from equation (20) with the following other parameter values: $c_1 = -0.1$, $c_2 = 0.1$, $\theta_1 = -0.8$ and $\theta_2 = 1.8$. We estimate the power of the test for two sample sizes $T = 100$ and $T = 200$. The results in Table 4. show that the test has good power against the alternative and that

Table 4: Empirical power of the test of SPS against STOPBREAK at significance level 0.05.

Sample Size	Empirical Power	Sample Size	Empirical Power
100	0.226	200	0.404

The table contains the empirical power of the SPS against STOPBREAK. The power is computed for a 0.05 nominal level

the power increases with the sample size. However, due to the fact that the size of the test deteriorates with the value of γ and that the test is not likely to be available at large values γ, we recommend caution when using it in applications.

5 Application

In this section we illustrate the use of the proposed test statistics and the SPS methodology. The application is borrowed from Engle and Smith (1999). We investigate whether the stock prices of companies that belong to the same market have the tendency to move together. In theory, such prices should move together if they follow the industry behaviour, and they might deviate from each other only temporarily and depending on industry-specific shocks. This theory therefore implies that the ratio of these stock prices should not follow a random walk.

We apply the SPS model to daily price series for Texaco, Mobil, IBM, Microsoft, General Motors, Ford, Coca-Cola and Pepsi. In all cases, stock prices are measured as the closing price of stocks listed in the US market. The observations cover a period from January 1988 to March 2004 and were obtained from Reuters 3000 Xtra.

The random walk test was computed using a grid search over γ and c. We include 100 different values for γ evenly spaced between 2 and 1000 and 25 values for c_1 and c_2 defined between -2.5 and 2.5. The Monte Carlo p-value was computed using 59 samples from a standard normal distribution. In order to compare our results with those of Engle and Smith (1999) we computed the random walk test for two different samples. The first sample comprises observations from January 1988 through December 1995, which is the sample used in Engle and Smith (1999), whereas the second sample contains observations from January 1988 to March 2004. The results are reported in Table 5. The table is divided into two panels, one for each sample. Each panel presents the results of the standard LM test and its robustified version, LM_R. The LM_R is robustified against heteroskedasticity by using the HAC estimator of the variance-covariance matrix. We prefer the results of the LM_R because the time series are highly heteroskcedastic. From the first panel of Table 5 it can be seen that the random walk hypothesis is not rejected at 5% significance level for the period 1988-1995 for the stock price ratios IBM/MSFT and Texaco/Mobil whereas it is rejected for Cola-Cola/Pepsi and General Motors/Ford. Using the complete sample size the results change. In fact, the random walk hypothesis is only rejected for the price ratio General Motors/Ford.

These results have to be interpreted with caution because the null hypothesis of random walk is the joint hypothesis $\theta_1 = \theta_2 = 0$ in (5). Thus it is possible that when the null hypothesis is not rejected the data follows a linear MA(1) process and not necessarily a random walk. It is therefore important to complement the results of the random walk test with those of the linearity test. These results can be found in Table 6. As before, the table is divided into two

Table 5: Random Walk test results

	(1988:1-1995:12)		(1988:1-2004:3)	
Ratio	Ordinary	Robust	Ordinary	Robust
IBM/MSFT	1.4(0.383)	0.7(0.733)	7.1(0.017)	1.9(0.167)
GM/FORD	6.7(0.034)	4.5(0.017)	33.1(0.017)	20.9(0.017)
COLA/PEPSI	5.1(0.017)	4.5(0.017)	4.4(0.033)	2.1(0.183)
Texaco/Mobil	6.6(0.017)	1.1(0.500)	6.0(0.017)	2.7(0.120)

Note: The table reports the results based on ExpLM test. P-values in parenthesis. The Monte Carlo p-values were computed using 59 samples from the normal distribution.

panels. The left-hand panel contains the results for the period 1988-1995 whereas the results for the complete sample are reported in the right-hand panel. Linearity is generally not rejected, the only exception being the Texaco/Mobil ratio in the sample 1988-1995.

Taking into account the results of both tests, the random walk and the linearity test it is seen that, the stock price ratios Coca-Cola/Pepsi and General Motors/Ford can be characterized by a linear stationary MA(1) while the ratio IBM/MSFT seems to follow a random walk. No evidence in favour of smooth permanent surge is found in any of the stock price ratios. There is no need to fit an SPS model to these price ratio series.

Table 6: Linearity test results

	(1988:1-1995:12)		(1988:1-2004:3)	
Ratio	LM	LM_R	LM	LM_R
IBM/MSFT	13.2(0.001)	5.8(0.344)	17.3(0.001)	2.1(0.344)
GM/FORD	1.4(0.484)	1.1(0.581)	0.3(0.838)	0.7(0.716)
COLA/PEPSI	0.1(0.975)	0.1(0.981)	10.0(0.001)	3.6(0.166)
Texaco/Mobil	19.1(0.001)	6.9(0.037)	10.2(0.001)	1.2(0.540)

Note: The column label LM_R contains the results of the robust LM test. The tests are robustified using the standard HAC estimator for the variance.

6 Conclusions

In this paper we have introduced the Smooth Permanent Surge mode which generalizes the STOPBREAK model by Engle and Smith (1999). The new parametrization overcomes a difficulty inherent in STOPBREAK model, namely, that all shocks to the model have permanent effect. In SPS models there is the possibility for shocks to have transitory effects.

