NORMING RATES AND LIMIT THEORY FOR SOME TIME-VARYING COEFFICIENT AUTOREGRESSIONS

By

Offer Lieberman and Peter C.B. Phillips

September 2013

COWLES FOUNDATION DISCUSSION PAPER NO. 1916

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Norming Rates and Limit Theory for Some Time-Varying Coefficient Autoregressions*

Offer Lieberman† and Peter C. B. Phillips‡

Revised, August 24, 2013

Abstract

A time-varying autoregression is considered with a similarity-based coefficient and possible drift. It is shown that the random walk model has a natural interpretation as the leading term in a small-sigma expansion of a similarity model with an exponential similarity function as its autoregressive coefficient. Consistency of the quasi-maximum likelihood estimator of the parameters in this model is established, the behaviors of the score and Hessian functions are analyzed and test statistics are suggested. A complete list is provided of the normalization rates required for the consistency proof and for the score and Hessian functions standardization. A large family of unit root models with stationary and explosive alternatives are characterized within the similarity class through the asymptotic negligibility of a certain quadratic form that appears in the score function. A variant of the stochastic unit root model within the class is studied and a large sample limit theory provided which leads to a new nonlinear diffusion process limit showing the form of the drift and conditional volatility induced

*Our thanks to two referees and the Editor for helpful comments on the earlier version of this paper.

†Bar-Ilan University. Support from Israel Science Foundation grant No. 396/10 is gratefully acknowledged. Correspondence to: Department of Economics and Research Institute for Econometrics (RIE), Bar-Ilan University, Ramat Gan 52900, Israel. E-mail: offer.lieberman@gmail.com

‡Yale University, University of Auckland, Southampton University, and Singapore Management University. Support is acknowledged from the NSF under Grant Nos. SES 0956687 and SES 1258258.

by this model . Some simulations and a brief empirical application to data on an Australian Exchange Traded Fund are included.

Key words and phrases: Autoregression; Consistency; Nonlinear diffusion; Nonstationarity; Similarity; Small sigma approximation; Stochastic unit root; Time-varying coefficients.

JEL Classification: C22

1 Introduction

First-order autoregressions with possible unit roots or roots that are in the vicinity of unity have attracted an enormous amount of interest over recent decades. The literature now provides a near comprehensive coverage of estimation and testing of the coefficient of the lag dependent variable in stationary, unit root, explosive, and many intermediate cases of near and mild integration, including models with or without fitted intercepts and trends. Traditional analysis of this model relates to data generating processes (DGP's) with fixed coefficients that are consistent with a single scenario. For instance, empirical studies frequently work under a null hypothesis that the DGP is a unit root process with drift, not that the process may have fluctuating, time-dependent parameters that are compatible with stationary behavior for some parts of the sample, unit root behavior for other parts, and mildly explosive behavior elsewhere. Most econometric software packages include common tests for a unit root which reflect this characterization. But recent empirical work, particularly on the global financial crisis, has shown the advantages of working with flexible systems that accommodate multiple regimes of stationary and nonstationary behavior and transition mechanisms between them (e.g. Phillips, Wu and Yu, 2011; Phillips and Yu, 2011).

A second trend in the literature involves models with time varying coefficients, such that the process is at least weakly stationary. See among others, Nicholls and Quinn (1980, 1981, 1982), Chen and Tsay (1993), Dahlhaus *et. al.* (1999), Dahlhaus (2000) and Lundbergh *et. al.* (2003). Some related work on explosive random coefficient autoregressive processes has been done by Hwang and Basawa (2005). In addition, Granger and Swanson (1997) introduced a stochastic unit root (STUR) model where the autoregressive root is in the vicinity of unity, is stochastic, and is driven by an independent stationary process - see also McCabe and Tremayne (1995). Some properties of that model were derived but no limit theory or estimation theory was established. Within the context of a wider class of models, the present paper considers a variant of the STUR model, provides a large sample limit theory and studies the discriminatory power of unit root testing against STUR alternatives. We do not cover in this paper an important line of the literature dealing with time varying coefficient models that are not autoregressions. Those models give rise to issues which are very different from the ones which surface in autoregressions.

Recently, Lieberman (2012) introduced a similarity-based model in the context of time varying coefficient autoregressions. That paper developed the asymptotic theory for quasi-maximum-likelihood estimation (QMLE) of this model and various statistical tests. Unlike earlier literature, the coefficient of the lag dependent variable in this model can fluctuate freely and, at any specific period t, the process may behave in a stationary, unit root, or explosive manner. This feature of the similarity model adds some flexibility to the prominent unit root model producing a system for which unit root effects may hold on average in a given sample but not necessarily at all points within the sample.

In this paper we develop the idea further by considering a larger class of models and by showing that the unit root model can be naturally interpreted as a small-σ asymptotic approximation to the similarity model. To fix ideas, we consider the process

$$\begin{aligned} Y_1 &= \mu + \varepsilon_1, \\ Y_t &= \mu + \beta_t(w) Y_{t-1} + \varepsilon_t, t = 2, ..., n, \end{aligned} \tag{1}$$

where $\mu \in \mathbb{R}$, w is an m-dimensional vector of unknown parameters, assumed to lie in a compact subset of R^m, $\beta_t(w) = \beta_t(x_t, x_{t-1}; w) \in \mathbb{R}_+$, $x_t = (X_{1t}, ..., X_{mt})'$ is an m-vector of explanatory variables, and $\{\varepsilon_t\}$ is a sequence of $iid(0, \sigma_\varepsilon^2)$ random variables with cumulants κ_r, $r \geq 3$. When there is no risk of ambiguity we shall simply write σ^2 in place of σ_ε^2. It is emphasized that μ can be zero or otherwise and that the set of permissible specifications for $\beta_t(x_t, x_{t-1}; w)$ is rich. Moreover, for a given t, $\beta_t(w)$ can be less than-, equal to- or greater than unity, so that the model can behave in a stationary-, unit-root- or explosive manner over subperiods.

