Cointegrating Smooth Transition Regressions With Applications to the Asian Currency Crisis*

Pentti Saikkonen[†] and In Choi[‡]

October, 2000

Abstract

This paper studies the smooth transition regression model where regressors are $I(1)$ and errors are $I(0)$. The regressors and errors are assumed to be dependent both serially and contemporaneously. Using the triangular array asymptotics, the nonlinear least squares estimator is shown to be consistent and its asymptotic distribution is derived. It is found that the asymptotic distribution involves a bias under the regressor-error dependence, which implies that the nonlinear least squares estimator is inefficient and unsuitable for use in hypothesis testing. Thus, this paper proposes a Gauss-Newton type estimator which uses the NLLS estimator as an initial estimator and is based on regressions augmented by leads-and-lags. Using leads-and-lags enables the Gauss-Newton estimator to eliminate the bias and have a mixture normal distribution in the limit, which makes it efficient and suitable for use in hypothesis testing. Simulation results indicate that the results obtained from the triangular array asymptotics provide reasonable approximations for the finite sample properties of the estimators and t-tests when sample sizes are moderately large. The cointegrating smooth transition regression model is applied to the Korean and Indonesian data from the Asian currency crisis of 1997. Significant nonlinear effects of interest rate on spot exchange rate are found to be present in the Korean data, which partially supports the interest Laffer curve hypothesis. But overall the effects of interest rate on spot exchange rate are shown to be quite negligible in both the nations.

*This paper was partly written while the first author was visiting Institute of Statistics and Econometrics at Humboldt University, Berlin. This author acknowledges financial support from the Alexander von Humboldt Foundation under a Humboldt Research Award, and the Yrjö Jahnsson Foundation. The second author wrote this paper while visiting the Cowles Foundation for Research in Economics, Yale University. This author thanks the faculty and staff of the Cowles Foundation, especially Don Andrews, John Geanakoplos, David Pearce, Peter Phillips and Nora Wiedenbach for their support and hospitality. Part of the data studied in this paper was provided by Chi-Young Song whom we like to thank.

[†]University Helsinki, Helsinki.

[‡]Kookmin University, Seoul and Cowles Foundation for Research in Economics, Yale University, New Haven.

1 Introduction

It is often perceived that economic agents may show different behavior depending on which regions some economic variables belong in, though it seems hard to find explicit economic theory supporting such behavior. For example, investors and households may make different decisions regarding their investments and savings, respectively, when interest rates are rising rapidly than when they are stable. Another possible example is that employees under recession may behave differently than under boom. Econometricians and statisticians have developed several methods to study such behavior empirically which include, among others, switching regression (cf. Goldfeld and Quandt, 1973), threshold autoregression (cf. Tong, 1983) and smooth transition regression (cf. Granger and Teräsvirta, 1993).

In this paper, we focus on the smooth transition regression (STR) model. As argued in Granger and Teräsvirta (1993), the STR model is useful in explaining the aggregate-level economy because the economy is likely to show smooth transition if each economic agent switches sharply at different times. Asymptotic theory for the STR model involving only stationary variables can be inferred from standard theory in nonlinear econometrics (e.g., Newey and McFadden, 1994; Pötscher and Prucha, 1997).

However, asymptotic theory for the STR model with unit root nonstationary variables has not been developed yet. Recent methods by Park and Phillips (1999, 2000) provide a general framework to study nonstationary and nonlinear time series, but these methods do not seem to be applicable to the STR model. Thus, this paper studies asymptotic theory of the nonlinear least squares (NLLS) estimator for the STR model with $I(1)$ regressors and $I(0)$ errors. This model will be called the cointegrating STR model in this paper. As in most cointegration models, the regressors and errors are assumed to be dependent both serially and contemporaneously. Because using the usual asymptotic scheme of sending sample sizes to infinity does not seem to work for the cointegrating STR model, we will use the triangular array asymptotics. The triangular array asymptotics has been used, among others, in Andrews and McDermott (1995) for nonlinear econometric models with deterministically trending variables and Park and Phillips (2000) for nonstationary and nonlinear regressions.

The asymptotic distribution of the NLLS estimator for the cointegrating STR model involves a bias under the regressor-error dependence, which implies that the NLLS estimator is inefficient and unsuitable for use in hypothesis testing. Therefore, we propose a Gauss-Newton type estimator which uses the NLLS estimator as an initial estimator and is based on nonlinear regressions augmented by leads-and-lags. Linear cointegrating regressions augmented by leads-and-lags are studied in Saikkonen (1991), Phillips and Loretan (1991) and Stock and Watson (1993). The Gauss-Newton estimator eliminates the bias and has a mixture normal distribution in the limit, which implies that it is efficient and that standard hypothesis tests can be performed by using the estimator.

Because the triangular array asymptotic methods have not often been used in econometrics, one may rightfully question the finite sample properties of the tests and estimators using the methods. Therefore, we report some simulation results, which indicate that the results obtained from the triangular array asymptotics provide reasonable approximations for the finite sample properties of the estimators and tests when sample sizes are moderately large.

The cointegrating STR regression model is applied to the Korean and Indonesian data from the Asian currency crisis of 1997. Significant nonlinear effects of interest rate on spot exchange rate are found to be present in the Korean data. This result partially supports the interest Laffer curve hypothesis which states that higher interest rates may depreciate a currency when interest rates are too high because excessively high interest rates may increase the default risk by increasing the borrowing cost of corporations, by depressing the economy and by weakening the banking system of an economy (cf. Goldfajn and Baig, 1998). But overall the effects of interest rate on spot rate are shown to be quite negligible in both the nations. Considering the ineffectiveness of high interest rates in stabilizing exchange rates and the high economic cost associated with keeping high interest rates, the appropriateness of tight monetary policy during the Asian currency crisis should come into question.

The STR model has been used for some economic applications. The applications are Teräsvirta and Anderson (1992) for modelling business cycle asymmetries; Granger, Teräsvirta and Anderson (1993) for forecasting GNP; and Sarno (1999) and Lütkepohl, Teräsvirta and Wolters (1999) for money demand function. Besides these, Luukkonen, Saikkonen and Teräsvirta (1988) consider testing linearity against the smooth transition autoregression model.

The rest of the paper is organized as follows. Section 2 introduces the model and basic assumptions. Section 3 studies asymptotic properties of the NLLS and Gauss-Newton efficient estimators. Section 4 reports some simulation results. Section 5 applies the STR model to the data from the Asian currency crisis. Section 6 contains further remarks. Appendices include auxiliary results and the proofs of theorems.

A few words on our notation: all limits are taken as $T \to \infty$. Weak convergence is denoted as $\Rightarrow$. For square matrices the inequality $A > B$ $(A \geq B)$ means that the difference $A - B$ is positive definite (semidefinite). For an arbitrary matrix A, $\|A\| = [tr(A'A)]^{1/2}$.

2 The Model and Assumptions

Consider the cointegrating smooth transition regression

$$\begin{aligned} y_t &= \mu + \nu g\left((x_t - c_1), ..., (x_t - c_l); \gamma\right) + \sum_{j=1}^{p} \alpha_j x_{jt} + \sum_{j=1}^{p} \delta_j x_{jt} g\left((x_t - c_1), ..., (x_t - c_l), \gamma\right) + u_t \\ &= \mu + \nu g\left(x_t; \theta\right) + \sum_{j=1}^{p} \alpha_j x_{jt} + \sum_{j=1}^{p} \delta_j x_{jt} g\left(x_t, \theta\right) + u_t, \quad t = 1, 2, \ldots, \end{aligned} \tag{1}$$

where x_{jt} is the $j-$th component of the $I(1)$ vector x_t $(p \times 1)$, u_t is a zero-mean stationary error term, $\theta = [c_1', ..., c_l', \gamma']'$ and $g\left(x_t; \theta\right)$ is a smooth real-valued transition function of the process x_t and the parameter vector θ.[1] Moreover, μ, ν, α_j, and δ_j are scalar parameters.

The STR model (1) has been used to describe economic relations which change smoothly depending on the location of some economic variables. In model (1), the relationship between x_t and y_t may change depending on where x_t is located relative to parameters $c_1, ..., c_l$. Parameter γ in model (1) determines the smoothness of transition in the economic relations. The reader is referred to Granger and Teräsvirta (1993) for more discussions on the STR model, although these authors do not explicitly consider the case of $I(1)$ processes.

We discuss some examples of model (1) by using the following simplified version of model (1) where nonlinearity appears only in the first regressor.

$$y_t = \mu + \alpha_1 x_{1t} + \delta_1 x_{1t} g\left((x_{1t} - c_1), ..., (x_{1t} - c_l); \gamma\right) + \sum_{j=2}^{p} \alpha_j x_{jt} + u_t, \quad t = 1, 2, \ldots \quad (2)$$

Example 1:

$$g\left((x_{1t} - c_1), ..., (x_{1t} - c_l); \gamma\right) = \frac{1}{1 + e^{-\gamma(x_{1t} - c)}}. \quad (3)$$

Here the transition function is a logistic function which makes regression coefficient for x_{1t} vary smoothly between α_1 and $\alpha_1 + \delta_1$. When the value of the regressor x_{1t} is sufficiently far below the value of the parameter c the regression coefficient takes a value close to α_1; and when the value of the regressor x_{1t} increases and exceeds the value of the parameter c the value of the regression coefficient changes and approaches $\alpha_1 + \delta_1$.

Example 2:

$$g\left((x_{1t} - c_1), ..., (x_{1t} - c_l); \gamma\right) = \frac{1}{1 + e^{-\gamma(x_{1t} - c_1)(x_{1t} - c_2)}}, \ c_1 < c_2. \quad (4)$$

This transition function can be used when one wants to allow for the possibility that the regression coefficient changes twice. When $\mid x_{1t} \mid$ is large, the function takes a value close to 1 so that the coefficient for x_{1t} approaches $\alpha_1 + \delta_1$. But when x_{1t} is approximately in between c_1 and c_2, the function takes a value close to zero which makes the coefficient for x_{1t} approach α_1. Instead of function (4), one may also use a linear combination of two logistic functions.

When $\gamma \to \infty$, functions (3) and (4) approach the indicator functions $\mathbf{1}\{x_{1t} \geq c\}$ and $\mathbf{1}\{x_{1t} \geq c_1 \text{ or } x_{1t} \leq c_2\}$, respectively, and model (2) becomes close to a threshold regression model. Then the change in the regression coefficient of x_{1t} is abrupt and not gradual as assumed in (2). Our results do not apply to threshold models because the transition function is not allowed to be discontinuous. Otherwise our treatment is fairly general, and applies to any sufficiently well behaved transition function.

[1] Although model (1) assumes that all the regressors have a nonlinear effect on the regressand, our theoretical results can readily be modified to the case where the nonlinearity only appears in some of the regressors. In addition, our set-up does not allow for the possibility that different transition functions are used for different regressors. But it would not be difficult to extend our results to that case as well. To simplify exposition, we have preferred to work with a single transition function.

We shall now discuss assumptions required for model (1). As already mentioned, we assume

Assumption 1

$$x_t = x_{t-1} + v_t, \quad t = 1, 2, \ldots, \tag{5}$$

where v_t is a zero-mean stationary process and the initial value x_0 may be any random variable with the property $E\|x_0\|^2 < \infty$.

Moreover, it will be convenient to assume that the $(p+1)-$dimensional process $w_t = [u_t \; v_t']'$ satisfies the following assumption employed by Hansen (1992) in a somewhat weaker form.

Assumption 2 *For some $r > 4$, $w_t = [u_t \; v_t']'$ is a stationary, zero-mean, strong mixing sequence with mixing coefficients of size $-4r/(r-4)$ and $E\|w_t\|^r < \infty$.*

Assumption 2 is fairly general and covers a variety of weakly dependent processes. It also permits the cointegrated system defined by (1) and (5) to have nonlinear short-run dynamics which is convenient because our cointegrating regression is nonlinear.

Choosing the real number p in Corollary 14.3 of Davidson (1994) as $2r/(r+2)$, we find that Assumption 2 implies that the serial covariances of the process w_t at lag $|j|$ are of size -2. Thus, we have the summability condition

$$\sum_{j=-\infty}^{\infty} |j| \left\| E w_t w_{t+j}' \right\| < \infty. \tag{6}$$

This implies that the process w_t has a continuous spectral density matrix $f_{ww}(\lambda)$ which we assume to satisfy

Assumption 3 *The spectral density matrix $f_{ww}(\lambda)$ is bounded away from zero or that*

$$f_{ww}(\lambda) \geq \varepsilon I_n, \quad \varepsilon > 0. \tag{7}$$

Assumption 3 specialized to the case $\lambda = 0$ implies that the components of the $I(1)$ process x_t are not cointegrated. In addition, it is required for the estimation theory of Section 2 that (7) also holds for other values of λ. Conformably to the partition of the process w_t, we write $f_{ww}(\lambda) = [f_{ab}(\lambda)]$ where $a, b \in \{u, v\}$.

Assumption 2 also implies the multivariate invariance principle

$$T^{-1/2} \sum_{j=1}^{[Ts]} w_j \Rightarrow B(s), \quad 0 \leq s \leq 1, \tag{8}$$

where $B(s)$ is a Brownian motion with covariance matrix $\Omega = 2\pi f(0)$ (see Hansen, 1992, the proof of Theorem 3.1). We partition $B(s) = \left[B_u(s) \;\; B_v(s)'\right]'$ and

$$\Omega = \begin{bmatrix} \omega_u^2 & \omega_{uv} \\ \omega_{vu} & \Omega_{vv} \end{bmatrix}$$

conformably with the partition of the process w_t.

As for the transition function $g(x; \theta)$, we assume

Assumption 4 *(i) The parameter space* Θ *of* θ *is a compact subset of an Euclidean space.*

(ii) $g(x;\theta)$ *is three times continuously differentiable on* $\mathsf{R}^p \times \Theta^*$ *where* Θ^* *is an open set containing* Θ*.*

This assumption may not be the weakest possible, but it is satisfied by the most commonly used transition functions and simplifies exposition. Thus, we shall not try to weaken it. The compactness of the parameter space Θ is a standard assumption in nonlinear regression, but no such assumption is needed for other parameters.

3 Estimation Procedures

The cointegrating regression (1) assumes serial and contemporaneous correlation between the $I(1)$ regressor x_t and the error term u_t. Adverse consequences of this on linear least squares estimation are well known and various modifications have therefore been devised. In this paper, we extend the leads-and-lags procedure of Saikkonen (1991) to the STR model discussed in the previous section. Because there are some theoretical difficulties with a direct extension of this procedure, we will first consider the NLLS estimation which can be utilized to develop a Gauss-Newton type leads-and-lags estimation.

3.1 Triangular array asymptotics

Before embarking on the subject of NLLS estimation, we will explain the motivation for the employed asymptotic methods in this subsection. Park and Phillips (2000) show that two types of asymptotics can be considered in nonlinear regressions with $I(1)$ regressors. One is the usual asymptotics, and the other is the so-called triangular array asymptotics in which the actual sample size is fixed at T_0, say, and the model is imbedded in a sequence of models depending on a sample size T which tends to infinity. The imbedding is obtained by replacing the $I(1)$ regressor by $(T_0/T)^{1/2} x_t$. This makes the regressand dependent on T and, when $T = T_0$, the original model is obtained. Thus, if T_0 is large, the triangular array asymptotics can be expected to give reasonable approximations for finite sample distributions of estimators and test statistics. The triangular array asymptotics is also used in Andrews and McDermott (1995) for nonlinear econometric models with deterministically trending variables. Related references can also be found in Andrews and McDermott.