The SPS model is also an alternative to STIMA models because, it allows for asymmetries in the long-run effect of the shocks. The continuity of the likelihood function permits the use of standard asymptotic theory when carrying out inference on the model parameters.

We describe three tests of the SPS model that can be used in the modelling process. The first test is a test of non linearity, the second test is a test of SPS against random walk and the final test is a test of SPS against STOPREAK process. The results of the Monte Carlo experiment indicate that the first two tests have good properties in small samples while the last test seems to be of little practical use.

References

Andrews, D. W. K. (1988): "Law of large number for dependent non-identically distributed random variables," *Econometric Theory*, 4, 458–467.

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

Brännäs, K., J. G. De Gooijer, and T. Teräsvirta (1998): "Testing linearity against nonlinear moving average models," *Communications in Statistics - Theory and Methods*, 27, 2025–2035.

Brooks, S. P., and B. Morgan (1995): "Optimization using simulated annealing," *The Statistician*, 44, 241–257.

Davidson, J. (1994): *Stochastic Limit theory.* Oxford University Press, Oxford.

Davies, R. B. (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.

Dufour, J.-M. (1995): "Monte carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics," *Discussion paper, C.R.D.E. Université de Montréal.*

Dufour, J.-M., and L. Khalaf (2001): “Monte Carlo test methods in econometrics,” in *A companion to theoretical econometrics*, ed. by B. Baltagi, pp. 495–519. Blackwell.

Eitrheim, Ø., and T. Teräsvirta (1996): “Testing the adequacy of smooth transition autoregressive models,” *Journal of Econometrics*, 74, 59–75.

Engle, R. F., and A. D. Smith (1999): “Stochastic permanent breaks,” *Review of Economics and Statistics*, 81, 553–574.

Goffe, W., G. D. Ferrier, and J. Rogers (1994): “Global optimization of statistical functions with simulated annealing,” *Journal of Econometrics*, 60, 65–99.

González, A., and T. Teräsvirta (2004): “Simulation-based finite-sample linearity test against smooth transition regression models,” *Work in progress.*

Gonzalo, J., and O. Martínez (2003): “Threshold integrated moving average process (Does size matter? Maybe so),” *Preliminary draft.*

Granger, C. W. J., and A. P. Andersen (1978): “On invertibility of time series models,” *Stochastic processes and their applications*, 8, 87–92.

Granger, C. W. J., and N. R. Swanson (1997): “An introduction to stochastic unit-root process,” *Journal of Econometrics*, 80, 35–62.

Granger, C. W. J., and T. Teräsvirta (1993): *Modelling nonlinear economic relationships.* Oxford University Press, Oxford.

Jansen, E. S., and T. Teräsvirta (1996): “Testing parameter constancy and super exogeneity in econometric equations,” *Oxford Bulletin of Economics and Statistics*, 58, 735–763.

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

Rahbek, A., and N. Shephard (2002): “Inference and ergodicity in the autoregressive root model,” *Working paper.*

Wooldridge, J. M. (1994): “Estimation and inference for dependent processes,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2640–2738, New York. Elsevier Science B.V.

A Proof of theorem 2

This appendix contains the proof of Theorem 2 stating consistency and asymptotic normality of the maximum likelihood estimators of the parameters of the SPS model. The log-likelihood function is

$$L_T = \sum_{t=1}^{T} q_t(y^t, \varphi) \tag{1}$$

where $q_t(y^t, \varphi) = -\frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\Delta y_t + \varpi_{t-1}\epsilon_{t-1})^2$ and $\varphi = (\theta_1, \theta_2, \gamma, c_1, c_2, \sigma^2)'$. Furthermore, ϖ_{t-1} and $g(\epsilon_{t-1}, \gamma, c_1, c_2)$ are defined in (7) and (8), respectively.

A.1 Consistency

Repeated substitution of ϵ_{t-1} reveals that $q_t(y_t, \varphi)$ is a function of the increasing sequence $y_i, i = 1, \dots, t$, and consequently it is a heterogenous sequence of y_t. In order to prove consistency we show that conditions (M.1) - (M.3) of Theorem 4.3 in Wooldridge (1994) are satisfied. To invoke theorem 4.3 in Wooldridge (1994) we need to show that $\{q_t(\epsilon_t, \varphi) : t = 1, \dots\}$ satisfies the uniform law of large numbers on Ψ and that $q_t(y^t, \varphi)$ is measurable for any $\varphi \in \Psi$. For the uniform law of large numbers to hold we need to verify the following conditions:

1. For each $\tilde{\varphi} \in \varphi$, $\{q_t(y^t, \tilde{\varphi}) : t = 1, 2, \dots\}$ satisfies the weak law of large numbers.

2. There exists a function $h_t(y^t) \geq 0$ such that

   (a) For all φ_1,$\varphi_2 \in \Psi$, $|q_t(y^t, \varphi_1) - q_t(y^t, \varphi_2)| \leq h(y^t) \parallel \varphi_1 - \varphi_2 \parallel$

   (b) $\{h(y^t)\}$ satisfies the weak law of large numbers.

The strategy to prove condition 1 is the following. First we will show that for any $\tilde{\varphi} \in \varphi$, $\tilde{\epsilon}$ is L_2-NED, which by Theorem 17.9 in Davidson (1994) implies that $\tilde{\epsilon}^2$ is L_1-NED . We conclude from Theorem 17.5 in Davidson (1994) that $\{\tilde{\epsilon}^2 - \tilde{\sigma}^2\}$ is an L_1-mixingale. Finally, if $\tilde{\epsilon}^2$ is uniformly integrable, it follows from Andrews's (1988) weak law of large numbers that

$$\frac{1}{T}\sum_{t=1}^{T} q_t(\tilde{\epsilon}_t, \tilde{\varphi}) \xrightarrow{p} \mathsf{E}\frac{1}{T}\sum_{t=1}^{T} q_t(\tilde{\epsilon}_t, \tilde{\varphi})$$

for any $\tilde{\varphi} \in \Psi$.