Of particular interest is the model with an exponential similarity function

$$\beta_t(x_t, x_{t-1}; w) = \exp(wu_t), \ m = 1, \tag{2}$$

where $u_t = \Delta x_t = \Delta X_{1t}$ is the source of variation in the coefficient. A natural choice for x_t would be an explanatory variable for Y_t, but other generators of coefficient randomness are possible. Suppose that u_t are i.i.d. copies of u. $M_u(\omega) = \mathbb{E}(\beta_t)$ is the moment generating function (mgf) of u when it exists, and under certain conditions that will be discussed in Section 2,

$$M_u(w) = 1 + O\left((w\sigma_u)^2\right), \tag{3}$$

where $\sigma_u^2 = \text{Var}(u_t)$, $\forall t$. Moreover, the sample path of the coefficient $\beta_t(w)$ has an average that converges pointwise in probability to $M_u(w)$. Thus, on average, the unit root specification, which is believed to be prevalent in economic and financial data, may be interpreted as the leading form in a small-σ expansion of the similarity model, with an error which is of the order $O(\sigma_u^2)$.

In studying the model (1), we prove consistency of the QMLE of w when $\beta_t(w)$ is allowed to be non-negative and with $\mu \in \mathbb{R}$. This extends the results of Lieberman (2012) by allowing for the case $\mu = 0$. To achieve the results, we introduce uniform norming factors which are functions of $n \times n$ matrices, one of which covers the $\mu = 0$ case and another the case $\mu \neq 0$. The behaviors of the score and Hessian functions are analyzed and, as with the consistency proof, separate uniform norming factors are given for the $\mu = 0$ case and for the $\mu \neq 0$ case. The simplest scenario, in which $\mu \neq 0$ and the process is approximately a unit root, leads to a score-based test which is asymptotically normal (see Lieberman 2012). A further case, where the slope coefficient $w = w_n$ in the similarity function (2) is local to zero and is allowed to be random, also approximates a unit root model. This STUR model is analyzed by weak convergence methods and its limit theory is related to that of a local to unity process but gives rise to a new nonlinear diffusion. The properties of this limit process reveal the explicit form of the conditional volatility induced by the similarity function. In many other cases the distribution theory is much more complicated.

The plan for the paper is as follows. In Section 2 we discuss connections and interpretations of the model and show how the random walk model can be interpreted as a small-σ approximation of the similarity model with an exponential similarity function. Notation and the main results are introduced in Section 3, followed by some discussion in Section 4. Section 5 studies a similarity-based STUR model, its limit theory, and the discriminatory power of unit root tests against STUR alternatives. Simulations and an empirical application are provided in Section 6 and Section 7 concludes. All proofs are contained in the Appendix.

2 Connections and Interpretations of the model

This section draws connections between the similarity model and existing time series models and provides some insights and interpretations of our approach.

2.1 A Unit Root Model as a Small-σ Approximation to the Similarity Model

Small σ asymptotics were originally developed in Kadane (1971) to approximate finite sample distributions in simultaneous equations models to compare k-class estimators in terms of their bias, variance and mean squared error. They may also be used to take expansions about the standard regression model, which applies in a limiting case where the variance of the endogenous regressors tends to zero. A related method was used by Samuelson (1970) to develop quadratic and higher order approximations useful in portfolio analysis under situations where there is less and less risk.

In the present case, we consider the unit root model as a small σ approximation of a more general system involving time varying random coefficients. For the exponential similarity function given in (2), assume that u_t are i.i.d., copies of u, each with mgf $M_u(w)$, zero mean and small variation, as when it is distributed as $U[-a, a]$, or $N(0, \sigma_u^2)$, with small a or σ_u^2, respectively. The import is that the coefficient of Y_{t-1} varies with u_t but that the fluctuations in the coefficient value are not too large. In the former case, as $\sigma_u^2 = a^2/3$,

$$\begin{aligned} M_u(w) &= \frac{e^{wa} - e^{-wa}}{2wa} \\ &= 1 + \frac{(wa)^2}{6} + \frac{(wa)^4}{120} + O\left((wa)^6\right) \\ &= 1 + \frac{(w\sigma_u)^2}{2} + \frac{3(w\sigma_u)^4}{40} + O\left((wa)^6\right) \end{aligned} \tag{4}$$

and the property of the limit shown in (3) follows. In the second case,

$$\begin{aligned} M_u(w) &= e^{(w\sigma_u)^2/2} \\ &= 1 + \frac{(w\sigma_u)^2}{2} + \frac{(w\sigma_u)^4}{8} + O\left((w\sigma_u)^6\right) \end{aligned}$$

so that (3) again holds. Of course, in both cases we can reparameterize with $wu_t = u_t^* \sim U[-w^*, w^*]$, $w^* = wa$ or $u_t^* \sim N(0, w^*)$, $w^* = w\sigma_u$. These choices of β_t are natural and reflect the principle that the average Y_{t-1}-coefficient value may be close to unity across the sample but will deviate from unity at any point on the trajectory. In fact, the sample path of the coefficient $\beta_t(w)$ in this setting has an average that converges pointwise in probability to $M_u(w)$. Figure 1 illustrates this point showing the QMLE of β_t for data on an Australian Exchange Traded Fund (ETF). We emphasize that a limiting case occurs when $x_t = c$, *a.s.* for all t, so that $u_t = 0$, *a.s.*, so that there is no random variation in the AR coefficient and the model reduces to a random walk. The unit root model can thus be viewed as a small-σ approximation to a flexible similarity model.