We will use the triangular array asymptotics for our cointegrating model, because we expect it to provide quite reasonable approximations for estimators and test statistics and because some parameters cannot be identified when the usual asymptotics is used. The identification issue can be explained intuitively by using a special case of model (1) – the model in Example 1. When the model in Example 1 is applied, a typical situation is that the observations can be divided into three groups with each group containing a reasonably large proportion of the data. In the first and third group the values of the regression coefficient for x_{1t} are essentially α_1 and $\alpha_1 + \delta_1$, respectively, whereas the second group contains part of the sample where the value of the regression coefficient changes between these two values. Since x_{1t} is an $I(1)$ process, the use of conventional asymptotics means that the variation of x_{1t} increases so that the proportion of observations in the first and third groups increases and that in the second group decreases. Eventually the proportion of observations in the second group becomes negligible. This suggests that these parameters are unidentifiable in the limit, because only observations in the second group provide information about the parameters γ and c. This can also be seen by noting that, for T large,

$$g(\gamma(x_{1t}-c)) = g\left(T^{1/2}\gamma\left(T^{-1/2}x_{1t}-T^{-1/2}c\right)\right) \approx \mathbb{1}\{T^{-1/2}x_{1t} \geq 0\}.$$

Thus, asymptotically the parameters γ and c vanish from the model and become unidentifiable. This discussion implies that the use of conventional asymptotics leads to a situation which is very different from what happens in the sample where a reasonably large proportion of observations belongs to the second group.

However, the triangular array asymptotics takes the second group and, therefore, the parameters γ and c into account. Recall that $g(\cdot;\theta)$ is the logistic function. Basing the asymptotic analysis on $g\left(\gamma\left((T_0/T)^{1/2}x_{1t}-c\right)\right) = g\left(T^{-1/2}\gamma\left(T_0^{1/2}x_{1t}-T^{1/2}c\right)\right)$ instead of $g(\gamma(x_{1t}-c))$ means that the slope of the logistic function is assumed to decrease so that the proportion of observations in the three groups remains essentially the same even though the variation of x_{1t} increases. In this respect the situation for the triangular array asymptotics remains the same as for the sample. It also makes sense that parameter c has to be of order $O\left(T^{1/2}\right)$, because, due to the increasing variation of x_{1t}, a nonzero value of c could otherwise be indistinguishable form zero. Finally, note that when $g\left(\gamma\left((T_0/T)^{1/2}x_{1t}-c\right)\right)$ is used in asymptotic analysis the process x_{1t} is standardized in such a way that it remains bounded in probability. In this context a possible interpretation is that when T tends to infinity observations of the standardized version of the series x_{1t} are obtained denser and denser within its observed range in the sample; and thereby the proportion of observations in each of the three groups remains essentially the same which makes information about parameters γ and c retained even asymptotically.

Although the above discussion gives a reasonable motivation for using the triangular array asymptotics, it would be imprudent to claim that the triangular array asymptotics would always work well. For instance, we already noticed that problems may occur if the value of the parameter γ in model (2) with specification (3) is large so that the model is close to a threshold model.

3.2 NLLS estimation

This subsection considers the triangular array asymptotics of the NLLS estimator for model (1). In order to use the triangular array asymptotics, we imbed model (1) in a sequence of models

$$y_{tT} = f(x_{tT};\theta)'\phi + u_t, \quad t = 1,\dots,T, \tag{9}$$

where $x_{tT} = (T_0/T)^{1/2} x_t$[2], $f(x_{tT};\theta) = [1 \;\; g(x_{tT};\theta) \;\; x'_{tT} \;\; g(x_{tT};\theta)\, x'_{tT}]'$ and $\phi = \left[\mu \;\; \nu \;\; \alpha' \;\; \delta'\right]'$ with $\alpha = [\alpha_1 \;\cdots\; \alpha_p]'$ and $\delta = [\delta_1 \;\cdots\; \delta_p]'$.

In what follows we set $\vartheta = \left[\theta' \; \phi'\right]'$ and let $\vartheta_0 = \left[\theta'_0 \; \phi'_0\right]'$ stand for the true value of ϑ. The NLLS estimator of parameter ϑ_0 is obtained by minimizing the function

$$Q_T(\vartheta) = \sum_{t=1}^{T} \left(y_{tT} - f(x_{tT};\theta)' \phi\right)^2 \tag{10}$$

with respect to ϑ.

The assumptions made so far do not ensure that a minimum of function (10) exists, even asymptotically. To be able to introduce further assumptions, we first use the multivariate invariance principle (8) to conclude that $x_{tT} \Rightarrow T_0^{1/2} B_v(s) \overset{def}{=} B_v^0(s)$ as $T \to \infty$. This fact and a standard application of the continuous mapping theorem show that, for every $\theta \in \Theta$,

$$T^{-1} \sum_{t=1}^{T} f(x_{tT};\theta) f(x_{tT};\theta)' \Rightarrow \int_0^1 f\left(B_v^0(s);\theta\right) \; f\left(B_v^0(s);\theta\right)' ds.$$

An assumption which together with our previous assumptions ensures that the function $Q_T(\vartheta)$ has a minimum for T large enough is:

Assumption 5 *For some $\varepsilon > 0$,*

$$\inf_{\theta \in \Theta} \lambda_{\min} \left(\int_0^1 f\left(B_v^0(s);\theta\right) \; f\left(B_v^0(s);\theta\right)' ds \right) \geq \varepsilon > 0 \quad (a.s.) \tag{11}$$

where $\lambda_{\min}(\cdot)$ signifies the smallest eigenvalue of a square matrix.

Assumption 5 guarantees that, with probability approaching one, a minimum of the function $Q_T(\vartheta)$ exists as shown in Appendices.[3] Since we are interested in asymptotic results, we may as usual assume that a minimum exists for all values of T and is attained at $\tilde{\vartheta}_T = \left[\tilde{\theta}'_T \; \tilde{\phi}'_T\right]'$.

In addition to Assumption 5, the following assumption is needed for the consistency of the least squares estimator $\tilde{\vartheta}_T$.

Assumption 6 *For some $s \in [0,1]$ and all $(\theta, \phi) \neq (\theta_0, \phi_0)$,*

$$f\left(B_v^0(s);\theta\right)' \phi \neq \; f\left(B_v^0(s);\theta_0\right)' \phi_0 \quad (a.s.). \tag{12}$$

This is an identification condition which ensures that the parameters θ and ϕ can be separated in the product $f(x_{tT};\theta)' \phi$. Taken together, Assumptions 5 and 6 ensure the identifiability of the parameter vector ϑ.

[2] In practice we always choose $T = T_0$, so that the transformation is not required. The transformation is made only to facilitate the development of asymptotic analysis.

[3] See Lemma 5 and the proof of Theorem 1.

It might be convenient to be able to formulate deterministic identification conditions which do not depend on the sample paths of the Brownian motion $B_v^0(s)$. However, in general, it seems difficult to remove the Brownian motion $B_v^0(s)$ from conditions (11) and (12).[4] Nevertheless, conditions (11) and (12) appear fairly easy to use. For instance, it can be checked by the conditions that model (2) with specification (3) is identified when $\delta_1 \neq 0$ and $\gamma > 0$.

It may also be argued that it makes sense to use identification conditions which depend on the sample paths of the Brownian motion $B_v^0(s)$ when the triangular array asymptotics is used. Indeed, in applications of model (2) with specification (3), one can typically divide the observations into three groups in such a way that a fair amount of observations belongs to each group and, when the triangular array asymptotics is used, this state of affairs prevails even asymptotically. Thus, since $x_{tT} \Rightarrow B_v^0(s)$, the triangular array asymptotics in a sense conditions on such sample paths of $B_v^0(s)$ for which the shape of the function $g\left(\gamma\left(B_v^0(s) - c\right)\right)$ is similar to what is observed in the sample. Due to this 'conditioning', it seems quite reasonable to use identification conditions which depend on the sample paths of the Brownian motion $B_v^0(s)$ and ensure identifiability when the specified nonlinearity is related to the sample paths of $B_v^0(s)$ in the same way as to the observed realizations of x_{tT} within the sample. This means, for instance, that in the case of model (2) with specification (3) we are not interested in identification in cases where sample paths of $B_v^0(s)$ are such that the function $g\left(\gamma\left(B_v^0(s) - c\right)\right)$ is effectively constant and identifiability is very weak although it is still achieved when $\delta_1 \neq 0$ and $\gamma > 0$. This point could be made even stronger by replacing the logistic function by a piecewise continuous analog so that for some realizations of $B_v^0(s)$ the function $g\left(\gamma\left(B_v^0(s) - c\right)\right)$ would actually be constant and identifiability would fail. Clearly, such cases would be of no interest if $g\left(\gamma\left(x_{tT} - c\right)\right)$ is highly nonlinear within the sample.

The following theorem shows the existence and consistency of the least squares estimator $\tilde{\vartheta}_T$.

Theorem 1 *Suppose that Assumptions 1-6 hold. Then, a NLLS estimator $\tilde{\vartheta}_T$ exists with probability approaching one and is consistent.*

Theorem 1 shows the existence and consistency of the least squares estimator $\tilde{\vartheta}_T$ when the triangular array asymptotics is used. The following theorem shows the limiting distribution of the estimator $\tilde{\vartheta}_T$. For this theorem we need an additional assumption

Assumption 7

$$\int_0^1 K\left(B_v^0(s)\right) K\left(B_v^0(s)\right)' ds > 0 \quad (a.s.) \tag{13}$$

where

$$K(x) = \begin{bmatrix} \left(\nu_0 + \delta_0' x\right) \partial g\left(x;\theta_0\right) / \partial\theta \\ f\left(x;\theta_0\right) \end{bmatrix}.$$

[4] When x_t is a scalar process, it is possible to formulate an identification condition by using the regression function with a deterministic argument (see Park and Phillips, 2000, Theorem 4.6).

Theorem 2 *Suppose that Assumptions 1-7 hold and that θ_0 is an interior point of Θ. Then,*

$$T^{1/2}\left(\tilde{\vartheta}_T-\vartheta_0\right)\Rightarrow\left(\int_0^1 K\left(B_v^0(s)\right)K\left(B_v^0(s)\right)'ds\right)^{-1}$$
$$\times\left(\int_0^1 K\left(B_v^0(s)\right)dB_u(s)+\int_0^1 K_1\left(B_v^0(s)\right)ds\kappa_{vu}\right)$$

where $K_1(x)=\partial K(x)/\partial x'$ and $\kappa_{vu}=\sum_{j=0}^{\infty}Ev_0u_j$.

The limiting distribution given in Theorem 2 depends on nuisance parameters in a complicated way which renders the NLLS estimator inefficient and, in general, makes it unsuitable for hypothesis testing. This difficulty is removed in a special case where the processes v_t and u_t are totally uncorrelated, because then the limiting distribution becomes mixed normal as can be easily checked.

In its general form, Theorem 2 shows that the NLLS estimator is consistent of order $O_p\left(T^{-1/2}\right)$. This will be used to obtain an efficient two-step estimator based on the leads-and-lags modification. The reason why the order of consistency differs from $O_p\left(T^{-1}\right)$ obtained in previous linear cases is that we employ the triangular array asymptotics in which the regressand is made bounded.

3.3 Efficient estimation

This subsection considers efficient estimation of model (1) by using a leads-and-lags regression. As in Saikkonen (1991), we can express the error term u_t as

$$u_t=\sum_{j=-\infty}^{\infty}\pi_j'v_{t-j}+e_t \tag{14}$$

where e_t is a zero-mean stationary process such that $Ee_tv_{t-j}'=0$ for all $j=0,\pm1,\ldots,$ and

$$\sum_{j=-\infty}^{\infty}(1+|j|)\left\|\pi_j\right\|<\infty. \tag{15}$$

That this summability condition holds follows from condition (6) and Theorem 3.8.3 in Brillinger (1975). Expressions for the spectral density function and long-run variance of the process e_t can be obtained from the well-known formulas $f_{ee}(\lambda)=f_{uu}(\lambda)-f_{uv}(\lambda)f_{vv}^{-1}(\lambda)f_{vu}(\lambda)$ and $\omega_e^2=\omega_u^2-\omega_{uv}\Omega_{vv}^{-1}\omega_{vu}$, respectively.

Using equations (5) and (14), we can write the cointegrating regression (1) as

$$y_t=\mu+\nu g(x_t;\theta)+\alpha'x_t+\delta'x_tg(x_t;\theta)+\sum_{j=-K}^{K}\pi_j'\Delta x_{t-j}+e_{Kt},\quad t=K+1,\ldots,T-K, \tag{16}$$

where Δ signifies the difference operator and

$$e_{Kt}=e_t+\sum_{|j|>K}\pi_j'v_{t-j}.$$

In order to eliminate errors caused by truncating the infinite sum in (14) we have to consider asymptotics in which the integer K tends to infinity with T. The condition $K = o\left(T^3\right)$ used in the linear case by Saikkonen (1991) can also be used here.

Since we continue with the same triangular array asymptotics as in the previous subsection, we imbed model (16) in a sequence of models

$$y_{tT} = f\left(x_{tT};\theta\right)'\phi + V_t'\pi + e_{Kt},\ t = K+1,\ldots,T-K, \tag{17}$$

where $V_t = \left[\Delta x_{t-K}' \ \ldots \ \Delta x_{t+K}'\right]'$ and $\pi = \left[\pi_{-K}' \ \ldots \ \pi_K'\right]'$. Combining the regressors as $q\left(x_{tT};\theta\right) = \left[f\left(x_{tT};\theta\right)' \quad V_t'\right]'$ we can write this model more compactly as

$$y_{tT} = q\left(x_{tT};\theta\right)'\beta + e_{Kt}, \quad t = K+1,\ldots,T-K, \tag{18}$$

where $\beta = \left[\phi' \quad \pi'\right]'$.

Instead of proper nonlinear least squares estimators of the parameters in (18) we shall consider two-step estimators based on the NLLS estimator of the previous section. These estimators are defined by

$$\begin{bmatrix} \hat{\vartheta}_T^{(1)} \\ \hat{\pi}_T^{(1)} \end{bmatrix} = \begin{bmatrix} \tilde{\vartheta}_T \\ 0 \end{bmatrix} + \left(\sum_{t=K+1}^{T-K} \tilde{p}_{tT}\tilde{p}_{tT}'\right)^{-1} \sum_{t=K+1}^{T-K} \tilde{p}_{tT}'\, \widetilde{u}_{tT} \tag{19}$$

where $\widetilde{u}_{tT} = y_{tT} - f\left(x_{tT};\tilde{\theta}_T\right)'\tilde{\phi}_T$ and $\tilde{p}_{tT} = \left[\tilde{K}\left(x_{tT}\right)' \ \ V_t'\right]'$ with

$$\tilde{K}\left(x_{tT}\right) = \begin{bmatrix} \left(\tilde{\nu}_T + \tilde{\delta}_T' x_{tT}\right)\partial g\left(x_{tT};\tilde{\theta}_T\right)/\partial\theta \\ f\left(x_{tT};\tilde{\theta}_T\right) \end{bmatrix}.$$

The latter term on the right hand side of (19) is obviously the least squares estimator obtained from a regression of $\widetilde{u}_{tT}$ on $\tilde{p}_{tT}$. This estimator will be called the Gauss-Newton estimator.