For any $\tilde{\varphi}$ we have that $\tilde{\epsilon}_t = \Delta y_t + \tilde{\theta}_{t-1}\tilde{\epsilon}_{t-1}$ where $\tilde{\theta}_{t-1} = \tilde{\theta}_1 + \tilde{\theta}_2 g(\epsilon_{t-1}, \tilde{\gamma}, \tilde{c})$. Recursive substitution of $\tilde{\epsilon}_t$ shows that $\tilde{\epsilon}_t = f_t(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2,} \dots)$ which is a continuous function of a mixing sequence which is not necessarily mixing. However, $\tilde{\epsilon}$ is L_2-NED. To see this we follow Example 17.4 in Davidson (1994) and approximate f_t with a Taylor expansion about zero with respect to ϵ_{t-j} for $j > m$, where m is a fixed real number. This yields,

$$\tilde{f}_t = \tilde{f}_t^m + \sum_{j=m+1}^{t} \left(\frac{\partial f_t}{\partial \epsilon_{t-j}}\right)^* \epsilon_{t-j}$$

where $*$ denotes the evaluation of the derivatives at points in the interval $[0, \epsilon_{t-j}]$. Note that f_t^m is a measurable approximation of $\mathsf{E}\left[\tilde{f}_t|\mathcal{F}_{t-m}^{t+m}\right]$. Consequently, from Theorem 10.12 in Davidson (1994) we have

$$\left\|\tilde{\epsilon}_t - \mathsf{E}\left[\tilde{f}_t|\mathcal{F}_{t-m}^{t+m}\right]\right\|_2 \;\leq\; \|\tilde{\epsilon}_t - f_t^m\|_2 \,.$$

which implies

$$\left\|\tilde{\epsilon}_t - \mathsf{E}\left[\tilde{f}_t|\mathcal{F}_{t-m}^{t}\right]\right\|_2 \;\leq\; \left\|\sum_{j=m+1}^{t} \left(\frac{\partial \tilde{f}_t}{\partial \epsilon_{t-j}}\right)^* \epsilon_{t-j}\right\|_2$$

Differentiating $\tilde{f}_t$ w.r.t ϵ_{t-j} yields,

$$\begin{aligned}
\frac{\partial \tilde{f}_t}{\partial \epsilon_{t-j}} &= \frac{\partial \Delta y_t}{\partial \epsilon_{t-j}} + \left(\tilde{\theta}_{t-1} + \frac{\partial \tilde{\theta}_{t-1}}{\partial \tilde{\epsilon}_{t-j}}\tilde{\epsilon}_{t-1}\right)\frac{\partial \tilde{\epsilon}_{t-1}}{\partial \epsilon_{t-j}} \\
&= \tilde{k}_{t-1}\tilde{k}_{t-2}\dots\tilde{k}_{t-j+1}\frac{\partial \Delta y_{t-j+1}}{\partial \epsilon_{t-j}} + \tilde{k}_{t-1}\tilde{k}_{t-2}\dots\tilde{k}_{t-j}\frac{\partial \Delta y_{t-j}}{\partial \epsilon_{t-j}} \\
&= \tilde{k}_{t-1}\tilde{k}_{t-2}\dots\tilde{k}_{t-j+1}\left(\varpi_{t-1} + \frac{\partial \theta_{t-j}}{\partial \epsilon_{t-1}}\epsilon_{t-1}\right) + \tilde{k}_{t-1}\tilde{k}_{t-2}\dots\tilde{k}_{t-j}
\end{aligned}$$

where $\tilde{k}_{t-1} = \left(\tilde{\theta}_{t-1} + \frac{\partial \tilde{\theta}_{t-1}}{\partial \tilde{\epsilon}_{t-1}}\tilde{\epsilon}_{t-1}\right)$. Under invertibility $|\tilde{k}_{t-i}| \leq |\bar{k}| < 1$ and consequently

$$\left|\frac{\partial \tilde{f}_t}{\partial \epsilon_{t-j}}\right| \;\leq\; 2\bar{k}^j$$

which implies

$$\begin{aligned}
\left\|\tilde{\epsilon}_t - \mathsf{E}\left[\tilde{f}_t|\mathcal{F}_{t-m}^{t}\right]\right\|_2 &\leq \sum_{j=m+1}^{t}\left|\left(\frac{\partial \tilde{f}_t}{\partial \epsilon_{t-j}}\right)^*\right| \|\epsilon_{t-j}\|_2 \\
&< 2\sum_{j=m+1}^{\infty}\bar{k}^j \left\|\epsilon_{t-j}\right\|_2 = v_m d_t
\end{aligned}$$

where $v_m = 2\sum_{j=t+m}^{\infty}\bar{k}^j$ and $d_t = \|\epsilon\|_2$. Consequently, $\{\tilde{\epsilon}_t\}$ is L_2-NED. It follows from Theorem 17.9 in Davidson (1994) that $\{\tilde{\epsilon}_t^2\} = \{\tilde{\epsilon}_t\tilde{\epsilon}_t\}$ is L_1-NED and from Theorem 17.5 in Davidson (1994), that $\{\tilde{\epsilon}_t^2\}$ is L_1-mixing.