2.2 CAPM

The capital asset pricing model (CAPM) has long been central to finance and provides a working foundation for more sophisticated models. Let $\Delta \log Y_t$ be the expected excess return of a certain capital asset and $\Delta \log Z_t$ be the market premium, both at time t. Without the error term, the CAPM model relates these quantities through the equation

$$\Delta \log Y_t = \beta \Delta \log Z_t. \tag{5}$$

Upon rearrangement, the model is

$$Y_t = \exp(\beta \Delta x_t) Y_{t-1},$$

where, in line with our notation so far, $x_t = X_{1t} = \log Z_t$. Thus, the CAPM model is just a similarity model with an exponential similarity function in which the value of Y_t is determined by its similarity to Y_{t-1} through the closeness of X_t to X_{t-1}[1].

2.3 Threshold Autoregression

The threshold case

$$\beta_t(x_t, x_{t-1}; w) = 1\{\|\Delta x_t\| < w_1\} + w_2 1\{\|\Delta x_t\| \geq w_1\},$$

where $\|\cdot\|$ is the Euclidean norm, is of particular interest. Here, Y_{t-1} receives a unit weight in the response only if its characteristics, x_{t-1}, are within w_1-Euclidean distance from x_t, the characteristics of Y_t. Otherwise, Y_{t-1} is considered to be too 'far' from Y_t and receives only a w_2-weight, where $|w_2| < 1$. This model essentially has the form of a threshold autoregression, see for instance, Tong (2011) and the references therein.

3 Notation and Main Results

This section contains the main theoretical results of the paper. Assumptions and proofs for what follows are given in the Appendix.

Let $C = C(w)$ be an $n \times n$ matrix with entries $[C(w)]_{t,t-1} = \beta_t(x_t, x_{t-1}; w)$, $t = 2, ..., n$ and $[C(w)]_{i,j} = 0$, otherwise, $x_t = (X_{1t}, ..., X_{mt})'$ is an m-vector of explanatory variables, I_n be the identity matrix of order n, $S = I_n - C$, $\theta_1 = \sigma^2$, $\theta_2' = (w_1, \ldots, w_m)$, $\theta = (\theta_1, \theta_2')'$, $y = (Y_1, ..., Y_n)'$ and $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$. For brevity, we shall write $\beta_t(w)$ in place of $\beta_t(x_t, x_{t-1}; w)$, or simply β_t. Following convention, the true values of θ, w, and μ are denoted θ_0, w_0 and μ_0, respectively. Similarly, we set $C_0 = C(w_0)$ and $S_0 = I_n - C_0$.

For the $\mu_0 = 0$ case, model (1) can be rewritten as $y = S^{-1}\varepsilon$ and as $\det(S) = 1$, the quasi-log-likelihood function is given by

$$l_n(\theta) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{y'S'Sy}{2\sigma^2}. \tag{6}$$

For the $\mu_0 \neq 0$ case, we concentrate the quasi-log-likelihood function by using $\hat{\mu}_n(\theta) = 1'S(\theta)y/n$ in place of μ, which leads to

$$(Sy - \hat{\mu}_n(\theta)1)'(Sy - \hat{\mu}_n(\theta)1) = y'S'MSy$$

[1]Strictly speaking, the inclusion of an additive error term in (5) results in the model $Y_t = \exp(\beta\Delta X_t)Y_{t-1}\exp(\varepsilon_t)$ - a similarity model with a multiplicative error term.

and therefore,

$$l_n^c(\theta) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{y'S'MSy}{2\sigma^2}, \tag{7}$$

where $M = I_n - P$, $P = 11'/n$ and $1'$ is an $n \times 1$ (row) vector of 1's. For brevity, $\hat{\theta}_n$ will denote the QMLE of θ using either (6), for the case $\mu_0 = 0$, or (7), for the case $\mu_0 \neq 0$.

We use the l_1-, spectral- and Frobenius norms of an $n \times n$ matrix A, denoted by $\|A\|_1$, $\|A\|_2$ and $\|A\|_F$, and given by $\|A\|_1 = \sum_{i,j=1}^n \left|[A]_{i,j}\right|$, $\|A\|_2 = \left(\sup_{|x|=1} x'A'Ax\right)^{1/2}$ and $\|A\|_F = (tr\,(A'A))^{1/2}$. Finally, by $O_e(\cdot)$ and $O_{p,e}(\cdot)$ we denote the exact order and exact order in probability, respectively.

The quantity

$$\rho_n = \frac{\left\|S_0^{-1}\right\|_F^2}{\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}$$

turns out to be central to the asymptotic analysis. Terms in the expansions of $l_n(\theta)$, $l_n^c(\theta)$, the score and Hessian functions based on them, can be conveniently grouped according to orders of magnitude of powers of ρ_n. As C_0 is nonnegative and nilpotent by Assumption A2,

$$S_0^{-1} = I_n + C_0 + \cdots + C_0^{n-1},$$

so that all the elements of S_0^{-1} are nonnegative. It follows that

$$\left\|S_0^{-1}\right\|_F^2 = \sum_{i=1}^n \left[S_0^{-1\prime}S_0^{-1}\right]_{i,i} \leq \sum_{i,j=1}^n \left[S_0^{-1\prime}S_0^{-1}\right]_{i,j} = \left\|S^{-1\prime}S^{-1}\right\|_1, \tag{8}$$

implying that

$$\rho_n \leq 1. \tag{9}$$

The behavior of ρ_n has been analyzed by Lieberman (2012, lemma 1) in some leading special cases. In particular, for fixed coefficient stable or explosive autoregressions ρ_n is bounded from below, whereas for the unit root model, $\rho_n = o(1)$.