To see the motivation of the Gauss-Newton estimator, subtract $f(x_{tT};\tilde{\theta}_T)'\tilde{\phi}_T$ from both sides of (17) and apply the mean value approximation $f(x_{tT};\theta)'\phi - f(x_{tT};\tilde{\theta}_T)'\tilde{\phi}_T \approx \tilde{K}\left(x_{tT}\right)'(\vartheta - \tilde{\vartheta}_T)$ to the right hand side. Thus, after linearization, we get the auxiliary regression model

$$\widetilde{u}_{tT} = \tilde{K}\left(x_{tT}\right)'(\vartheta - \tilde{\vartheta}_T) + V_t'\pi + error$$

which in conjunction with standard least squares theory gives estimator (19).

The following theorem describes asymptotic properties of the estimators $\hat{\vartheta}_T^{(1)}$ and $\hat{\pi}_T^{(1)}$. The limiting distribution of the estimator $\hat{\vartheta}_T^{(1)}$ requires a standardization by the square root of $T - 2K$, the effective number of observations in the regression of $\widetilde{u}_{tT}$ on $\tilde{p}_{tT}$. For convenience, we denote $N = T - 2K$.

Theorem 3 *Suppose that the assumptions of Theorem 2 hold and that $K \to \infty$ in such a way that $K^3/T \to 0$ and $T^{1/2}\sum_{|j|>K}\|\pi_j\| \to 0$. Then,*

(i)

$$N^{1/2}\left(\hat{\vartheta}_T^{(1)}-\vartheta_0\right)\Rightarrow\left(\int_0^1 K\left(B_v^0(s)\right)K\left(B_v^0(s)\right)'ds\right)^{-1}\int_0^1 K\left(B_v^0(s)\right)dB_e(s)$$

where $B_e(s)$ is a Brownian motion which is independent of $B_v(s)$ and has variance ω_e^2.

(ii) $\|\hat{\pi}_T-\pi_0\|=O_p\left(K^{1/2}/N^{1/2}\right)$.

The independence of the Brownian motions $B_e(s)$ and $B_v(s)$ implies that the limiting distribution in Theorem 3 is mixed normal. Furthermore, we can conclude from Saikkonen (1991) that the Gauss-Newton estimator $\hat{\vartheta}_T^{(1)}$ is asymptotically more efficient than the least squares estimator $\tilde{\vartheta}_T$ in general. In the same way as in Saikkonen (1991), we have also here been forced to supplement the previously mentioned condition $K=o\left(T^3\right)$ by an additional condition which implies that the integer K may not increase too slowly.

Theorem 3 indicates that we can estimate ω_e^2 consistently (see, for example, Andrews, 1991) by using the residuals from the regression model (16) with estimator (19). Thus, conventional tests like Wald and t-tests can be constructed in a straightforward manner and shown to have standard distributions in the limit.

4 Simulation

Implications of the theoretical results in Section 3 can be summarized as: (i) The NLLS and Gauss-Newton estimators are consistent. (ii) In large samples, the Gauss-Newton estimator eliminates the bias of the NLLS estimator and is more efficient than the NLLS estimator. (iii) The t-test based on the Gauss-Newton estimator follows a standard normal distribution in the limit. Because these results are based on the triangular array asymptotics where the sample size of the embedding model goes to infinity, it may not seem quite obvious whether these results hold when the sample size T_0 is large. Therefore, this section examines the aforementioned results by using simulation.

Data were generated by

$$\begin{aligned}
y_t &= \mu+\alpha x_t+\delta x_t\frac{1}{1+\exp(-(x_t-c))}+u_t, \quad t=1,...,T_0\\
\mu &= \alpha=\delta=1; c=5; x_t=x_{t-1}+v_t;\\
\begin{pmatrix} u_t \\ v_t \end{pmatrix} &= \varepsilon_t+B\varepsilon_{t-1}; B=\begin{bmatrix}\omega & \omega\\ 0 & \omega\end{bmatrix}; \omega=.2,.5,.8;\\
\varepsilon_t &\sim iidN\left(0,\begin{bmatrix}1 & \sigma_{12}\\ \sigma_{12} & 1\end{bmatrix}\right); \sigma_{12}=.5. \qquad (20)
\end{aligned}$$

Larger ω implies that the regressors and errors are more correlated both serially and contemporaneously.

Unreported simulation results indicate that it is difficult to estimate parameter γ accurately by the NLLS method unless either sample sizes are very large or parameter c is located close to the median of $\{x_t\}$; and that other parameter estimates are quite adversely affected by poor estimates of γ. Since the purpose of this section is to check the implications of the triangular array asymptotics, we do not want that our simulation results are affected by outliers produced by poor estimates of parameter γ. Therefore, we assume that the value of the transition parameter γ is known to be 1. Also, $\{x_t\}$ were generated such that c is located in between the 15th and 85th percentiles of $\{x_t\}$. The purpose of this scheme is the same as that of fixing the value of γ. When c is near endpoints of the sample, extremely poor estimates of parameter c are sometimes produced which affects other parameter estimates to the extent that evaluating their finite sample performance at different sample sizes becomes meaningless.

The estimators considered are the NLLS, one-step Gauss-Newton and two-step Gauss-Newton estimators.[5] The values of the leads-and-lags parameter for the Gauss-Newton estimators were set at $K = 1, 2, 3$. Table 1 reports the empirical biases and root mean squares errors (RMSEs) of the estimators at sample sizes 150 and 300.[6] The numbers of replications at $T_0 = 150$ and $T_0 = 300$ were 5,000 and 3,000, respectively. As for the method of minimization, the Polak-Ribiere conjugate gradient method[7] was used. The results in Table 1 can be summarized as follows.

- As sample size T_0 grows, the RMSEs of all the estimators decrease, which may be interpreted as evidence for consistency.
- The Gauss-Newton estimators reduce the magnitudes of bias and RMSE substantially in relation to the NLLS estimator as predicted by Theorem 3.
- As the regressors and errors are more correlated both serially and contemporaneously, the two-step Gauss-Newton estimator tends to improve the one-step Gauss-Newton estimator in terms of RMSE. But the two-step Gauss-Newton estimator is sometimes more biased than the one-step Gauss-Newton estimator, though the degree of the biases for both the estimators is quite mild.
- The choice of the parameter K does not seem to affect the results significantly.
- The nonlinear parameter c is subject to higher RMSE than other linear parameters, which may reflect the computational difficulties associated with estimating the nonlinear parameter.

Table 2 reports empirical sizes of the t-ratios using the Gauss-Newton estimators under the null hypotheses $\alpha = 1, \delta = 1$ and $c = 5$. Nominal sizes were chosen to be 5% and 10%, and the same experimental format as for Table 1 was used. The results in Table 2 can be summarized as follows.

- The t-ratios reject more often than they should in part (1). But increasing the

[5] The one-step Gauss-Newton and two-step Gauss-Newton estimators use the NLLS and one-step Gauss-Newton estimators as initial estimators, respectively.

[6] We do not report the results for the estimators of μ, because these are not the main concern in most applications.

[7] It was found that quasi-Newton methods tend to give more outliers. The maximum number of iterations for optimization was set at 100,000.

sample size T_0 to 300 improves the performance of the t-ratios.[8]

- When there are less serial and contemporaneous correlations between the regressors and errors at $T_0 = 300$, empirical sizes get closer to the corresponding nominal sizes. But this is not noticeable at $T_0 = 150$.
- The one-step and two-step Gauss-Newton estimators show similar performance.
- Choosing $K = 1$ and $K = 2$ at $T_0 = 150$ and $T_0 = 300$, respectively, tends to provide the best results.

In sum, the simulation results in Tables 1 and 2 seem to confirm that the results from the triangular array asymptotics in Section 2 can provide reasonable approximations for the finite sample properties of the estimators and tests when the sample size is moderately large.

5 Applications to the Asian Currency Crisis

One of the substantial controversies regarding the Asian currency crisis of 1997 has been whether tight monetary policy was effective in stabilizing foreign exchange rates during and in the aftermath of the crisis.[9] In fact, tight monetary policy constituted an essential part of the IMF rescue package for Asian countries, because it has conventionally been believed that higher interest rates reduce capital outflows by raising the cost of currency speculation and induce capital inflows by making domestic assets more attractive in the short run; and that they improve current account balance by reducing domestic absorption in the long-run.

However, as discussed in Goldfajn and Baig (1998), higher interest rates may depreciate a currency when interest rates are too high because excessively high interest rates may increase the default risk by increasing the borrowing cost of corporations, by depressing the economy and by weakening the banking system of an economy.[10] This hypothesis may be called the "interest Laffer curve" hypothesis because the effects of interest rates on spot exchange rates are hypothesized to depend on the levels of the interest rates. This section employs the model and asymptotic theory developed in previous sections to study the interest Laffer curve hypothesis and reports the magnitudes of interest elasticity of spot rate for Korea and Indonesia during the Asian currency crisis.

The uncovered interest rate parity relation predicts that log spot rate is related to the difference of domestic and foreign interest rates and log expected future spot rate.[11] Though the relation predicted by the interest rate parity condition is strictly

[8] Increasing the sample size to 500 further improves the empirical size of the t-ratio, though the results are not reported here.

[9] See Goldfajn and Baig (1998), Kaminsky and Schmukler (1998), Ghosh and Phillips (1998), Kraay (1998), Dekle, Hsiao and Wang (1999), Park, Wang and Chung (1999) and Choi and Park (2000) for empirical results regarding this issue.

[10] In addition, Feldstein (1998), Furman and Stiglitz (1998) and Radelet and Sachs (1998a,b), among others, argue that tight monetary policy in Asia was either ineffective in stabilizing exchange rates or that it may have even exacerbated the situation.

[11] The uncovered interest rate parity relation is written as $1 + i_t = (1 + i_t^*)\frac{S_{t+1}^e}{S_t}$, where i_t and i_t^* denote the domestic and the foreign interest rates at date t, respectively; and S_t and S_{t+1}^e denote the

linear, it indicates that the difference of the domestic and foreign interest rates and the log expected future spot rate may be considered as major variables explaining the spot rate. This consideration leads us to employ the difference of the domestic and foreign interest rates and the log expected future spot rate as regressors in our nonlinear regression. But because the expected future spot rate is not observable, forward exchange rate can be used as its substitute. More specifically, the STR model we use in this section is

$$y_t = \mu + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \delta x_{2t} \frac{1}{1 + e^{-\gamma(x_{2t} - c)}} + u_t, \tag{21}$$

where y_t and x_{1t} are the spot and forward rates, respectively and x_{2t} is the difference between the domestic and foreign interest rates (i.e., $i_t - i_t^*$). Because we are interested only in the nonlinear relation between the spot rate and the interest rate differential, the transition function includes only the interest rate differential. Equation (21) signifies that the relation between the spot rate and the interest rate differential changes when the latter is well above the level c unless γ is zero. Thus, the model is appropriate for studying the relation between the spot rate and the interest rate differential which may change depending on the level of the interest rate differential.

The spot exchange rate data which we use are daily nominal exchange rates of Korea and Indonesia vis-a-vis the U.S. dollar. For forward exchange rates, one-month maturity data are used. For Korea, we use the forward exchange rate from the NDF market.[12] For Indonesia, we use data from their onshore forward exchange markets.[13] For domestic interest rates, we use the overnight call rates of each country. Since the overnight call rates are the main monetary policy instruments of each country, they seem to best reflect monetary policy stances of each country and could be regarded as exogenous policy variables. The U.S. federal funds rate is used as the foreign interest rate.

The whole sample covers the 19 month periods 4/1/1997 - 10/30/1998 for Korea and 1/3/1997 - 1/24/1998 for Indonesia. The sample period for each country begins at about six to seven months before the eruption of its own currency crisis. The sample sizes for Korea and Indonesia are 343 and 406, respectively.

The results of the two-step Gauss-Newton estimation of model (21) are reported

spot exchange rate at date t and the expected future spot exchange rate at date $t+1$, respectively. Taking logs of both sides of the interest parity relation yields $\ln(S_t) = \ln(1+i_t^*) - \ln(1+i_t) + \ln(S_{t+1}^e) \approx i_t^* - i_t + \ln(S_{t+1}^e)$.

[12] The NDFs are non-deliverable forwards traded in the offshore market. Unlike the onshore forward exchange rates which has been influenced by direct regulation and heavy intervention of the Korean government, we believe that the NDF rates better reflect expectations of market participants.

[13] Since Indonesia had already liberalized domestic foreign exchange markets, the Indonesian rupiah were not traded in the NDF market.

in Table 3.[14] Instead of estimating parameter γ, we used $\gamma = 0.1, 0.2, ..., 3.5$[15] and reported the results which yielded the least sum of squared errors for the two-step Gauss-Newton estimator. This is because unreported simulation results indicate that the NLLS estimator of parameter γ is subject to high errors, which also affect the Gauss-Newton estimators adversely. The results for Korea show that all the regressors are significant and that significant nonlinear effects of the interest differential are present.[16] The location parameter c is estimated to lie in between 14.6 and 15.0 depending on the choice of the leads-and-lags parameter K. Though only the results for $\gamma = 0.6$ are reported here, we note that choosing other values of γ greater than 0.6 does not bring any qualitatively different results. For Indonesia, nonlinear effects of the interest differentials seem to be mild, though coefficients for the forward rate and the interest differentials are significant. The estimates of the location parameter c are similar in magnitudes to those for Korea, though Indonesia experienced much higher interest rates than Korea during the period of currency crisis.[17] As in the case of Korea, the estimation and test results are robust to other choices of γ.

The results in Table 3 indicate that the future rates are quite important in explaining the spot rates given the magnitudes of the coefficient estimates. But the coefficient estimates for the terms involving the interest differentials are close to zero whether they are statistically significant or not. To visualize the nonlinear effects of the interest differentials, we draw the interest differential elasticity of spot rate in Figures 1 and 2 by assuming that the estimation results in Table 3 (with $K = 2$) represent the true relation.[18]

Figures 1 and 2 show that when the interest differentials take values lower than approximately 12% and 14% for Korea and Indonesia, respectively, the conventional wisdom that increasing interest rate help stabilizing spot rate seems to be supported. But when the interest differentials take higher values up to approximately 20% and

[14] Elliot, Rothenberg and Stock's (1996) Dickey-Fuller-GLS$^{\mu}$ test and Choi's (1994) LM test were applied to the spot and forward rates and the interest differentials for both Korea and Indonesia. The results support the presence of a unit root at conventional levels, and hence the theoretical results in previous sections are relevant here. Prior to estimating the STR model, it is proper to perform linearity tests. But the linearity tests for models with $I(1)$ variables are not yet available, so we bypass the stage of hypothesis testing.

[15] We tried these values, because the NLLS estimates of γ without leads-and-lags take small values.

[16] These results are based on the assumptions that the error terms in equation (21) is $I(0)$ and that regressors are not cointegrated. Formal tests for cointegration for the STR model are not yet available. But fitting the AR(1) regression for the residuals from equation (21) using the values of γ in Table 3 and the two-step Gauss-Newton estimator, we obtained AR(1) coefficient estimates 0.53 and 0.33; and corresponding t-ratios -10.9 and -14.3 for Korea and Indonesia, respectively (for all the values of K in Table 3). These results were also found to be robust to other values of γ. Thus, it seems unlikely that the residuals are $I(1)$. In addition, we tested for cointegration between the future rates and interest differentials, but found no evidence of cointegration.

[17] Indonesia's maximum call rate during the sample period was 91.5%, and the average 29.4%. But the maximum and average for Korea were 35% and 15.6%, respectively.