For Andrews's (1988) weak law of large numbers to apply we need $\mathsf{E}|\tilde{\epsilon}_t|^{2p} \leq M < \infty$ for $p > 1$. This result follows from

$$\begin{aligned} \|\tilde{\epsilon}_t\|_{2p} &= \|\Delta y_t + \tilde{\theta}_{t-1}\tilde{\epsilon}_{t-1}\|_{2p} \\ &\leq \|\Delta y_t\|_{2p} + \|\tilde{\theta}_{t-1}\tilde{\epsilon}_{t-1}\|_{2p} \\ &\leq \|\Delta y_t\|_{2p} + \bar{k}\|\tilde{\epsilon}_{t-1}\|_{2p} \end{aligned}$$

hence $\mathsf{E}\tilde{\epsilon}_t{}^{2p} \leq \left(\frac{1}{1-\bar{k}}\|\Delta y_t\|_{2p}\right)^{2p}$. Note that $\|\Delta y_t\|_{2p}$ exists since $\|\epsilon_t\| \leq M < \infty$. It follows from Andrews's (1988) weak law of large numbers that

$$\frac{1}{T}\sum_{t=1}^{T} q_t(y^t, \tilde{\varphi}) \rightarrow \mathsf{E}\frac{1}{T}\sum_{t=1}^{T} q_t(y^t, \tilde{\varphi})$$

for any $\tilde{\varphi} \in \Psi$.

In other to prove that Condition 2 also holds for SPS models we have to find a dominant function $h(y^t)$ such that,

1. For all φ_1, $\varphi_2 \in \Psi$, $|q_t(y^t, \varphi_1) - q_t(y^t, \varphi_2)| \leq h(y^t) \parallel \varphi_1 - \varphi_2 \parallel$.
2. $\{h(y^t)\}$ satisfies the weak law of large numbers.

The mean value approximation to the likelihood function around φ_2 is,

$$q_t(y^t, \varphi) - q_t(y^t, \varphi_2) = \frac{\partial q_t(y^t, \tilde{\varphi})}{\partial \varphi}' (\varphi - \varphi_2)$$

where $\tilde{\varphi}$ indicates that the derivative is evaluated at a point between φ and φ_2. Evaluating the mean value approximation at φ_1 and using the triangle inequality we obtain

$$\left|q_t(y^t, \varphi_1) - q_t(y^t, \varphi_2)\right| \leq \left\langle \frac{\partial q_t(y^t, \tilde{\varphi})}{\partial \varphi'} \right\rangle \|\varphi_1 - \varphi_2\|$$

where $\langle x \rangle$ denotes the Euclidean norm of the vector x.

We have that,

$$\begin{aligned} \left\langle \frac{\partial q_t(y^t, \varphi)}{\partial \varphi} \right\rangle &= \Big[\tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{1t}^2 + \tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{2t}^2 + \tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{3t}^2 \\ &\quad + \tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{4t}^2 + \tilde{\sigma}^{-4}\tilde{\epsilon}_t^2\tilde{w}_{5t}^2 + \tilde{\sigma}^{-6}(\tilde{\epsilon}_t^2 - \tilde{\sigma}_t^2)^2\Big]^{1/2} \\ &\leq \tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{1t}| + \tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{2t}| + \tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{3t}| \\ &\quad + \tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{4t}| + \tilde{\sigma}^{-2}|\tilde{\epsilon}_t\tilde{w}_{5t}| + \tilde{\sigma}^{-3}|(\tilde{\epsilon}_t^2 - \tilde{\sigma}_t^2)| \end{aligned} \tag{2}$$

where $\tilde{w}_{it}$ for $i = 1, \ldots, 5$, are the derivatives of $\tilde{\epsilon}_t$ with respect to φ:

$$\begin{aligned}
\tilde{w}_{1t} &= \frac{\partial \tilde{\epsilon}_t}{\partial \theta_1} = \tilde{\epsilon}_{t-1} + \tilde{k}_{t-1}\tilde{w}_{1t-1} \\
\tilde{w}_{2t} &= \frac{\partial \tilde{\epsilon}_t}{\partial \theta_2} = \tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}}) + \tilde{k}_{t-1}\tilde{w}_{2t-1} \\
\tilde{w}_{3t} &= \frac{\partial \tilde{\epsilon}_t}{\partial \gamma} = \tilde{\theta}_2 \frac{\partial \tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}})}{\partial \gamma}\epsilon_{t-1} + \tilde{k}_{t-1}\tilde{w}_{3t-1} \\
\tilde{w}_{4t} &= \frac{\partial \tilde{\epsilon}_t}{\partial c_1} = \tilde{\theta}_2 \frac{\partial \tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}})}{\partial c_1}\epsilon_{t-1} + \tilde{k}_{t-1}\tilde{w}_{4t-1} \\
\tilde{w}_{5t} &= \frac{\partial \tilde{\epsilon}_t}{\partial c_2} = \tilde{\theta}_2 \frac{\partial \tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}})}{\partial c_2}\epsilon_{t-1} + \tilde{k}_{t-1}\tilde{w}_{5t-1}
\end{aligned}$$

where $\tilde{\mathbf{c}} = (c_1, c_2)'$ and $\tilde{g}_{t-1}(.) = g(\tilde{\epsilon}_{t-1}, \tilde{\gamma}, \tilde{\mathbf{c}})$.