Consistency of the QMLE is given in Theorem 1.

Theorem 1 *Under Assumptions A0-A3,* $\hat{\theta}_n \to_p \theta_0$.

The requirements for consistency seem quite weak. To establish the asymptotic distribution of the score and the Hessian behavior, we let D_n and D_n^c be $n \times n$ normalizing matrices corresponding to the $\mu_0 = 0$ and $\mu_0 \neq 0$ cases, respectively, such that,

$$D_n = \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\left\| S_0^{-1} \right\|_F} & & \\ \cdots & & \cdots & \\ & & & \\ 0 & & & \frac{1}{\left\| S_0^{-1} \right\|_F} \end{pmatrix}$$

and

$$D_n^c = \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{1/2}} & & \\ \cdots & & \cdots & \\ & & & \\ 0 & & & \frac{1}{\left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{1/2}} \end{pmatrix}.$$

The normalized concentrated score is

$$\begin{aligned} z_n(\theta) &= D_n z_n^*(\theta), \; z_n^*(\theta) = \frac{\partial l_n(\theta)}{\partial \theta}, \text{ if } \mu_0 = 0, \\ z_n^c(\theta) &= D_n^c \frac{\partial l_n^c(\theta)}{\partial \theta}, \; z_n^{c*}(\theta) = \frac{\partial l_n^c(\theta)}{\partial \theta}, \text{ if } \mu_0 \neq 0, \end{aligned}$$

with components $z_{nr}(\theta)$ and $z_{nr}^c(\theta)$, $r = 1, ..., m+1$. We have,

$$z_{n1}(\theta_0) = -\frac{\sqrt{n}}{2\sigma_0^2} + \frac{y' S_0' S_0 y}{2\sigma_0^4 \sqrt{n}}, \mu_0 = 0,$$

and

$$z_{n1}^c(\theta_0) = -\frac{\sqrt{n}}{2\sigma_0^2} + \frac{y' S_0' M S_0 y}{2\sigma_0^4 \sqrt{n}}, \mu_0 \neq 0. \tag{10}$$

For $r = 2, ..., m+1$, let $\dot{S}_{0r} = \partial S_0 / \partial \theta_r$,

$$\Lambda_{0r} = \dot{S}_{0r} S_0^{-1} + S_0^{-1\prime} \dot{S}_{0r}', \tag{11}$$

$$\Gamma_{0r} = M \dot{S}_{0r} S_0^{-1} + S_0^{-1\prime} \dot{S}_{0r}' M, \tag{12}$$

$$QF_{nr} = \varepsilon' \Lambda_{0r} \varepsilon, \tag{13}$$

$$QF_{nr}^{c} = \varepsilon' \Gamma_{0r} \varepsilon, \tag{14}$$

and

$$LF_{nr}^{c} = 1' S_0^{-1\prime} \dot{S}_{0r}' M \varepsilon. \tag{15}$$

The notations QF_{nr} and LF_{nr} stands for quadratic form and linear form, respectively and when the superscript c is used, it indicates that the score is based on $l_n^c(\theta)$. For $r = 2, ..., m+1$ then, we have

$$z_{nr}(\theta_0) = -\frac{QF_{nr}}{2\sigma_0^2 \left\| S_0^{-1} \right\|_F}, \mu_0 = 0, \tag{16}$$

$$z_{nr}^{c}(\theta_0) = -\frac{QF_{nr}^{c} + 2\mu_0 LF_{nr}^{c}}{2\sigma_0^2 \left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{1/2}}, \mu_0 \neq 0, \text{ if } \rho_n = O_e(1) \tag{17}$$

and

$$z_{nr}^{c}(\theta_0) = -\frac{\mu_0 LF_{nr}^{c}}{\sigma_0^2 \left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{1/2}} + o_p(1), \mu_0 \neq 0, \text{ if } \rho_n = o(1). \tag{18}$$

As is evident from (16)-(18), the leading terms of $z_{nr}(\theta)$ and $z_{nr}^{c}(\theta)$ depend on both the value of μ_0 and on the order of magnitude of ρ_n. In particular, in the $\mu_0 \neq 0$ case with $\rho_n = o(1)$, the leading term in (18) is linear in ε. This case corresponds, for instance, to a unit root with a drift process. On the other hand, in the $\mu_0 \neq 0$ with $\rho_n = O_e(1)$ case, $z_{nr}^{c}(\theta)$ $(r = 2, ..., m+1)$ involves both a linear and a quadratic form in ε, so the asymptotic distributions of the score in the two cases are very different.

Theorem 2 *Under Assumptions A0-A3:*

1. $z_{n1}(\theta_0)$ and $z_{n1}^{c}(\theta_0)$ converge in distribution to $N\left(0, \frac{1}{2\sigma_0^4}\left(1 + \frac{\kappa_4}{2\sigma_0^4}\right)\right)$.

2. $z_{nr}(\theta_0)$ and $z_{nr}^{c}(\theta_0)$ are non-negligible, $r = 2, ..., m+1$.