[18] Ignoring the error term in equation (21) and assuming that the parameter estimates are the true parameter values, the elasticity was calculated by using the formula $\frac{\partial y_t}{\frac{1}{x_{2t}}\partial x_{2t}} = \frac{\partial y_t}{\partial \ln(x_{2t})}$. Here the partial derivative is multiplied by x_{2t}, because log was taken for the spot rate but not for the interest differential.

18% for Korea and Indonesia, respectively, the elasticities become positive which implies that increasing interest rate has negative effects on stabilizing spot rate. When the interest differentials are above 20% and 18% for Korea and Indonesia, respectively, the elasticities become negative again.

Figures 1 and 2 partially support the interest Laffer curve hypothesis. But they also indicate that tight monetary policy is effective, though very weakly, when interest rates are very high. Notwithstanding this remark, we conclude from the magnitudes of the elasticities shown in Figures 1 and 2 that the effects of interest rate on spot rate are negligible in either direction.[19] For example, when the interest differential is 25% for Korea, the elasticity is only -.012. This implies that rasing the interest differential from 25% to 27.5% (i.e., 10% increase) has the effects of appreciating the Korean currency only by .12%. Considering the fact that the currency was depreciated approximately by 30% on the average during the sampling period, such meager effect is certainly unsatisfactory to the Korean economy. This is more so when one considers the negative effects of the interest raise of such magnitude on the corporations and banking system of the economy.

6 Further Remarks

We have analyzed and applied the cointegrated STR model in this paper. However, there are a couple of topics which deserve our attention but were not studied in this paper. First, methods for testing linearity in the presence of $I(1)$ variables are not yet available but useful for empirical analyses. Because nonlinear models are flexible, they may give a good in-sample fit even when the true model is linear. Thus, testing linearity prior to nonlinear model fitting is important. Second, testing for cointegration for the STR model should precede estimation, but relevant methods are not yet available. We hope that these topics can be studied in the future by the authors and other researchers.

[19] Choi and Park (2000) also report that interest differential did not cause spot rate both in the short and long runs during the Asian currency crisis.

Table 1: Biases and Root Mean Squared Errors

Notes: (i) GN1 and GN2 denote the one-step and two-step Gauss-Newton estimators, respectively.
(ii) The numbers of replications at T_0=150 and T_0=300 were 5,000 and 3,000, respectively.
(iii) Parameter ω signifies the degree of serial and contemporaneous correlation in the regressors and errors. Larger ω implies that the regressors and errors are more correlated both serially and contemporaneously. Parameter K denotes the numbers of leads and lags.

(1) $T_0 = 150$

	Estimator	α		δ		c	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
$\omega = 0.2$	NLLS	.048	.122	-.026	.113	.029	.421
	GN1 (K=1)	-.008	.103	.007	.098	-.009	.339
	GN1 (K=2)	-.005	.106	.005	.100	-.007	.344
	GN1 (K=3)	-.005	.109	.005	.103	-.007	.349
	GN2 (K=1)	-.010	.103	.010	.098	-.010	.330
	GN2 (K=2)	-.007	.106	.008	.100	-.008	.335
	GN2 (K=3)	-.007	.110	.008	.103	-.007	.341
$\omega = 0.5$	NLLS	.060	.143	-.033	.136	.008	.598
	GN1 (K=1)	-.010	.115	.007	.113	-.005	.463
	GN1 (K=2)	.002	.121	.001	.119	.001	.469
	GN1 (K=3)	-.003	.127	.003	.125	-.000	.474
	GN2 (K=1)	-.015	.112	.012	.111	-.012	.451
	GN2 (K=2)	-.003	.114	.006	.113	-.006	.456
	GN2 (K=3)	-.008	.118	.008	.115	-.007	.462
$\omega = 0.8$	NLLS	.068	.159	-.036	.153	-.048	.871
	GN1 (K=1)	-.012	.121	.007	.123	-.033	.644
	GN1 (K=2)	.005	.132	-.002	.133	-.025	.657
	GN1 (K=3)	-.006	.138	.003	.139	-.029	.663
	GN2 (K=1)	-.019	.119	.014	.121	-.044	.600
	GN2 (K=2)	-.001	.121	.005	.123	-.034	.618
	GN2 (K=3)	-.013	.124	.011	.126	-.042	.620

(2) $T_0 = 300$

	Estimator	α		δ		c	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
$\omega = 0.2$	NLLS	.030	.067	-.017	.064	-.004	.336
	GN1 (K=1)	-.003	.053	.003	.054	-.009	.264
	GN1 (K=2)	-.001	.053	.001	.054	-.008	.266
	GN1 (K=3)	-.001	.054	.002	.055	-.008	.267
	GN2 (K=1)	-.003	.053	.003	.054	-.009	.251
	GN2 (K=2)	-.002	.053	.002	.054	-.009	.253
	GN2 (K=3)	-.002	.054	.002	.055	-.008	.254
$\omega = 0.5$	NLLS	.033	.078	-.019	.078	-.036	.486
	GN1 (K=1)	-.006	.058	.004	.061	-.012	.366
	GN1 (K=2)	.000	.058	.001	.061	-.013	.370
	GN1 (K=3)	-.003	.058	.002	.061	-.012	.371
	GN2 (K=1)	-.007	.059	.005	.061	-.014	.345
	GN2 (K=2)	-.001	.058	.002	.061	-.015	.352
	GN2 (K=3)	-.004	.059	.003	.062	-.015	.351
$\omega = 0.8$	NLLS	.038	.078	-.021	.078	-.046	.636
	GN1 (K=1)	-.006	.056	.003	.060	-.006	.464
	GN1 (K=2)	.003	.056	-.002	.060	-.003	.468
	GN1 (K=3)	-.003	.056	.001	.061	-.004	.469
	GN2 (K=1)	-.007	.056	.005	.060	-.014	.432
	GN2 (K=2)	.002	.055	-.000	.059	-.009	.436
	GN2 (K=3)	-.004	.056	.003	.060	-.011	.438

Table 2: Empirical Sizes of the T-ratios

Notes: (i) The same experimental format as for Table 1 was used.
(ii) The long-run variance was estimated by using Andrews' (1991) methods with an AR(4) approximation for the prefilter.

(1) $T_0 = 150$

	Estimator	α		δ		c	
		5%	10%	5%	10%	5%	10%
	GN1 (K=1)	.089	.152	.084	.143	.075	.129
	GN1 (K=2)	.090	.156	.085	.148	.080	.132
$\omega = 0.2$	GN1 (K=3)	.096	.161	.090	.159	.084	.139
	GN2 (K=1)	.087	.148	.082	.142	.077	.135
	GN2 (K=2)	.089	.153	.086	.148	.083	.139
	GN2 (K=3)	.097	.157	.090	.157	.085	.143
	GN1 (K=1)	.086	.147	.093	.151	.077	.134
	GN1 (K=2)	.089	.152	.096	.154	.078	.134
$\omega = 0.5$	GN1 (K=3)	.094	.155	.100	.157	.081	.134
	GN2 (K=1)	.082	.143	.088	.149	.079	.141
	GN2 (K=2)	.088	.148	.092	.148	.082	.141
	GN2 (K=3)	.091	.154	.096	.154	.085	.146
	GN1 (K=1)	.089	.155	.089	.149	.085	.138
	GN1 (K=2)	.095	.152	.093	.149	.087	.142
$\omega = 0.8$	GN1 (K=3)	.098	.161	.096	.158	.090	.140
	GN2 (K=1)	.088	.154	.084	.150	.089	.141
	GN2 (K=2)	.090	.149	.091	.147	.086	.142
	GN2 (K=3)	.095	.159	.094	.155	.091	.145

(2) $T_0 = 300$

	Estimator	α		δ		c	
		5%	10%	5%	10%	5%	10%
$\omega = 0.2$	GN1 (K=1)	.069	.128	.065	.123	.064	.111
	GN1 (K=2)	.068	.132	.067	.126	.064	.117
	GN1 (K=3)	.069	.131	.066	.127	.064	.117
	GN2 (K=1)	.067	.129	.065	.119	.068	.117
	GN2 (K=2)	.065	.130	.066	.123	.068	.118
	GN2 (K=3)	.066	.133	.065	.124	.067	.121
$\omega = 0.5$	GN1 (K=1)	.078	.137	.075	.126	.061	.118
	GN1 (K=2)	.076	.132	.073	.127	.068	.114
	GN1 (K=3)	.077	.132	.073	.132	.067	.122
	GN2 (K=1)	.081	.137	.078	.127	.064	.119
	GN2 (K=2)	.077	.131	.073	.124	.068	.121
	GN2 (K=3)	.078	.133	.075	.131	.069	.122
$\omega = 0.8$	GN1 (K=1)	.079	.138	.070	.124	.063	.121
	GN1 (K=2)	.072	.130	.067	.122	.060	.118
	GN1 (K=3)	.075	.136	.069	.124	.060	.121
	GN2 (K=1)	.081	.142	.067	.128	.065	.120
	GN2 (K=2)	.072	.129	.065	.123	.064	.120
	GN2 (K=3)	.075	.134	.068	.127	.067	.125

Table 3: Two-Step Gauss-Newton Estimation Results

Notes: (i) Daily data covering the periods 4/1/1997 - 10/30/1998 and 1/3/1997-7/24/1998 were used for Korea and Indonesia, respectively.

(ii) Parameter K denotes the numbers of leads and lags.

(iii) The numbers in parentheses denote t-ratios.

(iv) (*): significant at the 5% level; (**): significant at the 1% level.

(1) Korea ($T_0 = 343$)

		α_1	α_2	δ	c
$\gamma = .6$	$K = 1$	.9762 (56.2**)	-.0045 (-2.48**)	.0039 (2.54**)	15.0
	$K = 2$	.9737 (55.2**)	-.0045 (-2.44**)	.0040 (2.56**)	14.8
	$K = 3$	.9722 (57.4**)	-.0045 (-2.52**)	.0041 (2.68**)	14.6

(2) Indonesia ($T_0 = 406$)

		α_1	α_2	δ	c
$\gamma = .5$	$K = 1$	1.007 (299**)	-.0016 (-2.72**)	.0008 (1.48)	14.8
	$K = 2$	1.007 (300**)	-.0016 (-2.80**)	.0008 (1.51)	15.0
	$K = 3$	1.008 (296**)	-.0016 (-2.85**)	.0008 (1.55)	15.2

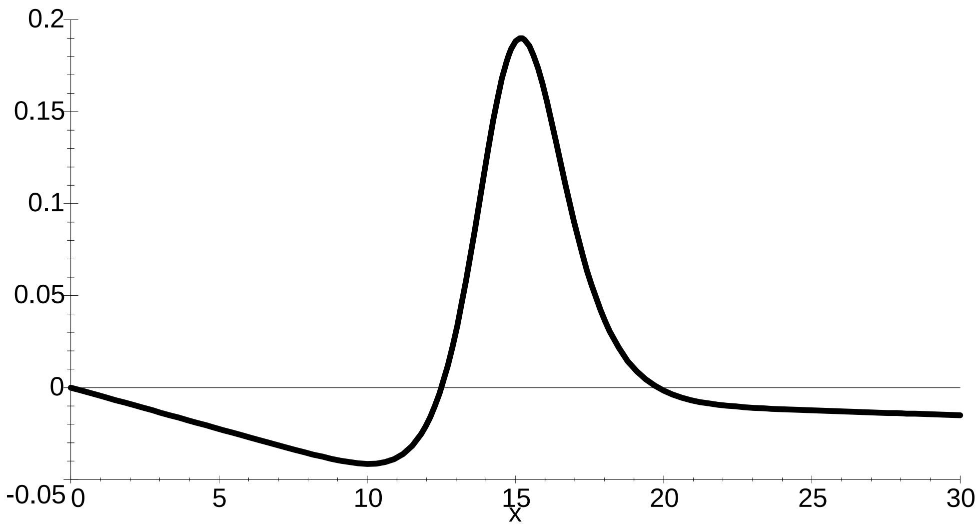

Figure 1: Interest Elasticity of Spot Rate (Korea)

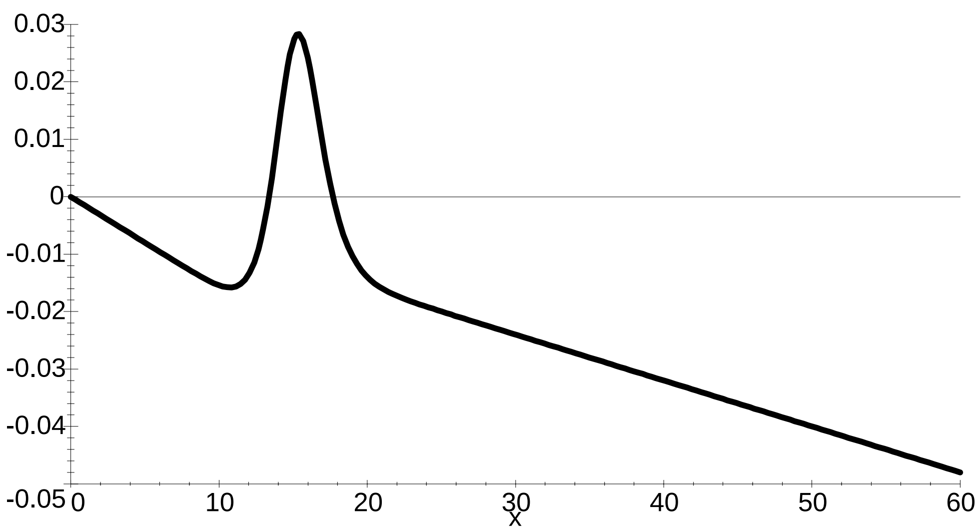

Figure 2: Interest Elasticity of Spot Rate (Indonesia)

7 Appendix I: Auxiliary Lemmas

We shall first prove some auxiliary results which may also have applications elsewhere. Recall the notation $N = T - 2K$ and note that a (possibly) matrix-valued function $h(x)$ defined on R^d is said to be locally bounded if $\|h(x)\|$ is bounded on compact subsets of R^d.

Lemma 1 *Let $h(x)$ be a locally bounded, vector-valued function defined on R^d $(d < \infty)$ and let $\{\epsilon_t, \mathcal{F}_t^\epsilon\}$ be a square integrable martingale difference sequence such that $\sup_t E\|\epsilon_t\|^2 < \infty$. Let $\zeta_{tT}^{(1)}$ $(d \times 1)$ and $\zeta_{tT}^{(2)}$ $(t = 1, \ldots, T)$ be random vectors defined on the same probability space as ϵ_t. Assume that $\max_{1\le t\le T}\left\|\zeta_{tT}^{(1)}\right\| = O_p(1)$ and $\sup_{t,T} E\left\|\zeta_{tT}^{(2)}\right\| < \infty$. Then,*

$$(i)\quad \max_{1\le t\le T}\left\|h(\zeta_{tT}^{(1)})\right\| = O_p(1)$$

$$(ii)\quad N^{-3/2}\sum_{j=-K}^{K}\sum_{t=K+1}^{T-K}\left\|h(\zeta_{tT}^{(1)})\right\|\left\|\zeta_{t+j,T}^{(2)}\right\| = O_p\left(K/N^{1/2}\right)$$

and

$$(iii)\quad \sum_{j=1}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K} h(\zeta_{tT}^{(1)})\epsilon_{t+j}'\right\| = O_p\left(K/N^{1/2}\right)$$

when $\zeta_{tT}^{(1)}$ is measurable with respect to the $\sigma-$algebra $\mathcal{F}_t^\epsilon$. The third result also holds with $\zeta_{tT}^{(1)}$ replaced by $\zeta_{t+j-1,T}^{(1)}$.