Consider

$$\begin{aligned}
|\tilde{\epsilon}_t| &\leq |\Delta y_t| + \bar{k}|\epsilon_{t-1}| \\
&\leq \sum_{i=1}^{t} \bar{k}^{i-1}|\Delta y_{t-i}| \leq (1-\bar{k})\sum_{i-1}^{t} \bar{k}^{i-1}|\epsilon_{t-i}|
\end{aligned}$$

and similarly,

$$\begin{aligned}
|\tilde{w_{1t}}| &\leq |\tilde{\epsilon}_t| + \bar{k}|\tilde{w}_{1t-1}| \\
&\leq \sum_{i=1}^{t} \bar{k}^{i-1}|\tilde{\epsilon}_{t-i}| \leq \sum_{i=1}^{t} \bar{k}^{i-1} \sum_{j=1}^{t-i} \bar{k}^{j-1}|\epsilon_{t-i-j}|
\end{aligned}$$

$$\begin{aligned}
|\tilde{w_{2t}}| &\leq |\tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}})| + \bar{k}|\tilde{w}_{2t-1}| \\
&\leq \sum_{i=1}^{t} \bar{k}^{i-1}|\tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}})| \leq \sum_{i=1}^{t} \bar{k}^{i-1}
\end{aligned}$$

$$\begin{aligned}
|\tilde{w}_{3t}| &\leq \left|\tilde{\theta}_2 \frac{\partial \tilde{g}_{t-1}(\tilde{\gamma}, \tilde{\mathbf{c}}}{\partial \gamma}\right| + \bar{k}|\tilde{w}_{3t-1}| \\
&\leq \sum_{i=1}^{t} \bar{k}^{i-1}\left|\tilde{\theta}_2 \frac{\partial \tilde{g}_{t-i}(\tilde{\gamma}, \tilde{\mathbf{c}})}{\partial \gamma}\right| \\
&\leq K_1 \sum_{i=1}^{t} \bar{k}^{i-1}\left|\frac{\partial \tilde{g}_{t-i}(\tilde{\gamma}, \tilde{\mathbf{c}})}{\partial \gamma}\right| < K(\tilde{\gamma}, \tilde{\mathbf{c}})\sum_{i=1}^{t} \bar{k}^{i-1} \qquad (3)
\end{aligned}$$

The last equality in (3) holds since $|\frac{\partial \tilde{g}_{t-i}(\tilde{\gamma},\tilde{\mathbf{c}})}{\partial \gamma}| < K(\tilde{\gamma},\tilde{\mathbf{c}}) < \infty$. Similarly, $|\tilde{w}_{4t}| \leq K_2(\tilde{\gamma},\tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1}$ and $|\tilde{w}_{5t}| \leq K_3(\tilde{\gamma},\tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1}$ with $K_i(\tilde{\gamma},\tilde{\mathbf{c}}) < \infty$ for $i = 1, 2, 3$. Thus (2) becomes

$$\begin{aligned}
\langle \frac{\partial q_t(y^t,\varphi)}{\partial \varphi} \rangle \leq \frac{1}{\tilde{\sigma}^2} \Bigg(& \sum_{i=1}^{t} \bar{k}^{i-1} \sum_{j=1}^{t-i} \bar{k}^{i-1} |\epsilon_{t-i-j}| + K_1(\tilde{\gamma},\tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1} \\
& + K_2(\tilde{\gamma},\tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1} + K_3(\tilde{\gamma},\tilde{\mathbf{c}}) \sum_{i=1}^{t} \bar{k}^{i-1} \Bigg) \Bigg(\sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| \Bigg) \\
& + \frac{1}{\tilde{\sigma}^3} \Big(\sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| \Big)^2 \leq h(y^t)
\end{aligned} \tag{4}$$

where $h(y^t) \equiv \sup_{\varphi \in \varphi} \langle \frac{\partial q_t(y^t,\varphi)}{\partial \varphi} \rangle$. Thus, $h(y^t) \equiv A \left(\sum_{i=1}^{t} \bar{k}^{i-1} \sum_{j=1}^{t-i} \bar{k}^{i-1} |\epsilon_{t-i-j}| + B \right) \times \sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| + \left(\sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| \right)^2$ where A and B are finite constants. It turns out that $h(y^t)$ is a function of $b_t = \sum_{i=1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}|$ which is L_2-NED in ϵ_t. To see this note that

$$\begin{aligned}
\|b_t - \mathsf{E}(b_t|\mathcal{F}_{t-m}^{t+m})\|_2 &= \left\| \sum_{i=m+1}^{t} \bar{k}^{i-1} |\epsilon_{t-i}| \right\|_2 \\
&\leq \sum_{i=m+1}^{t} \bar{k}^{i-1} \|\epsilon_{t-i}\|_2 \\
&< \sum_{i=m+1}^{\infty} \bar{k}^{i-1} \|\epsilon_{t-i}\|_2 = v_m d_t
\end{aligned} \tag{5}$$

From (5) it follows that b_t is L_2-NED with constants $d_t = \|\epsilon_t\|_2$ and $v_m = \sum_{i=m+1}^{\infty} \bar{k}^{i-1}$. It follows from Theorem 17.9 in Davidson (1994) that b_t^2 is L_1-NED. Moreover, since $\|\epsilon_t\|_{2p} < \infty$, b_t^2 is uniformly integrable. It follows that $h(y^t)$, which is a function of b_t^2, obeys a uniform weak law of large numbers. Consequently, $q(y^t,\varphi)$ satisfies the conditions of Theorem 4.3 in Wooldridge (1994) and we have that $\varphi \xrightarrow{p} \varphi_0$.