3. In the case $\mu_0 \neq 0$, $\rho_n = o(1)$ and a Gaussian ε, $z_{nr}^{c}(\theta_0)$ is asymptotically normal. In all other cases $z_{nr}(\theta)$ and $z_{nr}^{c}(\theta)$ involve quadratic forms in ε.

Since, in general, $z_{nr}(\theta)$ and $z_{nr}^c(\theta)$ involve quadratic forms, their asymptotic distributions cannot be determined without additional structure on the model. See Theorem 2 of Lieberman (2012).

Continuing, the normalized Hessian is given by

$$H_n(\theta) = D_n H_n^*(\theta) D_n,\ H_n^*(\theta) = \frac{\partial^2 l_n(\theta)}{\partial\theta\partial\theta'},\text{ if } \mu_0 = 0,$$

and

$$H_n^c(\theta) = D_n^c H_n^{c*}(\theta) D_n^c,\ H_n^{c*}(\theta) = \frac{\partial^2 l_n^c(\theta)}{\partial\theta\partial\theta'},\text{ if } \mu_0 \neq 0,$$

with components $H_{nr,s}(\theta)$ and $H_{nr,s}^c(\theta)$, $r,s = 1,...,m+1$. Let

$$H_{n1,1}(\theta_0) = \frac{1}{2\sigma_0^4} - \frac{y'S_0'S_0y}{n\sigma_0^6},$$

$$H_{n1,r}(\theta_0) = \frac{\varepsilon'\Lambda_{0r}\varepsilon}{2\sigma_0^4\sqrt{n}\left\|S_0^{-1}\right\|_F},\ r = 2,...,m+1,$$

$$H_{nr,s}^L(\theta_0) = -\frac{\varepsilon'S_0^{-1}\dot{S}_{0s}'\dot{S}_{0r}S_0^{-1}\varepsilon}{\sigma_0^2\left\|S_0^{-1}\right\|_F^2}, r,s = 2,...,m+1,$$

$$H_{n1,1}^c(\theta_0) = \frac{1}{2\sigma_0^4} - \frac{y'S_0'MS_0y}{n\sigma_0^6},$$

$$H_{n1,r}^c = \frac{\varepsilon'\Gamma_{0r}\varepsilon}{2\sigma_0^4\sqrt{n}\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}},\ r = 2,...,m+1,$$

$$H_{nr,s}^{c,L_1}(\theta_0) = -\frac{\mu_0^2 1'S_0^{-1}\dot{S}_{0s}'M\dot{S}_{0r}S_0^{-1}1}{\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1},\ r,s = 2,...,m+1,$$

$$H_{nr,s}^{c,L_2}(\theta_0) = -\frac{2\mu_0 1'S_0^{-1\prime}\dot{S}_{0s}'M\dot{S}_{0r}S_0^{-1}\varepsilon}{\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1},\ r,s = 2,...,m+1,$$

and

$$H_{nr,s}^{c,L_3}(\theta_0) = -\frac{\varepsilon'S_0^{-1}\dot{S}_{0s}'M\dot{S}_{0r}S_0^{-1}\varepsilon}{\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1},\ r,s = 2,...,m+1.$$

Theorem 3 *Under Assumptions A0-A4,*

1. $H_{n1,1}(\theta_0)$ *and* $H_{n1,1}^c(\theta_0)$ *converge in probability to* $-\left(2\sigma_0^2\right)^{-1}$, $\forall\mu \in \mathbb{R}$.

For $r = 2, ..., m+1$:

2. $H_{n1,r}(\theta_0)$ *and* $H^c_{n1,r}(\theta_0)$ *converge in probability to* 0, $\forall \mu \in \mathbb{R}$.

3. $H_{nr,s}(\theta_0) = H^L_{nr,s}(\theta_0) + o_p(1)$, *with* $H^L_{nr,s}(\theta_0) = O_{p,e}(1)$, *if* $\mu_0 = 0$.

4. $H^c_{nr,s}(\theta_0) = H^{c,L_1}_{nr,s}(\theta_0) + H^{c,L_2}_{nr,s}(\theta_0) + H^{c,L_3}_{nr,s}(\theta_0) + o_p(1)$, *with* $H^{c,L_1}_{nr,s}(\theta_0) = O_{p,e}(1)$, $H^{c,L_1}_{nr,s}(\theta_0) = O_p\left(\sqrt{\rho_n}\right)$ *and* $H^{c,L_3}_{nr,s}(\theta_0) = O_{p,e}(\rho_n)$, *if* $\mu_0 \neq 0$ *and* $\rho_n = O_e(1)$.

5. $H^c_{nr,s}(\theta_0) = H^{c,L_1}_{nr,s}(\theta_0) + o_p(1)$, *with* $H^{c,L_1}_{nr,s}(\theta_0) = O_{p,e}(1)$, *if* $\mu_0 \neq 0$ *and* $\rho_n = o(1)$.