Proof To prove the first assertion, let $\varepsilon > 0$ and use the assumption $\max_{1\le t\le T}\left\|\zeta_{tT}^{(1)}\right\| = O_p(1)$ to choose $m > 0$ such that $P\left\{\max_{1\le t\le T}\left\|\zeta_{tT}^{(1)}\right\| > m\right\} < \varepsilon$ for all T large. Next, use the assumption that $h(x)$ is locally bounded to conclude that $H_m = \sup_{\|x\|\le m}\|h(x)\|$ is finite. Then, the desired result follows because for all T large

$$P\left\{\max_{1\le t\le T}\left\|h(\zeta_{tT}^{(1)})\right\| > H_m\right\} \le P\left\{\max_{1\le t\le T}\left\|\zeta_{tT}^{(1)}\right\| > m\right\} < \varepsilon.$$

The second result is an immediate consequence of the first result and the moment condition imposed on $\zeta_{tT}^{(2)}$. To prove the third assertion, first note that an application of the triangular inequality yields

$$\begin{aligned}
&\sum_{j=1}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K} h(\zeta_{tT}^{(1)})\epsilon_{t+j}'\right\| \\
\le\ &\sum_{j=1}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K} \mathbf{1}\left\{\left\|\zeta_{tT}^{(1)}\right\| \le m\right\} h(\zeta_{tT}^{(1)})\epsilon_{t+j}'\right\| \\
&+\sum_{j=1}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K} \mathbf{1}\left\{\left\|\zeta_{tT}^{(1)}\right\| > m\right\} h(\zeta_{tT}^{(1)})\epsilon_{t+j}'\right\| \\
&\overset{def}{=} A_{1T} + A_{2T}.
\end{aligned}$$

Now, let $\varepsilon > 0$ and define m and H_m in the same way as in the proof of (i). Then, for every $M > 0$ and T large

$$P\left\{\left(N^{1/2}/K\right)|A_{2T}| > M/2\right\} \leq P\left\{\max_{1\leq t\leq T}\left\|\zeta_{tT}^{(1)}\right\| > m\right\} < \varepsilon.$$

As for A_{1T}, use the assumptions that $\{\epsilon_t, \mathcal{F}_t^{\epsilon}\}$ is a square integrable martingale difference sequence and that $\zeta_{tT}^{(1)}$ is measurable with respect to the $\sigma-$algebra $\mathcal{F}_t^{\epsilon}$ to obtain

$$\begin{aligned}
E|A_{1T}| &\leq \sum_{j=1}^{K}\left(E\left\|N^{-1}\sum_{t=K+1}^{T-K}\mathbf{1}\left\{\left\|\zeta_{tT}^{(1)}\right\|\leq m\right\}h(\zeta_{tT}^{(1)})\epsilon_{t+j}'\right\|^2\right)^{1/2} \\
&= \sum_{j=1}^{K}\left(N^{-2}\sum_{t=K+1}^{T-K}E\left(\mathbf{1}\left\{\left\|\zeta_{tT}^{(1)}\right\|\leq m\right\}h(\zeta_{tT}^{(1)})'h(\zeta_{tT}^{(1)})\epsilon_{t+j}'\epsilon_{t+j}\right)\right)^{1/2} \\
&\leq H_m\sum_{j=1}^{K}\left(N^{-2}\sum_{t=K+1}^{T-K}E\left(\epsilon_{t+j}'\epsilon_{t+j}\right)\right)^{1/2} \\
&\leq CH_mK/N^{1/2}, \quad C<\infty.
\end{aligned}$$

Hence, $P\left\{\left(N^{1/2}/K\right)|A_{1T}| > M/2\right\} \leq 2CH_m/M$ by Markov's inequality and we can conclude that for every M and T large

$$\begin{aligned}
P\left\{\left(N^{1/2}/K\right)|A_{1T}+A_{2T}| > M\right\} &\leq P\left\{\left(N^{1/2}/K\right)|A_{1T}| > M/2\right\} \\
&\quad +P\left\{\left(N^{1/2}/K\right)|A_{2T}| > M/2\right\} \\
&< 2CH_m/M+\varepsilon.
\end{aligned}$$

For $M > 2CH_m/\varepsilon$ the last expression is smaller than 2ε, which proves the stated result. A similar proof shows the final assertion. ¥

Note that the first two results of Lemma 1 obviously hold when $h(x)$ and $\zeta_{t+j,T}^{(2)}$ are matrix-valued and that the third result improves Lemma A.4(c) of Park and Phillips (2000) by relaxing the exponentially boundedness assumption used therein to local boundedness.

The first two results of Lemma 1 can be applied with the process

$$z_t = z_{t-1} + w_t, \quad t = 1, 2, \ldots \tag{A.1}$$

where w_t is as in Assumption 2 and z_0 may be any random vector such that $E\|z_0\|^2 < \infty$. In this case $\zeta_{tT}^{(1)} = z_{tT} = T^{-1/2}z_t$ and $\max_{1\leq t\leq T}\|z_{tT}\| = O_p(1)$ is an immediate consequence of the invariance principle (8). This definition of z_{tT} will be assumed in subsequent lemmas. The proofs of these lemmas make use of the fact that, due to Assumption 2, we can write

$$w_t = \eta_t - \Delta\xi_t \tag{A.2}$$

where

$$\eta_t = \sum_{j=0}^{\infty} \left(E_t w_{t+j} - E_{t-1} w_{t+j} \right) \quad \text{and} \quad \xi_t = \sum_{j=1}^{\infty} E_t w_{t+j}$$

with E_t the conditional expectation operator with respect to the $\sigma-$algebra $\mathcal{F}_t = \sigma\left(w_s, s \leq t\right)$ (cf. Hansen, 1992). Since $\{\eta_t, \mathcal{F}_t\}$ is a stationary martingale difference sequence equation (A.2) is analogous to the so-called Beveridge-Nelson decomposition which has been used extensively in asymptotic analysis of linear processes (see e.g. Phillips and Solo, 1992). Therefore, we shall refer to equation (A.2) as the Beveridge-Nelson decomposition also in the present context. In our applications of the third result of Lemma 1 the martingale difference sequence ϵ_t will be η_t. For these applications, as well as other subsequent derivations, it is worth noting that the (stationary) processes η_t and ξ_t have finite moments of order 4 (see the proof of Theorem 3.1 of Hansen (1992)).

Lemma 2 *Let $h(x;\theta)$ be a (possibly) vector valued continuously differentiable function defined on $\mathsf{R}^{p+1} \times \Theta^*$ where Θ^* is an open set in an Euclidean space. Let $\Theta \subset \Theta^*$ be a compact set containing the point θ_0 in its interior and assume that $K^2/T \longrightarrow 0$. Then,*

$$(i) \quad \sum_{j=-K}^{K} \left\| N^{-1} \sum_{t=K+1}^{T-K} h\left(z_{tT};\theta\right) w'_{t+j} \right\| = O_p\left(K/N^{1/2}\right) \quad \textit{for every fixed } \theta \in \Theta$$

and

$$(ii) \quad \sum_{j=-K}^{K} \left\| N^{-1} \sum_{t=K+1}^{T-K} h(z_{tT};\dot{\theta}_T) w'_{t+j} \right\| = O_p\left(K\right) \left\| \dot{\theta}_T - \theta_0 \right\| + O_p\left(K/N^{1/2}\right)$$

where $\dot{\theta}_T$ is a random vector such that $\dot{\theta}_T = \theta_0 + o_p(1)$.

Proof We shall first prove the latter assertion and then note how the first one can be obtained from the employed arguments. Without loss of generality, assume that $h(x;\theta)$ is real-valued and use the Beveridge-Nelson decomposition (A.2) in conjunction with the triangular inequality to obtain

$$\begin{aligned} \sum_{j=-K}^{K} \left\| N^{-1} \sum_{t=K+1}^{T-K} h(z_{tT};\dot{\theta}_T) w'_{t+j} \right\| &\leq \sum_{j=-K}^{K} \left\| N^{-1} \sum_{t=K+1}^{T-K} h(z_{tT};\dot{\theta}_T) \eta'_{t+j} \right\| \\ &\quad + \sum_{j=-K}^{K} \left\| N^{-1} \sum_{t=K+1}^{T-K} h(z_{tT};\dot{\theta}_T) \Delta\xi'_{t+j} \right\| \\ &\overset{def}{=} A_{3T}(\dot{\theta}_T) + A_{4T}(\dot{\theta}_T). \end{aligned}$$

First, consider $A_{4T}(\dot{\theta}_T)$ and use partial summation to obtain

$$N^{-1} \sum_{t=K+1}^{T-K} h(z_{tT};\dot{\theta}_T) \Delta\xi'_{t+j} \quad = \quad N^{-1} h(z_{T-K,T};\dot{\theta}_T) \xi'_{T-K+j} - N^{-1} h(z_{K,T};\dot{\theta}_T) \xi'_{K+j}$$

$$-N^{-1}\sum_{t=K+1}^{T-K}\left[h(z_{tT};\dot{\theta}_T)-h(z_{t-1,T};\dot{\theta}_T)\right]\xi_{t-1+j}'.$$

Hence, using the triangular inequality we find that

$$\begin{aligned}\left|A_{4T}(\dot{\theta}_T)\right| &\leq N^{-1}\sup_{\theta\in\Theta}\left\|h\left(z_{T-K,T};\theta\right)\right\|\sum_{j=-K}^{K}\left\|\xi_{T-K+j}\right\| \\ &\quad +N^{-1}\sup_{\theta\in\Theta}\left\|h\left(z_{K,T};\theta\right)\right\|\sum_{j=-K}^{K}\left\|\xi_{K+j}\right\| \\ &\quad +\sup_{\theta\in\Theta}\sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K}\left[h\left(z_{tT};\theta\right)-h\left(z_{t-1,T};\theta\right)\right]\xi_{t-1+j}'\right\|.\end{aligned}$$

Since $\sup_{\theta\in\Theta}\left\|h\left(x;\theta\right)\right\|$ is locally bounded, the first two terms on the right hand side are easily seen to be of order $O_p\left(K/N\right)$. For the third term we can use a standard mean value expansion to get

$$h\left(z_{tT};\theta\right)-h\left(z_{t-1,T};\theta\right)=T^{-1/2}H_1\left(\bar{z}_{t-1,T};\theta\right)w_t$$

where $H_1\left(x;\theta\right)=\partial h\left(x;\theta\right)/\partial x'$ and $\left\|\bar{z}_{t-1,T}-z_{tT}\right\|\leq\left\|z_{t-1,T}-z_{tT}\right\|=T^{-1/2}\left\|w_t\right\|.$ Thus, we can write

$$\begin{aligned}\left|A_{4T}(\dot{\theta}_T)\right| &\leq \sup_{\theta\in\Theta}\sum_{j=-K}^{K}\left\|N^{-3/2}\sum_{t=K+1}^{T-K}H_1\left(\bar{z}_{t-1,T};\theta\right)w_t\xi_{t-1+j}'\right\|+O_p\left(K/N\right) \\ &\leq N^{-3/2}\sum_{j=-K}^{K}\sum_{t=K+1}^{T-K}\sup_{\theta\in\Theta}\left\|H_1\left(\bar{z}_{t-1,T};\theta\right)\right\|\left\|w_t\xi_{t-1+j}'\right\|+O_p\left(K/N\right). \\ &= O_p\left(K/N^{1/2}\right). \qquad \text{(A.3)}\end{aligned}$$

Here the latter inequality is justified by the triangular inequality whereas the equality follows from Lemma 1(ii) because $\sup_{\theta\in\Theta}\left\|H_1\left(x;\theta\right)\right\|$ is locally bounded, $\max_{1\leq t\leq T}\left\|\bar{z}_{t-1,T}\right\|=O_p\left(1\right)$, and $E\left\|w_t\xi_{t-1+j}'\right\|$ is a finite constant. For later purposes we note that above we actually showed that $A_{4T}\left(\theta\right)=O_p\left(K/N^{1/2}\right)$ holds uniformly in $\theta\in\Theta$.

Next, consider $A_{3T}(\dot{\theta}_T)$. Since θ_0 is an interior point of Θ and $\dot{\theta}_T=\theta_0+o_p\left(1\right)$, we can use the mean value expansion

$$h(z_{tT};\dot{\theta}_T)=h\left(z_{tT};\theta_0\right)+H_2\left(z_{tT};\bar{\theta}_T\right)\left(\dot{\theta}_T-\theta_0\right)$$

where $H_2\left(x;\theta\right)=\partial h\left(x;\theta\right)/\partial\theta'$ and $\left\|\bar{\theta}_T-\theta_0\right\|\leq\left\|\dot{\theta}_T-\theta_0\right\|$. Thus, using the triangular inequality one obtains

$$\begin{aligned}\left|A_{3T}(\dot{\theta}_T)\right| &\leq \sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K}h\left(z_{tT};\theta_0\right)\eta_{t+j}'\right\| \\ &\quad +\sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K}H_2\left(z_{tT};\bar{\theta}_T\right)(\dot{\theta}_T-\theta_0)\eta_{t+j}'\right\|.\end{aligned}$$

The first term on the right hand side is $A_{3T}(\theta_0)$, and the second term can be bounded by

$$N^{-1}\sum_{j=-K}^{K}\sum_{t=K+1}^{T-K}\sup_{\theta\in\Theta}\|H_2(z_{tT};\theta)\|\left\|\dot{\theta}_T-\theta_0\right\|\|\eta_{t+j}\|=O_p(K)\left\|\dot{\theta}_T-\theta_0\right\|.$$

Here the equality is again obtained from Lemma 1(ii) because $\sup_{\theta\in\Theta}\|H_2(x;\theta)\|$ is locally bounded, $\max_{1\leq t\leq T}\|z_{tT}\|=O_p(1)$, and $E\left\|\eta_{t+j}\right\|$ is constant. Thus, to complete the proof, we have to show that $A_{3T}(\theta_0)=O_p\left(K/N^{1/2}\right)$.