A.2 Asymptotic normality of the maximum likelihood estimator

Using the mean value expansion, the maximum likelihood estimator can be approximated as follows,

$$\begin{aligned}
\sqrt{T}(\hat{\varphi}-\varphi^0) &= -\left(\frac{\partial^2 L_T(y^T,\tilde{\varphi})}{\partial\varphi\partial\varphi'}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q(y^T,\varphi^0)}{\partial\varphi} \\
&= -H_0^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q(y^T,\varphi^0)}{\partial\varphi} \\
&+ \left[H_0^{-1}-\left(\frac{\partial^2 L_T(y^T,\tilde{\varphi})}{\partial\varphi\partial\varphi'}\right)^{-1}\right]\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q(y^T,\varphi^0)}{\partial\varphi} \\
&= -H_0^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q(y^T,\varphi^0)}{\partial\varphi}+o_p(1)
\end{aligned}$$

The last equality holds if the sample Hessian evaluated at $\tilde{\varphi}$ converges in probability to the population Hessian. To verify this, we need to show that the hessian obeys the uniform law of large numbers.

Writing the parameter vector as $\varphi = (\varphi_1, \sigma^2)'$ with $\varphi_1 = (\theta_1, \theta_2, \gamma, c_1, c_2)'$ the average Hessian can be written as

$$-\frac{\partial L_T(y^T,\tilde{\varphi})}{\partial\varphi\partial\varphi'} = \begin{bmatrix} \tilde{H}_1 & \tilde{H}_3 \\ \tilde{H}_3' & \tilde{H}_2 \end{bmatrix}$$

with

$$\begin{aligned}
\tilde{H}_1 &= \frac{1}{T}\sum_{t=1}^{T}h_{1t} = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}\left(\tilde{w}_t\tilde{w}_t' + \tilde{\epsilon}_t\frac{\partial\tilde{w}_t}{\partial\varphi_1}\right) \\
\tilde{H}_3 &= \frac{1}{T}\sum_{t=1}^{T}h_{3t} = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}\tilde{w}_t\tilde{\epsilon}_t \\
\tilde{H}_2 &= \frac{1}{T}\sum_{t=1}^{T}h_{2t} = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}\left(3\frac{\tilde{\epsilon}_t^2}{\tilde{\sigma}^2}-1\right)
\end{aligned}$$

where $\tilde{w}_t = (\frac{\partial\tilde{\epsilon}_t}{\partial\theta_1}, \frac{\partial\tilde{\epsilon}_t}{\partial\theta_2}, \frac{\partial\tilde{\epsilon}_t}{\partial\gamma}, \frac{\partial\tilde{\epsilon}_t}{\partial c_1}, \frac{\partial\tilde{\epsilon}_t}{\partial c_2})'$.

As in the proof of consistency, we show that $\tilde{H}_1$,$\tilde{H}_2$, $\tilde{H}_3$ obey the uniform law of large numbers. For this purpose, we shall show that the following conditions hold,

1. For each $\tilde{\varphi} \in \Psi$, $\{\tilde{h}_{it} : t = 1, 2, \dots\}$ for $i = 1, 2, 3$, satisfies the weak law of large numbers.
2. For each element $\tilde{h}_{it}^{(j)}$ of $\tilde{H}_i$ for $i = 1, 2, 3$ there exists a function $r(y^t) \geq 0$ such that
   (a) For all φ_1, $\varphi_2 \in \Psi$, $|h_{it}^{(j)}(\varphi_1) - h_{it}^{(j)}(\varphi_2)| \leq r(y^t)\|\varphi_1 - \varphi_2\|$
   (b) $\{r(y^t)\}$ satisfies the weak law of large numbers.

In order to show that condition 1 is satisfied we follow a similar strategy to the one used in the consistency proof. That is, first we show that $\tilde{w}_t$ and $\frac{\partial \tilde{w}_t}{\partial \varphi_1}$ are L_2-NED in ϵ_t. This implies that the summands h_{it} are L_1-NED in ϵ_t. It follows from Theorem 17.5 in Davidson (1994) that they are L_1-mixingales. Finally, the assumption $\|\epsilon_t\|_{4p} \leq M < \infty$ guarantees uniform integrability. Consequently, H_i for $i = 1, 2, 3$, satisfy Andrews's (1988) weak law of large numbers.

In order to show that the summands in H_i for $i = 1, 2, 3$ are L_1-NED in ϵ_t we show that $\tilde{w}_{it}$ for $i = 1, \ldots, 5$ are L_2-NED. In particular, we show that,

$$\|\tilde{w}_{1t} - \mathsf{E}[w_{1t}|\mathcal{F}_{t-m}^{t+m}]\|_2 \leq \sum_{j=m+1}^{\infty} \left\| \left(\frac{\partial \tilde{w}_{1t}}{\partial \epsilon_{t-j}} \right)^* \epsilon_{t-j} \right\|_2$$