We remark that unlike the stationary fixed coefficient case, $H_n(\theta_0)$ may not converge to a fixed matrix. For example, in the $\mu_0 = 0$ and $\beta_t(w) = 1$, $\forall t$ case, $-H_n(\theta_0)$ converges to the random variable $2\int_0^1 W(r)^2 dr$, see Phillips (1987, Theorem 3.1). Nevertheless, we may still use Theorems 1-3 to construct hypothesis tests of the form $H_0 : \theta = \theta_0$ by adopting random norming, see, among others, Heyde (1975), Feigin (1976) and Lieberman (2010). To do so, we recall that by the mean value theorem

$$z_n(\theta_0) = -H_n\left(\bar{\theta}_n\right) D_n^{-1}\left(\hat{\theta}_n - \theta_0\right), \text{ if } \mu_0 = 0, \tag{19}$$

where $\bar{\theta}_n$ satisfies $\left\|\bar{\theta}_n - \hat{\theta}_n\right\| \leq \left\|\hat{\theta}_n - \theta_0\right\|$. Let

$$A^*_n(\theta_0) = E_{\theta_0}\left(\frac{\partial l_n(\theta_0)}{\partial \theta}\frac{\partial l_n(\theta_0)}{\partial \theta'}\right),$$

and

$$A_n(\theta_0) = D_n A^*_n(\theta_0) D_n.$$

Multiplying both sides of (19) by $A_n^{-1/2}(\theta_0)$ we obtain

$$A_n^{-1/2}(\theta_0) z_n(\theta_0) = -A_n^{-1/2}(\theta_0) H_n\left(\bar{\theta}_n\right) D_n^{-1}\left(\hat{\theta}_n - \theta_0\right), \text{ if } \mu_0 = 0, \tag{20}$$

or simply,

$$A_n^{*-1/2}(\theta_0) z^*_n(\theta_0) = -A_n^{*-1/2}(\theta_0) H^*_n\left(\bar{\theta}_n\right)\left(\hat{\theta}_n - \theta_0\right), \text{ if } \mu_0 = 0, \tag{21}$$

the difference between (20) and (21) being the cancellation of the normalization matrix, D_n. The last equation forms the basis of our test statistic for the hypothesis $H_0 : \theta = \theta_0$. The suggested test is

$$T_n = \left(\hat{\theta}_n - \theta_0\right)' H_n^* (\theta_0) (A_n^* (\theta_0))^{-1} H_n^* (\theta_0) \left(\hat{\theta}_n - \theta_0\right), \text{ if } \mu_0 = 0.$$

Similarly, for the $\mu \neq 0$ case, the suggested test statistic is

$$T_n^c = \left(\hat{\theta}_n - \theta_0\right)' H_n^{c*} (\theta_0) (A_n^{c*} (\theta_0))^{-1} H_n^{c*} (\theta_0) \left(\hat{\theta}_n - \theta_0\right),$$

where $H_n^{c*} (\theta_0)$ and A_n^{c*} are analogous to $H_n^* (\theta_0)$ and A_n^*, respectively, except that the former are based on $l_n^c (\theta_0)$ in place of $l_n (\theta_0)$ everywhere.

The statistics in (20) and (21) are vectors of normalized quadratic forms with mean zero and unit covariance matrix. In some special cases, such as in the stationary fixed coefficient setting, they are asymptotically $N (0, I_{m+1})$. Note that in the construction of T_n and T_n^c we have replaced $H_n^* \left(\bar{\theta}_n\right)$ and $H_n^{c*} \left(\bar{\theta}_n\right)$ by $H_n^* (\theta_0)$ and $H_n^{c*} (\theta_0)$. The validity of this step requires uniform boundedness of the normalized third-order log-likelihood derivatives, which is given in Lemma 7 of the Appendix. Formally, denote by F_q, F_q^c be the asymptotic distributions of the quadratic forms

$$q_n (\theta) = z_n^* (\theta_0)' (A_n^* (\theta_0))^{-1} z_n^* (\theta_0)$$

and

$$q_n^c (\theta) = z_n^{c*} (\theta_0)' (A_n^{c*} (\theta_0))^{-1} z_n^{c*} (\theta_0),$$

respectively. Let

$$b'_{0r} = M \dot{S}_{0r} S_0^{-1} 1.$$

We have:

Lemma 4 *Under Assumptions A0-A3:*

1. $[A_n]_{1,1} = \frac{1}{2\sigma_0^4} \left(1 + \frac{\kappa_4}{2\sigma_0^4}\right)$ *and* $[A_n^c]_{1,1} = [A_n]_{1,1} + o(1)$.
2. $[A_n]_{1,r} = 0$ *and* $[A_n^c]_{1,r} = o(1)$, $r = 2, ..., m+1$.
   For $r, s = 2, ..., m+1$:
3. $[A_n]_{r,s} = \frac{1}{2} (tr (\Lambda_{0r} \Lambda_{0s}))$, *if* $\mu_0 = 0$.

4. *If* $\mu_0 \neq 0$ *and* $\rho_n = O_e(1)$,

$$[A_n^c]_{r,s} = \frac{1}{4\sigma_0^4}\{2\sigma_0^4 tr(\Lambda_{0r}\Lambda_{0s}) + \kappa_4 \sum_{i=1}^{n} [\Gamma_{0r}]_{i,i}[\Gamma_{0s}]_{i,i} + 4\mu_0^2\sigma_0^2 b_{0r}' b_{0s} + 2\mu_0\kappa_3 \sum_{i=1}^{n} ([\Gamma_{0r}]_{i,i}[b_{0s}]_i + [\Gamma_{0s}]_{i,i}[b_{0r}]_i)\} + o(1).$$

5. $[A_n^c]_{r,s} = \frac{\mu_0^2}{\sigma_0^2} b_{0r}' b_{0s}$, *If* $\mu_0 \neq 0$ *and* $\rho_n = o(1)$.