By the definition of $A_{3T}(\theta_0)$,

$$\begin{aligned}A_{3T}(\theta_0) &= \sum_{j=1}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K}h(z_{tT};\theta_0)\eta_{t+j}'\right\|+\sum_{j=0}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K}h(z_{tT};\theta_0)\eta_{t-j}'\right\| \\ &\overset{def}{=} A_{31T}(\theta_0)+A_{32T}(\theta_0).\end{aligned}$$

Lemma 1(iii) implies that $A_{31T}(\theta_0)=O_p\left(K/N^{1/2}\right)$, so we need to show that the same holds true for $A_{32T}(\theta_0)$. To this end, use the Beveridge-Nelson decomposition (A.2) and the definition of z_{tT} to give

$$z_{tT}=s_{tT}-T^{-1/2}\xi_t+T^{-1/2}(\xi_0-z_0)$$

where $s_{tT}=T^{-1/2}\sum_{j=1}^{t}\eta_j$. Thus, a mean value expansion yields

$$h(z_{tT};\theta)=h(s_{t-j-1,T};\theta)+T^{-1/2}H_1(\bar{s}_{t-j-1,T};\theta)r_{tj}$$

where $r_{tj}=\sum_{i=0}^{j}\eta_{t-j+i}-\xi_t+\xi_0+z_0$ and $\|\bar{s}_{t-j-1,T}-z_{tT}\|\leq\|s_{t-j-1,T}-z_{tT}\|=T^{-1/2}\|r_{tj}\|$. This identity and the triangular inequality imply

$$\begin{aligned}|A_{32T}(\theta_0)| &\leq \sum_{j=0}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K}h(s_{t-j-1,T};\theta_0)\eta_{t-j}'\right\| \\ &\quad +\sum_{j=0}^{K}\left\|N^{-3/2}\sum_{t=K+1}^{T-K}H_1(\bar{s}_{t-j-1,T};\theta_0)r_{tj}\eta_{t-j}'\right\| \\ &= \sum_{j=0}^{K}\left\|N^{-3/2}\sum_{t=K+1}^{T-K}H_1(\bar{s}_{t-j-1,T};\theta_0)r_{tj}\eta_{t-j}'\right\|+O_p(K/N^{1/2}).\end{aligned}$$

Here the equality is obtained from Lemma 1(iii) which obviously applies despite the differences in subscripts. To analyze the first term in the last expression, it suffices to replace r_{jt} in turn by each of the four components in its definition. Thus, consider the quantity

$$\sum_{j=0}^{K}\left\|N^{-3/2}\sum_{t=K+1}^{T-K}H_1(\bar{s}_{t-j-1,T};\theta_0)\sum_{i=0}^{j}\eta_{t-j+i}\eta_{t-j}'\right\|$$

$$\leq \sum_{j=0}^{K} \left\| N^{-3/2} \sum_{t=K+1}^{T-K} H_1\left(\bar{s}_{t-j-1,T};\theta_0\right) \eta_{t-j}\eta_{t-j}' \right\| + \sum_{j=1}^{K} \left\| N^{-3/2} \sum_{t=K+1}^{T-K} H_1\left(\bar{s}_{t-j-1,T};\theta_0\right) \sum_{i=1}^{j} \eta_{t-j+i}\eta_{t-j}' \right\|.$$

Arguments similar to those used for $A_{4T}(\dot{\theta}_T)$ in (A.3) show that the first term on the right hand side is of order $O_p\left(K/N^{1/2}\right)$. These arguments also apply when the last three terms in the definition of r_{tj} are considered. Thus, to complete the proof we only need to show that the latter term in the last expression is of order $O_p\left(K/N^{1/2}\right)$. Using the triangular inequality, one obtains

$$\sum_{j=1}^{K} \left\| N^{-3/2} \sum_{t=K+1}^{T-K} H_1\left(\bar{s}_{t-j-1,T};\theta_0\right) \sum_{i=1}^{j} \eta_{t-j+i}\eta_{t-j}' \right\| \leq \sum_{j=1}^{K}\sum_{i=1}^{j} \left\| N^{-3/2} \sum_{t=K+1}^{T-K} H_1\left(\bar{s}_{t-j-1,T};\theta_0\right) \eta_{t-j+i}\eta_{t-j}' \right\|. \tag{A.4}$$

To show that the last quantity is of order $O_p\left(K/N^{1/2}\right)$, we can make use of a similar truncation argument as in the proof of Lemma 1(iii) and replace the function $H_1(x;\theta_0)$ by $\mathbf{1}\{\|x\| \leq m\} H_1(x;\theta_0)$ with an appropriately chosen real number m. Thus, since $H_1(x;\theta_0)$ is locally bounded $\mathbf{1}\{\|x\| \leq m\} H_1(x;\theta_0)$ is bounded and we can proceed by assuming that the function $H_1(x;\theta_0)$ itself is bounded. Assuming this shows that for $i \geq 1$

$$\begin{aligned} & E\left\| N^{-3/2} \sum_{t=K+1}^{T-K} H_1\left(\bar{s}_{t-j-1,T};\theta_0\right) \eta_{t-j+i}\eta_{t-j}' \right\| \\ \leq & \left(E\left\| N^{-3/2} \sum_{t=K+1}^{T-K} H_1\left(\bar{s}_{t-j-1,T};\theta_0\right) \eta_{t-j+i}\eta_{t-j}' \right\|^2 \right)^{1/2} \\ = & O\left(N^{-1}\right) \end{aligned}$$

where the equality follows because the terms in the preceding sum are uncorrelated with bounded second moments. Thus, the right hand side of (A.4) is of order $O_p\left(K^2/N\right)$, which proves the desired result and completes the proof of the second assertion.

To prove the first assertion, notice that we need to show that $A_{3T}(\theta)$ and $A_{4T}(\theta)$ are of order $O_p\left(K/N^{1/2}\right)$ for every fixed θ. For $A_{4T}(\theta)$ we showed that this holds even uniformly in θ. As for $A_{3T}(\theta)$, it suffices to consider $A_{31T}(\theta)$ and $A_{32T}(\theta)$ separately. In the above proof we showed that $A_{31T}(\theta_0)$ and $A_{32T}(\theta_0)$ are of order $O_p\left(K/N^{1/2}\right)$ and an inspection of the proof reveals that θ_0 can be replaced with any $\theta \in \Theta$ without changing the result. This completes the proof of Lemma 2. ¥

It would be useful to be able to show that the pointwise result of Lemma 2(i) also holds uniformly in θ but we have been unable to obtain this extension. The following result is not difficult to obtain, however.

Lemma 3 *Suppose the assumptions of Lemma 2 hold and let $R_T = [R_{-KT} \ \cdots \ R_{KT}]$ be a (possibly) stochastic matrix with* $p+1$ *rows and such that, for some finite constant* c, $\|R_T\| \leq c$ *(a.s.). Then,*

$$\sup_{\theta\in\Theta}\left\|\sum_{j=-K}^{K} N^{-1}\sum_{t=K+1}^{T-K} h\left(z_{tT};\theta\right) w'_{t+j} R_{jT}\right\| = o_p\left(1\right).$$

Proof Without loss of generality assume that $c = 1$ and that $h(x;\theta)$ is real valued and R_{jT} is a vector. Since $\|R_{jT}\| \leq 1$ for all j, we have for every fixed $\theta \in \Theta$

$$\begin{aligned}\left\|\sum_{j=-K}^{K} N^{-1}\sum_{t=K+1}^{T-K} h\left(z_{tT};\theta\right) w'_{t+j} R_{jT}\right\| &\leq \sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K} h\left(z_{tT};\theta\right) w'_{t+j} R_{jT}\right\| \\ &\leq \sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K} h\left(z_{tT};\theta\right) w'_{t+j}\right\| \\ &= o_p\left(1\right)\end{aligned}$$

where the equality is due to Lemma 2(i). Thus, the problem is to strengthen this pointwise convergence in probability to uniform convergence in probability. Since Θ is a compact set it suffices to show that the quantity whose norm is taken is stochastically equicontinuous (see e.g. Davidson, 1994, p. 337). To this end, let θ_1 and θ_2 be arbitrary points of Θ and consider the quantity

$$\begin{aligned}&\left\|\sum_{j=-K}^{K} N^{-1}\sum_{t=K+1}^{T-K} h\left(z_{tT};\theta_1\right) w'_{t+j} R_{jT} - \sum_{j=-K}^{K} N^{-1}\sum_{t=K+1}^{T-K} h\left(z_{tT};\theta_2\right) w'_{t+j} R_{jT}\right\| \\ =& \left\|N^{-1}\sum_{t=K+1}^{T-K}\left[h\left(z_{tT};\theta_1\right) - h\left(z_{tT};\theta_2\right)\right]\sum_{j=-K}^{K} w'_{t+j} R_{jT}\right\| \qquad (A.5) \\ \leq& \left(N^{-1}\sum_{t=K+1}^{T-K}\left\|h\left(z_{tT};\theta_1\right) - h\left(z_{tT};\theta_2\right)\right\|^2\right)^{1/2}\left(N^{-1}\sum_{t=K+1}^{T-K}\left\|\sum_{j=-K}^{K} w'_{t+j} R_{jT}\right\|^2\right)^{1/2}\end{aligned}$$

where the inequality follows from the Cauchy-Schwarz inequality. For the difference in the last expression we can use the mean value expansion

$$h\left(z_{tT};\theta_1\right) - h\left(z_{tT};\theta_2\right) = H_2\left(z_{tT};\bar{\theta}\right)\left(\theta_1 - \theta_2\right)$$

where $H_2(x;\theta) = \partial h(x;\theta)/\partial\theta'$ and $\left\|\bar{\theta} - \theta_1\right\| \leq \|\theta_1 - \theta_2\|$. Thus,

$$\begin{aligned}&\left(N^{-1}\sum_{t=K+1}^{T-K}\left\|h\left(z_{tT};\theta_1\right) - h\left(z_{tT};\theta_2\right)\right\|^2\right)^{1/2} \\ \leq& \|\theta_1 - \theta_2\|\left(N^{-1}\sum_{t=K+1}^{T-K}\sup_{\theta\in\Theta}\left\|H_2\left(z_{tT};\theta\right)\right\|^2\right)^{1/2} \\ =& \|\theta_1 - \theta_2\|\, O_p\left(1\right)\end{aligned}$$

where the equality is justified by Lemma 1(i) because $\sup_{\theta \in \Theta} \|H_2(x;\theta)\|^2$ is locally bounded and $\max_{1 \le t \le T} \|z_{tT}\| = O_p(1)$. Hence, the desired stochastic equicontinuity follows in a straightforward manner from (A.5) if we show that the latter factor in the last expression therein is of order $O_p(1)$. To see this, define the matrix

$$\Gamma_T = \left[N^{-1} \sum_{t=K+1}^{T-K} w_{t+i} w'_{t+j} \right], \quad i,j = -K,\ldots,K,$$

and let $\lambda_{\max}(\cdot)$ denote the largest eigenvalue of the indicated matrix. With these definitions we have

$$\begin{aligned} \left(N^{-1} \sum_{t=K+1}^{T-K} \left\| \sum_{j=-K}^{K} w'_{t+j} R_{jT} \right\|^2 \right)^{1/2} &= \left(tr\left(R'_T \Gamma_T R_T \right) \right)^{1/2} \\ &\le \left(\lambda_{\max}(\Gamma_T)\, tr\left(R'_T R_T \right) \right)^{1/2} \\ &\le \lambda_{\max}^{1/2}(\Gamma_T) \\ &= O_p(1). \end{aligned}$$

Here the last relation is a straightforward consequence of the fact that the spectral density matrix of the process w_t is bounded and the preceding one follows from the assumption $\|R_T\| \le 1$ (a.s.). Thus, the proof is complete. ¥

The results of Lemmas 2 and 3 also hold with a fixed value of K. In that case R_{jT} in Lemma 3 may be replaced by an identity matrix, as can easily be checked from the given proofs.

In the following lemma we use the notation $C(\Theta)^{a \times b}$ to signify the space of all continuous functions from the compact set Θ to $\mathsf{R}^{a \times b}$ endowed with the uniform metric. In $\mathsf{R}^{a \times b}$ the usual Euclidean metric is assumed.

Lemma 4 *Let $H(x,\theta)$ $(a \times b)$ be a matrix valued continuous function defined on $\mathsf{R}^{p+1} \times \Theta$. Then, if $K/T \to 0$*

$$N^{-1} \sum_{t=K+1}^{T-K} H(z_{tT};\theta) \Rightarrow \int_0^1 H(B(s);\theta)\, ds$$

where the convergence holds in the function space $C(\Theta)^{a \times b}$.

Proof Since $z_{tT} \Rightarrow B(s)$ by (8) the proof can be obtained in the same way as the first result in Theorem 3.1 of Park and Phillips (2000). ¥

Lemma 4 can be used to prove the following.

Lemma 5 *Let $f(x;\theta)$, $\theta \in \Theta$, and x_{tT} be as in Subsection 3.2. Then there exists an $\varepsilon > 0$ such that with probability approaching one*

$$\inf_{\theta \in \Theta} \lambda_{\min} \left(N^{-1} \sum_{t=K+1}^{T-K} f(x_{tT};\theta) f(x_{tT};\theta)' \right) \ge \varepsilon.$$

Proof The stated result follows from condition (11), Lemma 4, and the continuity of eigenvalues and the infimum function. ¥

Lemma 6 *Let $h(x)$ be a vector valued twice continuously differentiable function defined on R^{p+1}. Then,*

$$T^{-1/2}\sum_{t=1}^{T}h(z_{tT})\,w_t' \Rightarrow \int_0^1 h(B(s))\,dB(s)' + \int_0^1 H_1(B(s))\,ds\Lambda$$

where $H_1(x) = \partial h(x)/\partial x'$ and $\Lambda = \sum_{j=0}^{\infty} Ew_0 w_j'$. Moreover, this weak convergence holds jointly with that in (8).

Proof Using the Beveridge-Nelson decomposition (A.2), one obtains

$$T^{-1/2}\sum_{t=1}^{T}h(z_{tT})\,w_t' = T^{-1/2}\sum_{t=1}^{T}h(z_{tT})\,\eta_t' - T^{-1/2}\sum_{t=1}^{T}h(z_{tT})\,\Delta\xi_t'. \qquad \text{(A.6)}$$

First, consider the latter term on the right hand side. By partial summation,

$$\begin{aligned}-T^{-1/2}\sum_{t=1}^{T}h(z_{tT})\,\Delta\xi_t' &= -T^{-1/2}h(z_{TT})\,\xi_T' + T^{-1/2}h(z_{0T})\,\xi_0' \\ &\quad + T^{-1/2}\sum_{t=1}^{T}\left[h(z_{tT}) - h(z_{t-1,T})\right]\xi_{t-1}' \\ &= T^{-1/2}\sum_{t=1}^{T}\left[h(z_{tT}) - h(z_{t-1,T})\right]\xi_{t-1}' + o_p(1)\end{aligned}$$

where the latter equality is an immediate consequence of the assumptions. Thus, a standard mean value expansion and the fact $\Delta z_{tT} = T^{-1/2}w_t$ yield

$$-T^{-1/2}\sum_{t=1}^{T}h(z_{tT})\,\Delta\xi_t' = T^{-1}\sum_{t=1}^{T}H_1(\bar{z}_{t-1,T})\,w_t\xi_{t-1}' + o_p(1)$$

where the notation is as before so that $H_1(\bar{z}_{t-1,T})$ signifies a matrix whose each row is evaluated at a possibly different intermediate point in the line segment between z_{tT} and $z_{t-1,T}$. Since the function $H_1(x)$ is continuously differentiable by assumption, we have $\|H_1(\bar{z}_{t-1,T}) - H_1(z_{tT})\| \leq T^{-1/2}\bar{H}_T\|w_t\|$ where $\bar{H}_T$ is determined by the second partial derivatives of the function $h(x)$ and, as a straightforward consequence of Lemma 1(i), $\bar{H}_T = O_p(1)$. Hence, since $E\|w_t\|\left\|w_t\xi_{t-1}'\right\|$ is a finite constant, we can write

$$\begin{aligned}-T^{-1/2}\sum_{t=1}^{T}h(z_{tT})\,\Delta\xi_t' &= T^{-1}\sum_{t=1}^{T}H_1(z_{tT})\,w_t\xi_{t-1}' + o_p(1) \\ &= T^{-1}\sum_{t=1}^{T}H_1(z_{tT})\,w_t\xi_t' + T^{-1}\sum_{t=1}^{T}H_1(z_{tT})\,w_t w_t' \\ &\quad - T^{-1}\sum_{t=1}^{T}H_1(z_{tT})\,w_t\eta_t' + o_p(1)\end{aligned}$$

where the latter equality follows from the Beveridge-Nelson decomposition (A.2). Theorems 3.2 and 3.3 of Hansen (1992) imply that replacing $w_t\xi_t'$ in the first term of the last expression by its expectation causes an error of order $o_p(1)$. To see that a similar replacement can be done in the second term of the last expression, observe that, by Assumption 2 and the mixing inequality in Davidson (1994, p. 211), $w_t w_t' - Ew_t w_t'$ is a stationary L_1-mixingale. Hence, the desired result follows from Theorem 3.3 of Hansen (1992). As a whole we can thus conclude that

$$-T^{-1/2}\sum_{t=1}^{T} h(z_{tT})\,\Delta\xi_t' = T^{-1}\sum_{t=1}^{T} H_1(z_{tT})\,\Lambda - T^{-1}\sum_{t=1}^{T} H_1(z_{tT})\,w_t\eta_t' + o_p(1) \quad \text{(A.7)}$$

where the result $Ew_t\xi_t' + Ew_t w_t' = \Lambda$ is a simple consequence of the definition of the matrix Λ and the process ξ_t (cf. Hansen, 1992, the proof of Theorem 4.1).