we have that

$$\begin{aligned}
\left| \frac{\partial \tilde{w}_{1t}}{\partial \epsilon_{t-j}} \right| &\leq \bar{k} \left| \frac{\partial \tilde{w}_{1t-1}}{\partial \epsilon_{t-j}} \right| + \left(1 + \left| \frac{\partial \tilde{k}_{t-1}}{\partial \tilde{\epsilon}_{t-1}} \right| |\tilde{w}_{1t-1}| \right) \left| \frac{\partial \tilde{\epsilon}_{t-1}}{\partial \epsilon_{t-j}} \right| \\
&\leq \sum_{i=1}^{j} \bar{k}^{i-1} \left(1 + \left| \frac{\partial \tilde{k}_{t-i}}{\partial \tilde{\epsilon}_{t-i}} \right| |\tilde{w}_{1t-i}| \right) \left| \frac{\partial \tilde{\epsilon}_{t-i}}{\partial \epsilon_{t-j}} \right| \\
&\leq \bar{k}^{j-1} \sum_{i=1}^{j} \left(1 + \left| \frac{\partial \tilde{k}_{t-i}}{\partial \tilde{\epsilon}_{t-i}} \right| |\tilde{w}_{1t-i}| \right)
\end{aligned}$$

where,

$$\begin{aligned}
\left| \frac{\partial \tilde{k}_{t-i}}{\partial \tilde{\epsilon}_{t-i}} \right| &\leq \left| (1 + \tilde{\epsilon}_{t-i}) \frac{\partial \tilde{\theta}_{t-i}}{\partial \tilde{\epsilon}_{t-i}} + \frac{\partial^2 \tilde{\theta}_{t-i}}{\partial^2 \tilde{\epsilon}_{t-i}} \right| \\
&\leq \left| (1 + \tilde{\epsilon}_{t-i}) \tilde{\theta}_2 \frac{\partial \tilde{g}_{t-i}}{\partial \tilde{\epsilon}_{t-i}} + \tilde{\theta}_2 \frac{\partial^2 \tilde{g}_{t-i}}{\partial^2 \tilde{\epsilon}_{t-i}} \right| \\
&\leq K_1 + K_2 |\tilde{\epsilon}_{t-i}| \qquad (6)
\end{aligned}$$

In (6), K_1 and K_2 are finite constants. The last inequality in (6) follows from the fact that $|\frac{\partial g_{t-i}}{\partial \epsilon_{t-i}}|$ and $|\frac{\partial^2 g_{t-1}}{\partial^2 \epsilon_{t-i}}|$ are bounded functions of $|\epsilon_{t-i}|$. Moreover, from the proof of consistency we have, $|\tilde{w}_t| \leq \sum_{i=1}^{t} \bar{k}^{i-1} |\tilde{\epsilon}_{t-i}|$. Thus,

$$\begin{aligned}
\|\tilde{w}_{1t} - \mathsf{E}[\tilde{w}_{1t}|\mathcal{F}_{t-m}^{t+m}]\| &\leq \sum_{j=m+1}^{t} \bar{k}^{j-1} \left[\sum_{i=1}^{j} (K_1 + K_2 \|\tilde{\epsilon}_{t-i}\|_2) \sum_{i=1}^{t-j} \bar{k}^{i-1} \|\tilde{\epsilon}_{t-j-i}\|_2 \right] \|\epsilon_{t-j}\|_2 \\
&< K \sum_{j=m+1}^{\infty} j \bar{k}^{j-1} \|\epsilon_{t-j}\|_2 = \quad v_m d_t
\end{aligned}$$

with $v_m = K \sum_{j=m+1}^{\infty} j \bar{k}^{j-1}$ and $d_t = \|\epsilon_t\|_2$. That is, $\tilde{w}_{1t}$ is L_2-NED on ϵ_t. The results for $\tilde{w}_{it}, i = 2, \ldots, 5$ follow from similar derivations. Consequently, $\tilde{w}_{it}$ for $i = 1, \ldots, 5$ are L_2-NED on ϵ_t.

We shall show that $\partial\tilde{w}_t/\partial\varphi_1$ are L_2-NED. The distinct elements of $\partial\tilde{w}_t/\partial\varphi_1$ are

$$\frac{\partial\tilde{w}_{1t}}{\partial\varphi_i} = \tilde{k}_{t-1}\frac{\partial\tilde{w}_{1t-1}}{\partial\varphi_i} + \tilde{w}_{it-1} + \frac{\partial\tilde{k}_{t-1}}{\partial\varphi_i}\tilde{w}_{1t-1},$$

for $i = 1, \ldots, 5$.

$$\frac{\partial\tilde{w}_{2t}}{\partial\varphi_i} = \tilde{k}_{t-1}\frac{\partial\tilde{w}_{2t-1}}{\partial\varphi_i} + \frac{\partial\tilde{g}_{t-1}}{\partial\varphi_i} + \frac{\partial\tilde{k}_{t-1}}{\partial\varphi_i}\tilde{w}_{2t-1},$$

for $i = 2, \ldots, 5$.

$$\frac{\partial\tilde{w}_{ht}}{\partial\varphi_i} = \tilde{k}_{t-1}\frac{\partial\tilde{w}_{ht-1}}{\partial\varphi_i} + \frac{\partial\tilde{k}_{t-1}}{\partial\varphi_i}\tilde{w}_{ht-1} + \tilde{\theta}_2\frac{\partial^2\tilde{g}_{t-1}}{\partial\gamma\partial\varphi_i}\tilde{\epsilon}_{t-1} + \tilde{\theta}_2\frac{\partial^2\tilde{g}_{t-1}}{\partial\gamma\partial\varphi_i}\tilde{w}_{it-1},$$

for $i \geq h = 3, 4, 5$, where

$$\frac{\partial\tilde{k}_{t-1}}{\partial\varphi_i} = \frac{\partial\tilde{\theta}_{t-1}}{\partial\varphi_i} + \frac{\partial^2\tilde{\theta}_{t-1}}{\partial\tilde{\epsilon}_{t-1}\partial\varphi_i}\tilde{\epsilon}_{t-1} + \frac{\partial\tilde{\theta}_{t-1}}{\partial\tilde{\epsilon}_{t-1}}\tilde{w}_{it-1}$$