Theorem 5 *Under Assumptions A0-A4, the statistics* T_n *and* T_n^c *are asymptotically distributed* F_q *and* F_q^c*, respectively.*

To construct a simple test of the form $H_0 : \theta_r = \theta_{0r}$, we may use

$$T_n = \left(\hat{\theta}_{nr} - \theta_{0r}\right)^2 \left[H_n^*(\theta_0)(A_n^*(\theta_0))^{-1} H_n^*(\theta_0)\right]_{r,r}, \text{ if } \mu_0 = 0,$$

or

$$T_n^c = \left(\hat{\theta}_{nr} - \theta_{0r}\right)^2 \left[H_n^{*c}(\theta_0)(A_n^{c*}(\theta_0))^{-1} H_n^{c*}(\theta_0)\right]_{r,r}, \text{ if } \mu_0 \neq 0.$$

The tests can be applied in principle by comparing the calculated T_n or T_n^c values to the simulated p-values of $q_n(\theta_0)$ or $q_n^c(\theta_0)$, respectively.

To complete the limit theory, we provide a consistency theorem for $\hat{\mu}_n$ and discuss its asymptotic distribution.

Theorem 6 *Under Assumptions A0-A3,* $\hat{\mu}_n \to_p \mu_0$.

As is clear from the proof of Theorem 6,

$$\sqrt{n}(\hat{\mu}_n - \mu_0) = \frac{1'\left(C_0 - C\left(\hat{\theta}_n\right)\right) y}{\sqrt{n}} + \sqrt{n}\bar{\varepsilon}_n. \tag{22}$$

Therefore, the asymptotic distribution of $\hat{\mu}_n$ depends critically on the behaviour of the first term on the rhs of (22) and may therefore be non-normal. To illustrate, consider the case where u_t is stochastic, $\beta_t(x_t, x_{t-1}; w)$ is given by (2), $\eta_t = (u_t, \varepsilon_t) \sim iid(0, \Sigma)$ with finite fourth moments, $\text{vech}(\Sigma) = (\sigma_u^2, \sigma_{u\varepsilon}, \sigma_\varepsilon^2)$, $n^{-1/2}\sum_{t=1}^{\lfloor nr \rfloor} \eta_t \Rightarrow B(r) \equiv \text{BM}(\Sigma)$, where $B = (B_u, B_\varepsilon)'$, $\mu_0 = 0$, and $w_0 = 0$. In this example and in what follows in Section 5 it is convenient to use the notation σ_ε^2 in place of σ^2 for $\mathbb{E}\left(\varepsilon_t^2\right)$. In this case, some detailed but standard weak convergence arguments reveal the following explicit limit theory for $\hat{w}_n$

$$\sqrt{n}\hat{w}_n \Rightarrow \xi := \frac{\sigma_{\varepsilon u} \int \tilde{B}_\varepsilon}{\left\{\sigma_u^2 + \mathbb{E}\left(\varepsilon_t u_t^2\right)\right\} \int \tilde{B}_\varepsilon^2}, \tag{23}$$

where $\tilde{B}_\varepsilon(r) = B_\varepsilon(r) - \int B_\varepsilon$ and all integrals are taken over the unit interval $[0,1]$. Then

$$\begin{aligned}
n^{-1/2} 1' \left(C_0 - C\left(\hat{\theta}_n\right)\right) y &= n^{-1/2} \sum_{t=1}^{n-1} \left(\beta_{t+1}(w_0) - \beta_{t+1}(\hat{w}_n)\right) Y_t \\
= \quad & n^{-1/2} \sum_{t=1}^{n-1} \left(1 - e^{\hat{w}_n u_{t+1}}\right) Y_t \\
= \quad & -\left(\sqrt{n}\hat{w}_n\right) \left(\frac{1}{n} \sum_{t=1}^{n-1} u_{t+1} Y_t\right) - \frac{\sigma_u^2 \left(\sqrt{n}\hat{w}_n\right)^2}{2} \left(\frac{1}{n^{3/2}} \sum_{t=1}^{n-1} Y_t\right) \\
& - \frac{\left(\sqrt{n}\hat{w}_n\right)^2}{2} \left(\frac{1}{n^{3/2}} \sum_{t=1}^{n-1} \left(u_{t+1}^2 - \sigma_u^2\right) Y_t\right) + o_p(1) \\
\sim \quad & -\xi \int B_\varepsilon dB_u - \frac{\sigma_u^2 \xi^2}{2} \int B_\varepsilon + o_p(1),
\end{aligned}$$

showing nonnormality in the limit theory and leading in this case to

$$\sqrt{n}\left(\hat{\mu}_n - \mu_0\right) \Rightarrow -\xi \int B_\varepsilon dB_u - \frac{\sigma_u^2 \xi^2}{2} \int B_\varepsilon + B_\varepsilon(1).$$

This example illustrates how the asymptotic distribution of $\sqrt{n}\left(\hat{\mu}_n - \mu_0\right)$ generally depends intimately on that of $\hat{\theta}_n$ when $\sigma_{u\varepsilon} \neq 0$. When $\sigma_{u\varepsilon} = 0$, $\sqrt{n}\left(\hat{\mu}_n - \mu_0\right) \Rightarrow B_\varepsilon(1)$ and the limit theory does not depend on that of $\hat{\theta}_n$.

In the nonstochastic case where $u_t = \frac{t}{n}$ we find in place of (23) the limit theory

$$n\hat{w}_n \Rightarrow \xi := \frac{\int r B_\varepsilon(r)\, dB_\varepsilon}{\int r^2 B_\varepsilon^2(r)}. \tag{24}$$

Then

$$n^{-1/2}1'\left(C_0 - C\left(\hat{\theta}_n\right)\right)y = n^{-1/2}\sum_{t=1}^{n-1}\left(1 - e^{\hat{w}_n \frac{t+1}{n}}\right)Y_t$$

$$= \quad -\left(n\hat{w}_n\right)\left(\frac{1}{n^{3/2}}\sum_{t=1}^{n-1}\frac{t+1}{n}Y_t\right) - \frac{\left(n\hat{w}_n\right)^2}{2}\left(\frac{1}{n^{5/2}}\sum_{t=1}^{n-1}\left(\frac{t+1}{n}\right)^2 Y_t\right)$$

$$\Rightarrow \quad -\xi\int rB_{\varepsilon}\left(r\right).$$

So $\sqrt{n}\left(\hat{\mu}_n - \mu_0\right) \Rightarrow B_{\varepsilon}\left(1\right) - \xi\int rB_{\varepsilon}\left(r\right)$ and the limit distribution is again non-normal and depends on that of $\hat{\theta}_n$.