Now consider the first term on the right hand side of (A.6) and use the same mean value expansion as above to write

$$T^{-1/2}\sum_{t=1}^{T} h(z_{tT})\,\eta_t' = T^{-1/2}\sum_{t=1}^{T} h(z_{t-1,T})\,\eta_t' + T^{-1}\sum_{t=1}^{T} H_1(\bar{z}_{t-1,T})\,w_t\eta_t'. \quad \text{(A.8)}$$

In the same way as above, we can also here replace $H_1(\bar{z}_{t-1,T})$ by $H_1(z_{tT})$ and combine equations (A.7) and (A.8) with (A.6). This gives

$$T^{-1/2}\sum_{t=1}^{T} h(z_{tT})\,w_t' = T^{-1/2}\sum_{t=1}^{T} h(z_{t-1,T})\,\eta_t' + T^{-1}\sum_{t=1}^{T} H_1(z_{tT})\,\Lambda + o_p(1).$$

To complete the proof, notice that η_t is a stationary square integrable martingale difference sequence and that an invariance principle holds jointly for the processes z_t and $\sum_{j=1}^{t}\eta_j$ (see Hansen, 1992, the proof of Theorem 3.1). Hence, the stated result is obtained from Theorem 2.1 of Hansen (1992). ¥

8 Appendix II: Proofs of Main Results

Proof of Theorem 1: We shall first demonstrate the existence of the estimators $\tilde{\theta}_T$ and $\tilde{\phi}_T$. For any fixed value of θ, the least squares estimator of ϕ, denoted by $\tilde{\phi}_T(\theta)$, exists and is unique with probability approaching one. This is an immediate consequence of the definition of the estimator $\tilde{\phi}_T(\theta)$ and Lemma 5. Thus, we have

$$Q_T(\theta,\phi) \geq Q_T(\theta,\tilde{\phi}_T(\theta)) \geq \inf_{\theta\in\Theta} Q_T(\theta,\tilde{\phi}_T(\theta)).$$

It is straightforward to check that, when the estimator $\tilde{\phi}_T(\theta)$ exists and is unique, $Q_T(\theta,\tilde{\phi}_T(\theta))$ is a continuous function of θ so that, by the assumed compactness of the parameter space Θ, there exists $\tilde{\theta}_T$ such that $Q_T(\tilde{\theta}_T,\tilde{\phi}_T(\tilde{\theta}_T))$ equals the above infimum. Thus, $\tilde{\theta}_T$ and $\tilde{\phi}_T = \tilde{\phi}_T(\tilde{\theta}_T)$ are the desired least squares estimators.

The next step is to show that $\tilde{\phi}_T$ is bounded in probability. To this end, notice that

$$\begin{aligned}\tilde{\phi}_T &= \left(T^{-1}\sum_{t=1}^{T} f(x_{tT};\tilde{\theta}_T)f(x_{tT};\tilde{\theta}_T)'\right)^{-1} \\ &\quad \times T^{-1}\sum_{t=1}^{T} f(x_{tT};\tilde{\theta}_T)\left[u_t + f\left(x_{tT};\theta_0\right)'\phi_0\right].\end{aligned}$$

Lemma 5 implies that the largest eigenvalue of the inverse on the right hand side is of order $O_p(1)$. Thus, we have to show that the latter factor on the right hand side is of order $O_p(1)$. To see this, note that the assumptions imply that $\sup_{\theta\in\Theta}\|f(x;\theta)\|$ is locally bounded. Therefore, by Lemma 1(i) we have $\max_{1\le t\le T}\left\|f(x_{tT};\tilde{\theta}_T)\right\| \le \max_{1\le t\le T}\sup_{\theta\in\Theta}\|f(x_{tT};\theta)\| = O_p(1)$ and similarly with $\tilde{\theta}_T$ replaced by θ_0. Hence, it follows that $\tilde{\phi}_T = O_p(1)$. Moreover, since $\tilde{\theta}_T = O_p(1)$ holds trivially by the compactness of the parameter space Θ, we have $\tilde{\vartheta}_T = O_p(1)$ which means that the sequence of estimators $\tilde{\vartheta}_T$ is tight.

To prove the consistency of the estimators $\tilde{\theta}_T$ and $\tilde{\phi}_T$, use the definitions to write

$$\begin{aligned}0 &\ge T^{-1}Q_T(\tilde{\theta}_T,\tilde{\phi}_T) - T^{-1}Q_T\left(\theta_0,\phi_0\right) \\ &= T^{-1}\sum_{t=1}^{T}\left[f(x_{tT};\tilde{\theta}_T)'\tilde{\phi}_T - f\left(x_{tT};\theta_0\right)'\phi_0\right]^2 \\ &\quad -2T^{-1}\sum_{t=1}^{T}\left[f(x_{tT};\tilde{\theta}_T)'\tilde{\phi}_T - f\left(x_{tT};\theta_0\right)'\phi_0\right]u_t \\ &= T^{-1}\sum_{t=1}^{T}\left[f(x_{tT};\tilde{\theta}_T)'\tilde{\phi}_T - f\left(x_{tT};\theta_0\right)'\phi_0\right]^2 + o_p(1).\end{aligned}$$

Since $\tilde{\phi}_T = O_p(1)$ the latter equality follows from Lemma 3 with $K = 0$. Now suppose that $\tilde{\vartheta}_T \xrightarrow{p} \vartheta_0$ does not hold. Then, by the tightness of the sequence $\tilde{\vartheta}_T$, we can find a subsequence $\tilde{\vartheta}_{T_j}$ which converges weakly to $\vartheta_* = \left[\theta_*' \ \phi_*'\right]'$, say, and $\vartheta_* \neq \vartheta_0$ with a positive probability (see Billingsley, 1968, Theorem 6.1). Thus, we can conclude that

$$\begin{aligned}0 &\ge T^{-1}Q_{T_j}(\tilde{\theta}_{T_j},\tilde{\phi}_{T_j}) - T^{-1}Q_{T_j}\left(\theta_0,\phi_0\right) \\ &\Rightarrow \int_0^1\left[f\left(B_v^0(s);\theta_*\right)'\phi_* - f\left(B_v^0(s);\theta_0\right)'\phi_0\right]^2 ds\end{aligned}$$

where the weak convergence is justified by Lemma 4 and Lemma A.2 of Saikkonen (2000). (The latter lemma requires that the relevant quantities converge jointly which can be guaranteed by redefining the subsequence if necessary.) When $\vartheta_* \neq \vartheta_0$ it follows from condition (12) that the difference in the weak limit above is nonzero for some value of s and, by continuity, in an open interval. Thus, the last expression is positive with a positive probability. This gives a contradiction so that we must have $\vartheta_* = \vartheta_0$. This completes the proof. ¥

Proof of Theorem 2: For simplicity, denote $h(x_{tT};\vartheta) = f(x_{tT};\theta)'\phi$ so that $Q_T(\vartheta) = \sum_{t=1}^{T}[y_{tT} - h(x_{tT};\vartheta)]^2$. Since θ_0 is assumed to be an interior point of Θ, the consistency of the estimator $\tilde{\vartheta}_T$ justifies the mean value expansion

$$\partial Q_T(\vartheta_0)/\partial\vartheta = -\left(\partial^2 Q_T\left(\bar{\vartheta}_T\right)/\partial\vartheta\partial\vartheta'\right)\left(\tilde{\vartheta}_T - \vartheta_0\right) \tag{A.9}$$

where the notation is as before so that $\partial^2 Q_T\left(\bar{\vartheta}_T\right)/\partial\vartheta\partial\vartheta'$ signifies a matrix whose each row is evaluated at a possibly different intermediate point in the line segment between $\tilde{\vartheta}_T$ and ϑ_0. The partial derivatives can be expressed as

$$\partial Q_T(\vartheta)/\partial\vartheta = -2\sum_{t=1}^{T}\left(\partial h(x_{tT};\vartheta)/\partial\vartheta\right)[y_{tT} - h(x_{tT};\vartheta)]$$

and

$$\begin{aligned}\partial^2 Q_T(\vartheta)/\partial\vartheta\partial\vartheta' &= 2\sum_{t=1}^{T}\left(\partial h(x_{tT};\vartheta)/\partial\vartheta\right)\left(\partial h(\vartheta)/\partial\vartheta'\right)\\ &\quad -2\sum_{t=1}^{T}\left(\partial^2 h(x_{tT};\vartheta)/\partial\vartheta\partial\vartheta'\right)[y_{tT} - h(x_{tT};\vartheta)].\end{aligned}$$

Next, note that

$$\begin{aligned}& T^{-1}\sum_{t=1}^{T}\left(\partial^2 h\left(x_{tT};\bar{\vartheta}_T\right)/\partial\vartheta\partial\vartheta'\right)\left[y_{tT} - h\left(x_{tT};\bar{\vartheta}_T\right)\right]\\ = \; & T^{-1}\sum_{t=1}^{T}\left(\partial^2 h\left(x_{tT};\bar{\vartheta}_T\right)/\partial\vartheta\partial\vartheta'\right)u_t\\ & -T^{-1}\sum_{t=1}^{T}\left(\partial^2 h\left(x_{tT};\bar{\vartheta}_T\right)/\partial\vartheta\partial\vartheta'\right)\left[h\left(x_{tT};\bar{\vartheta}_T\right) - h(x_{tT};\vartheta_0)\right].\end{aligned}$$

Since the function $f(x;\theta)$ is three times continuously differentiable by assumption, it follows from the consistency of the estimator $\tilde{\vartheta}_T$ and Lemma 2(ii) with K fixed that the first term on the right hand side is of order $o_p(1)$. It can be seen that the same is true for the second term by taking a mean value expansion of the difference in the brackets and using the local boundedness of the resulting summands in conjunction with Lemma 1(i) and the consistency of the estimator $\tilde{\vartheta}_T$. Thus, we can write

$$\begin{aligned}T^{-1}\partial^2 Q_T\left(\bar{\vartheta}_T\right)/\partial\vartheta\partial\vartheta' &= 2T^{-1}\sum_{t=1}^{T}\left(\partial h\left(x_{tT};\bar{\vartheta}_T\right)/\partial\vartheta\right)\left(\partial h\left(\bar{\vartheta}_T\right)/\partial\vartheta'\right) + o_p(1)\\ &\Rightarrow 2\int_0^1 K\left(B_v^0(s)\right)K\left(B_v^0(s)\right)' ds.\end{aligned} \tag{A.10}$$

Here the weak convergence can be justified by using the consistency of the estimator $\tilde{\vartheta}_T$, Lemma 4, and Lemma A.2 of Saikkonen (2000). The expression of the limit follows from the definitions.

To complete the proof, use Lemma 6 and the definitions to conclude that

$$\begin{aligned} T^{-1/2}\partial Q_T(\vartheta_0)/\partial\vartheta &= -2T^{-1/2}\sum_{t=1}^{T}\left(\partial h(x_{tT};\vartheta_0)/\partial\vartheta\right)u_t \qquad (A.11)\\ &\Rightarrow -2\left(\int_0^1 K\left(B_v^0(s)\right)dB_u(s)+\int_0^1 K_1\left(B_v^0(s)\right)ds\kappa_{vu}\right) \end{aligned}$$

where the weak convergence holds jointly with that in (A.10). Thus, since the weak limit in (A.10) is positive definite (a.s.) by assumption the result of the theorem is an immediate consequence of (A.9)-(A.11) and the continuous mapping theorem. ¥

Proof of Theorem 3: Denote again $f(x_{tT};\theta)\phi = h(x_{tT};\vartheta)$ and conclude from the definitions that

$$\begin{aligned} \tilde{u}_{tT} &= u_t - \left[h(x_{tT};\tilde{\vartheta}_T) - h(x_{tT};\vartheta_0)\right]\\ &= V_t'\pi_0 + e_{Kt} - H_2\left(x_{tT};\bar{\vartheta}_T\right)(\tilde{\vartheta}_T-\vartheta_0) \end{aligned}$$

where $H_2(x_{tT};\vartheta) = \partial h(x;\vartheta)/\partial\vartheta'$ and $\left\|\bar{\vartheta}_T-\vartheta_0\right\| \le \left\|\tilde{\vartheta}_T-\vartheta_0\right\|$. For simplicity, denote

$$\tilde{M}_T = N^{-1}\sum_{t=K+1}^{T-K}\tilde{p}_{tT}\tilde{p}_{tT}'.$$

Then,

$$\begin{aligned} \begin{bmatrix}\hat{\vartheta}_T^{(1)}-\vartheta_0\\ \hat{\pi}_T^{(1)}-\pi_0\end{bmatrix} &= \begin{bmatrix}\tilde{\vartheta}_T-\vartheta_0\\ -\pi_0\end{bmatrix} + \tilde{M}_T^{-1}N^{-1}\sum_{t=K+1}^{T-K}\tilde{p}_{tT}\,V_t'\pi_0 + \tilde{M}_T^{-1}N^{-1}\sum_{t=k+1}^{T-K}\tilde{p}_{tT}\,e_{Kt}\\ &\quad -\tilde{M}_T^{-1}N^{-1}\sum_{t=K+1}^{T-K}\tilde{p}_{tT}\,H_2\left(x_{tT};\bar{\vartheta}_T\right)(\tilde{\vartheta}_T-\vartheta_0)\\ &= \tilde{M}_T^{-1}N^{-1}\sum_{t=K+1}^{T-K}\tilde{p}_{tT}\,e_{Kt} \qquad (A.12)\\ &\quad -\tilde{M}_T^{-1}N^{-1}\sum_{t=K+1}^{T-K}\tilde{p}_{tT}\left[H_2\left(x_{tT};\bar{\vartheta}_T\right)-H_2(x_{tT};\tilde{\vartheta}_T)\right](\tilde{\vartheta}_T-\vartheta_0). \end{aligned}$$

The latter equality is obtained by replacing $H_2\left(x_{tT};\bar{\vartheta}_T\right)$ in the second expression by $H_2(x_{tT};\tilde{\vartheta}_T) = \tilde{K}(x_{tT})$ and observing that $\tilde{p}_{tT} = \left[\tilde{K}(x_{tT})'\ \ V_t'\right]'$. We shall show next that

$$\left\|N^{-1}\sum_{t=K+1}^{T-K}\tilde{p}_{tT}\left[H_2\left(x_{tT};\bar{\vartheta}_T\right)-H_2(x_{tT};\tilde{\vartheta}_T)\right]\right\| = O_p\left(K/N^{1/2}\right). \qquad (A.13)$$

To this end, notice that, since the function $H_2(x;\theta)$ is continuously differentiable by assumption, a mean value expansion and an application of Lemma 1(i) show that

$$\max_{K+1\le t\le T-K}\left\|H_2\left(x_{tT};\bar{\vartheta}_T\right)-H_2(x_{tT};\tilde{\vartheta}_T)\right\| = O_p(1)\left\|\bar{\vartheta}_T-\tilde{\vartheta}_T\right\| = O_p\left(T^{-1/2}\right)$$

where the latter equality is due to the $T^{1/2}-$consistency of the estimator $\tilde{\vartheta}_T$ obtained from Theorem 2. Thus, since $\tilde{K}(x_{tT}) = H_2(x_{tT};\tilde{\vartheta}_T)$, the local boundedness of $\sup_{\theta\in\Theta}\|H_2(x;\vartheta)\|$ and Lemma 1(i) similarly yield $\max_{K+1\leq t\leq T-K}\left\|\tilde{K}(x_{tT})\right\| = O_p(1)$. Hence, (A.13) holds with $\tilde{p}_{tT}$ replaced by $\tilde{K}(x_{tT})$ and we need to show that it also holds with $\tilde{p}_{tT}$ replaced by V_t. This can be seen by observing that

$$\begin{aligned}
&\left\|N^{-1}\sum_{t=K+1}^{T-K} V_t\left[H_2\left(x_{tT};\bar{\vartheta}_T\right)-H_2(x_{tT};\tilde{\vartheta}_T)\right]\right\|^2 \\
=\ & \sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K} v_{t+j}\left[H_2\left(x_{tT};\bar{\vartheta}_T\right)-H_2(x_{tT};\tilde{\vartheta}_T)\right]\right\|^2 \\
\leq\ & \left(\sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K} v_{t+j}\left[H_2\left(x_{tT};\bar{\vartheta}_T\right)-H_2(x_{tT};\tilde{\vartheta}_T)\right]\right\|\right)^2 \\
=\ & O_p\left(K^2/N\right)
\end{aligned}$$

where the last relation follows from Lemma 2(ii) and the $T^{1/2}-$consistency of the estimator $\tilde{\vartheta}_T$. Thus, we have established (A.13).