By proceeding in a manner analogous to that for $w_{it}, i = 1, \ldots, 5$ above, we can show that the elements of $\partial\tilde{w}_t/\partial\varphi_1$ are L_2-NED on ϵ_t. Thus, the summands of H_1 are L_1-NED on ϵ_t. Moreover, since $\tilde{w}_t$, $\partial\tilde{w}_t/\partial\varphi_1$, $\tilde{\epsilon}_t$ have more that two finite moments, we can invoke the weak law of large numbers for L_1-mixingale processes. Thus $\tilde{H}_1 \xrightarrow{p} \mathsf{E}(\tilde{H}_1)$ for all $\tilde{\varphi} \in \Psi$.

Finally, we have to show that there exists a function $r(y^t)$ such that

$$|h_{1t}^{(i,j)}(\varphi_1) - h_{1t}^{(i,j)}(\varphi_2)| \leq r(y^t)\langle\varphi_1 - \varphi_2\rangle$$

where $h_{1t}^{(i,j)}(\varphi_1)$ denotes the (i, j)-th element of $\tilde{H}_1$ evaluated at φ_1. As before, $r(y^t)$ is such that $\sup_{\varphi\in\Psi}\left\langle\frac{\partial h1it(\varphi^*)}{\partial\varphi}\right\rangle \leq r(y^t)$.

We have that

$$\left\langle\frac{\partial\tilde{h}_{1t}^{(i,j)}(\tilde{\varphi})}{\partial\varphi}\right\rangle \leq \sum_{s=1}^{6}\left|\frac{\partial\tilde{h}_{1t}^{(i,j)}(\tilde{\varphi})}{\partial\varphi_s}\right|$$

The elements $|(\partial\tilde{h}_{1t}^{(i,j)}/\partial\varphi_s)|$ are not bounded functions of $\{\tilde{\epsilon}_t\}$, but, $\sup_{\varphi\in\Psi}|\partial\tilde{h}_{1t}^{(i,j)}/\partial\varphi_s|$ is function of the L_1-NED process $\{b_t^2\}$. It follows that $H_1 \xrightarrow{p} \mathsf{E}(H_1)$ uniformly in Ψ.

Consider

$$\tilde{H}_2 = \frac{1}{T\tilde{\sigma}^2}\sum_{t=1}^{T}\tilde{w}_t\tilde{\epsilon}_t.$$

From the proof of consistency we have that $\{\tilde{w}_t\}$ and $\tilde{\epsilon}_t$ are L_2-NED on $\{\epsilon_t\}$. From this it follows that $\{\tilde{w}_t\tilde{\epsilon}_t\}$ is L_1-NED on $\{\epsilon_t\}$ and satisfies the weak law of large number for L_1-mixingale processes. It is also possible to show that there exists a dominant function $r(y^t)$

independent of φ that satisfies the weak law of large numbers. That is, $H_2 \xrightarrow{p} \mathsf{E}(H_2)$ uniformly on Ψ. Finally, $H_3 \xrightarrow{p} \mathsf{E}(H_3)$ uniformly on Ψ since $\{\tilde{\epsilon}^2\}$ obeys a weak law of large numbers and $\|\epsilon_t\|_{4p} \leq M < \infty$.

Since $\left(\partial^2 L_T(y^T, \tilde{\varphi})/\partial\varphi\partial\varphi'\right) \xrightarrow{p} \mathsf{E}\left[\partial^2 L_T(y^T, \tilde{\varphi})/\partial\varphi\partial\varphi'\right]$ uniformly on Ψ, and $\left(\partial^2 L_T(y^T, \tilde{\varphi})/\partial\varphi\partial\varphi'\right)$ is a continuous function of φ and $\hat{\varphi} \xrightarrow{p} \varphi^0$ it follows that,

$$\sqrt{T}(\hat{\varphi} - \varphi^0) \quad = \quad -H_0^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q(y^T, \varphi^0)}{\partial\varphi} + o_p(1)$$

where

$$H_0 = -\mathsf{E}\left[\frac{\partial^2 L_t(y^T, \varphi^0)}{\partial\varphi\varphi'}\right] = \begin{bmatrix} \frac{1}{T\sigma_0^2}\sum_{t=1}^{T}\mathsf{E}w_t w_t' & 0 \\ 0 & \frac{2}{\sigma_0^2} \end{bmatrix}$$

Now consider

$$\begin{aligned} \lambda' H_0^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q(y^T, \varphi^0)}{\partial\varphi} &= \lambda' H_0^{-1/2}\frac{1}{\sqrt{T}}\begin{bmatrix} -\frac{1}{\sigma_0^2\sqrt{T}}\sum_{t=1}^{T} w_t\epsilon_t \\ \frac{1}{\sigma_0^3\sqrt{T}}\sum_{t=1}^{T}(\epsilon_t^2 - \sigma_0^2) \end{bmatrix} \\ &= \lambda' H_0^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \end{aligned}$$

with $\lambda'\lambda = 1$. Using the same argument as Engle and Smith (1999), we have that

$$H_0^{-1/2}T^{-1/2}\frac{\partial L_T(y^T, \varphi)}{\partial\varphi}|_{\varphi=\varphi_0} \xrightarrow{d} N(0, 1).$$