The next step is to observe that

$$\left\|\tilde{M}_T^{-1}-\bar{M}_T^{-1}\right\|_1 = O_p\left(K/N^{1/2}\right) \tag{A.14}$$

where, denoting $\lambda_{\max}(A)$ as the largest eigenvalue of matrix A, $\|A\|_1 = (\lambda_{\max}(A'A))^{1/2}$ and

$$\bar{M}_T = diag\left[N^{-1}\sum_{t=K+1}^{T-K}\tilde{K}(x_{tT})\tilde{K}(x_{tT})' \quad N^{-1}\sum_{t=K+1}^{T-K}V_tV_t'\right].$$

To see this, first note that

$$\begin{aligned}
\left\|N^{-1}\sum_{t=K+1}^{T-K}\tilde{K}(x_{tT})V_t'\right\|^2 &= \sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K}H_2(x_{tT};\tilde{\vartheta}_T)v_{t+j}'\right\|^2 \\
&\leq \left(\sum_{j=-K}^{K}\left\|N^{-1}\sum_{t=K+1}^{T-K}H_2(x_{tT};\tilde{\vartheta}_T)v_{t+j}'\right\|\right)^2 \\
&= O_p\left(K^2/N\right)
\end{aligned}$$

again by Lemma 2(ii) and the $T^{1/2}-$consistency of the estimator $\tilde{\vartheta}_T$. This and the well-known fact $\|\cdot\|_1 \leq \|\cdot\|$ imply that $\left\|\tilde{M}_T-\bar{M}_T\right\|_1 = O_p\left(K/N^{1/2}\right)$, and we need to show that a similar result holds for the corresponding inverses. By Lemma A.2 of Saikkonen and Lütkepohl (1996), this holds true if $\left\|\bar{M}_T^{-1}\right\|_1 = O_p(1)$ or if $\left\|\left(N^{-1}\sum_{t=K+1}^{T-K}\tilde{K}(x_{tT})\tilde{K}(x_{tT})'\right)\right\|_1^{-1} = O_p(1)$ and $\left\|\left(N^{-1}\sum_{t=K+1}^{T-K}V_tV_t'\right)^{-1}\right\|_1 =$

$O_p(1)$. The former requirement can be obtained from condition (13), the consistency of the estimator $\tilde{\vartheta}_T$, Lemma 5, and Lemma A2 of Saikkonen (2000) whereas the latter can be deduced from Lemmas A2-A4 of Saikkonen (1991). Since the assumptions used in Saikkonen (1991) were slightly different from the present ones we note that these lemmas, as well as Lemmas A5 and A6 of that paper, can also be proved under the present assumptions. For Lemmas A3 and A5 the previous proofs apply whereas Lemma A2 and, consequently, Lemmas A4 and A6 can be proved by using Lemma A.5(i) of this paper and the fact that, for some finite constant C independent of $j = -K, \ldots, K$,

$$E\left\|N^{-1}\sum_{t=K+1}^{T-K}\left(v_t v_{t+j}' - E v_t v_{t+j}'\right)\right\| \leq N^{-1}C.$$

This follows from Assumption 2 and Lemma 6.19 of White (1984). For later purposes we also note that the above discussion implies that $\left\|\tilde{M}_T^{-1}\right\|_1 = O_p(1)$.

Next, note that $N^{-1/2}\sum_{t=K+1}^{T-K}\tilde{K}(x_{tT})\,e_{Kt} = O_p(1)$ and $\left\|N^{-1/2}\sum_{t=K+1}^{T-K}V_t e_{Kt}\right\| = O_p\left(K^{1/2}\right)$. The former result will become evident below while the latter is obtained from Lemmas A5 and A6 of Saikkonen (1991). Since $\left\|\tilde{M}_T^{-1}\right\|_1 = O_p(1)$ we can use (A.13), (A.14), and the $T^{1/2}-$consistency of the estimator $\tilde{\vartheta}_T$ to conclude from (A.12) that

$$\begin{bmatrix} \hat{\vartheta}_T^{(1)} - \vartheta_0 \\ \hat{\pi}_T^{(1)} - \pi_0 \end{bmatrix} = \bar{M}_T^{-1}N^{-1}\sum_{t=K+1}^{T-K}\tilde{p}_{tT}\,e_{Kt} + O_p\left(K^{3/2}/N\right).$$

Since $K^{3/2}/N^{1/2} \to 0$ by assumption this implies that

$$N^{1/2}(\hat{\vartheta}_T^{(1)} - \vartheta_0) = \left(N^{-1}\sum_{t=K+1}^{T-K}\tilde{K}(x_{tT})\,\tilde{K}(x_{tT})'\right)^{-1} N^{-1/2}\sum_{t=K+1}^{T-K}\tilde{K}(x_{tT})\,e_{Kt} + o_p(1) \tag{A.15}$$

and

$$\begin{aligned}\left\|\hat{\pi}_T^{(1)} - \pi_0\right\| &= \left\|\left(N^{-1}\sum_{t=K+1}^{T-K}V_t V_t'\right)^{-1} N^{-1}\sum_{t=K+1}^{T-K}V_t e_{Kt}\right\| + O_p\left(K^{3/2}/N\right) \\ &= O_p\left((K/N)^{1/2}\right).\end{aligned}$$

Here the last equality follows from results obtained in the Appendix of Saikkonen (1991) and already used above. To show that the limiting distribution of $\hat{\vartheta}_T^{(1)}$ is as stated in the theorem and thereby to complete the proof, first note that the arguments used for (A.10) in the proof of Theorem 2 show that the inverse on the right hand side of (A.15) converges weakly to the inverse in the theorem. Thus, we need to consider

$$N^{-1/2}\sum_{t=K+1}^{T-K}\tilde{K}(x_{tT})\,e_{Kt} = N^{-1/2}\sum_{t=K+1}^{T-K}\tilde{K}(x_{tT})\,e_t + o_p(1)$$

$$= \quad N^{-1/2} \sum_{t=K+1}^{T-K} H\left(x_{tT};\vartheta_0\right) e_t + o_p\left(1\right) \qquad \text{(A.16)}$$

where the equalities can be justified as follows. First, recall that

$$e_{Kt} = e_t + \sum_{|j|>K} \pi_j' v_{t-j} \overset{def}{=} e_t + a_{Kt}$$

and note that $E\left\|a_{Kt}\right\|^2 = o_p\left(T^{-1}\right)$ for all t, as shown in the proof of Lemma A5 of Saikkonen (1991). Thus, the first equality in (A.16) follows because $\max_{K+1\leq t\leq T-K}\left\|\tilde{K}\left(x_{tT}\right)\right\| = O_p\left(1\right)$, as already noticed. To justify the second equality, recall that $\tilde{K}\left(x_{tT}\right) = H_2(x_{tT};\tilde{\vartheta}_T)$, take a mean value expansion of $H_2(x_{tT};\tilde{\vartheta}_T)$ about ϑ_0, and use the $T^{1/2}-$consistency of the estimator $\tilde{\vartheta}_T$ in conjunction with Lemma 3 with $K = 0$.

To complete the proof we have to show that the first term in the last expression of (A.16) converges weakly to the stochastic integral in the theorem and that this holds jointly with the weak convergence of the inverse on the right hand side of (A.15). If the process $[v_t' \;\; e_t']'$ fulfilled the conditions of Assumption 2 this would follow from Lemma A.9 but, since the process e_t is not guaranteed to be strong mixing, this reasoning does not apply directly. However, using L to denote the usual lag operator we may write $e_t = a\left(L\right)' w_t$ where $a\left(L\right)' = \sum_{j=-\infty}^{\infty} a_j' L^j = \left[1 \quad -\pi\left(L\right)'\right]$ and $\pi\left(L\right) = \sum_{j=-\infty}^{\infty} \pi_j L^j$. In view of the summability condition (15) and Lemma A.9 we can use Theorem 4.2 of Saikkonen (1993) and obtain the needed weak convergence results. The assumptions required to apply this theorem are straightforward consequences of Assumption 2 which, in addition to the summability condition (6) and the invariance principle (8), also implies that the first and second sample moments of w_t are consistent estimators of their theoretical counterparts. This completes the proof. ¥

9 References

Andrews, D.W.K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

Andrews, D.W.K. and C.J. McDermott (1995): "Nonlinear Econometric Models with Deterministically Trending Variables," *Review of Economic Studies* 62, 343-360.

Billingsley, P. (1968): *Convergence of Probability Measures.* John Wiley & Sons: New York

Brillinger, D.R. (1975): *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston: New York.

Choi, I. (1994): "Residual Based Tests for the Null of Stationarity with Applications to U.S. Macroeconomic Time Series," *Econometric Theory* 19, 720-746.

Choi, I. and D.K. Park (2000): "Causal Relation Between Interest and Exchange Rates in the Asian Currency Crisis," mimeo, University of Southern California.

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

Dekle, R., C. Hsiao and S. Wang (1999): "Interest Rate Stabilization of Exchange Rates and Contagion in the Asian Crisis Countries," mimeo, University of Southern California.

Elliott, G., Rothenberg, T.J. and Stock, J. H. (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica* 64, 813-836.

Feldstein, M. (1998): "Refocusing the IMF," *Foreign Affairs* 77, 20-33.

Furman, J. and J.E. Stiglitz (1998): "Economic Crises: Evidence and Insights from East Asia," *Brookings Papers on Economic Activity* 2, 1-135.

Ghosh, A. and S. Phillips (1998): "Interest Rates, Stock Market Prices, and Exchange Rates in East Asia," mimeo, International Monetary Fund.

Goldfajn, I. and T. Baig (1998): "Monetary Policy in the Aftermath of Currency Crises: The Case of Asia," IMF Working Paper, No. 98-170.

Goldfeld, S.M. and R. Quandt (1973): "The Estimation of Structural Shifts by Switching Regressions," *Annals of Economic and Social Measurement* 2, 475-485.

Granger, C.W.J. and T. Teräsvirta (1993): *Modelling Non-linear Economic Relationships.* Oxford University Press: Oxford.

Granger, C.W.J., T. Teräsvirta and H. Anderson (1993): "Modelling Non-Linearity Over the Business Cycle," in J.H. Stock and M.W. Watson (eds.) *New Reserch on Business Cycles, Indicators and Forecasting*, 311-325. Chicago University Press: Chicago.

Hansen, B.E. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory* 8, 489-500.

Kaminsky, G. and S. Schmukler (1998): "The Relationship Between Interest Rates and Exchange Rates in Six Asian Countries," mimeo, World Bank.

Kraay, A. (1998): "Do High Interest Rates Defend Currencies During Speculative Attacks?," mimeo, World Bank.

Lütkepohl, H. T. Teräsvirta and J. Wolters (1999): "Investigating Stability and Linearity of a German M1 Demand Function," *Journal of Applied Econometrics* 14, 511-525.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988): "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika* 75, 491-499.

Newey, W. and D.L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing" in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* Vol. IV, pp. 2113-2445. Elsevier Science: Amsterdam.

Park, J. and P.C.B. Phillips (2000): "Nonlinear Regressions With Integrated Time Series," *Econometrica* (forthcoming).

Park, J. and P.C.B. Phillips (1999): "Asymptotics for Nonlinear Transformations of Time Series," *Econometric Theory* 15, 269-298.

Park, Y.C., Y. Wang, C.-S. Chung (1999): "Exchange Rate Policies in Korea: Has Exchange Rate Volatility Increased After the Crisis?", mimeo, Korea Institute for Economic Policy.

Phillips, P. C. B. and M. Loretan (1991): "Estimating Long Run Economic Equilibria," *Review of Economic Studies*, 58, 407–436.

Phillips, P.C.B. and V. Solo (1992): "Asymptotics for Linear Processes," *Annals of Statistics* 20, 971-1001.

Pötscher, B. and I.R. Prucha (1997): *Dynamic Nonlinear Econometric Models.* Springer: New York.

Radelet, S. and J. Sachs (1998a): "The East Asian Financial Crisis: Diagnosis, Remedies, Prospects", *Brookings Papers on Economic Activity* , 1-74.

Radelet, S. and J. Sachs (1998b): "The Onset of the East Asian Financial Crises," mimeo, Harvard Institute for International Development.

Saikkonen, P. (1991): "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory* 7, 1-21.

Saikkonen, P. (1993): "Continuous weak convergence and stochastic equicontinuity results for integrated processes with an application to the estimation of a regression model," *Econometric Theory* 9, 155-188.

Saikkonen, P. (2000): "Consistent Estimation in Cointegrated Vector Autoregressive Processes With Nonlinear Time Trends in Cointegrating Relations," *Econometric Theory* (forthcoming).

Saikkonen, P. and H. Lütkepohl (1996): "Infinite Order Cointegrated Vector Autoregressive Processes: Estimation and Inference," *Econometric Theory* 12, 814-844.

Sarno, L. (1999): "Adjustment Costs and Nonlinear Dynamics in the Demand for Money: Italy, 1861-1991," *International Journal of Finance and Economics* 4, 155-177.

Stock, J. H. and M. W. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

Teräsvirta, T. and H.M. Anderson (1992): "Characterizing Nonlinearities in business Cycles Using Smooth Transition Autoregressive Models," *Journal of Applied Econometrics* 7, S119-S136.

Tong, H. (1983): *Threshold Models in Non-Linear Time Series Analysis.* Springer-Verlag: New York.

White, H. (1984): *Asymptotic Theory for Econometricians.* Academic Press, Inc.: San Diego.