4 Discussion

Some of the implications of these findings are as follows.

1. The normalization rates required for $l_n\left(\theta\right)$ and $l_n^c\left(\theta\right)$ in the proof of Theorem 1 are detailed in Table 1. For the case $\mu_0 = 0$, we require an $\left\|S_0^{-1}\right\|_F^{-2}$- or an $\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{-1}$- normalization when $\rho_n = O_e\left(1\right)$ and an $\left\|S_0^{-1}\right\|_F^{-2}$-normalization when $\rho_n = o\left(1\right)$. In terms of the fixed coefficient framework, as the unit root model is characterized by the condition $\rho_n = o\left(1\right)$, an $\left\|S_0^{-1}\right\|_F^{-2}$-normalization corresponds to an n^{-2}-normalization. For the fixed coefficient stable or explosive AR(1) model, both which are characterized by $\rho_n = O_e\left(1\right)$, an $\left\|S_0^{-1}\right\|_F^{-2}$- or an $\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{-1}$- normalization corresponds to an n^{-1}- and a β^{-2n}- normalization in the stable- and explosive cases, respectively. These rate results are well known – see Evans and Savin (1984), Phillips (1987) and Lemma 1 of Lieberman (2012).

2. In the $\mu_0 \neq 0$ case, an $\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{-1}$- normalization is required for the consistency proof based on $l_n^c\left(\theta\right)$, regardless of the order of magnitude of ρ_n. This corresponds to an n^{-1}- and a β^{-2n}- normalization in the stable- and explosive cases, respectively and an n^{-3}-normalization for the unit root model. Thus, unlike the stable and explosive cases in which the normalization is uniform in μ_0, different normalizations are required for the $\mu_0 = 0$ and $\mu_0 \neq 0$ cases in the unit root case.

3. To establish the behavior of the score, we use the normalizations summarized in Table 2. In the $\mu_0 = 0$ case, the score is a scalar multiple of QF_n and needs to be normalized by $\left\| S_0^{-1} \right\|_F^{-1}$. In the $\mu_0 \neq 0$ and $\rho_n = O_e(1)$ case, both QF_n^c and LF_n^c are of the same order of magnitude whereas in the $\mu_0 \neq 0$ and $\rho_n = o(1)$ case the term $QF_n / \left\| S^{-1\prime} S^{-1} \right\|_1^{1/2}$ is negligible. As the condition $\rho_n = o(1)$ characterizes a unit root type process, the vanishing of the latter term is indicative that the process is unit root and not a stable or an explosive process. Moreover, since LF_n is the dominant term in the $\mu_0 \neq 0$ and $\rho_n = o(1)$, case, if ε is Gaussian, the normalized score, being linear in ε, is also Gaussian. This is the case discussed in Lieberman (2012). For all other cases, the normalized score involves a quadratic form in ε and therefore is not asymptotically Gaussian. It is clear that in general it is not possible to determine the asymptotic distribution of the normalized score without additional structure on the model.

4. In the special case $\mu_0 = 0$ and a (fixed coefficient)unit root model, the normalized score given by $QF_n / \left\| S_0^{-1} \right\|_F$ is easily seen to converge to a $\left(\chi^2(1) - 1 \right)/2$ variate, which agrees with the result of Phillips (1987). See also Lieberman (2010).

5. The present setting assumes exogenous regressors Z_t. This assumption is common in time series regression where additional regressors are introduced and occurs in models such as the CAPM, ARMAX, and error correction models. In the following section, we consider a stochastic unit root model where the regressors may be dependent.

6. We have given two examples that show the asymptotic distribution of $\sqrt{n}(\hat{\mu}_n - \mu_0)$ may be non-normal. Furthermore, in sufficient generality the asymptotic distributions of T_n and T_n^c are non-normal. In both cases, simulation based approaches, such as the bootstrap, can be applied to generate p-values for hypothesis testing.

5 A Similarity STUR Model

As a further example of a similarity autoregression we consider a model that belongs to a class most closely related to the stochastic unit root (STUR) model studied in Granger and Swanson (1997). We use the no-intercept similarity autoregression

$$\begin{aligned} Y_1 &= \varepsilon_1, \\ Y_t &= \beta_t(w_n) Y_{t-1} + \varepsilon_t, \ t = 2, ..., n, \end{aligned} \tag{25}$$

with an exponential similarity function $\beta_t(x_t, x_{t-1}; w_n) = e^{w_n u_t}$, where in line with the notation following eq'n (2), $u_t = \Delta x_t$ is the source of the variation in the autoregressive coefficient. In this formulation, $w_n = \frac{a}{\sqrt{n}}$ is local to zero, so that

$$\beta_t = \exp\left(\frac{a}{\sqrt{n}} u_t\right) = 1 + \frac{a}{\sqrt{n}} u_t + O_p\left(\frac{1}{n}\right) \to 1, \tag{26}$$

is local to unity as $n \to \infty$. However, unlike the usual constant coefficient local to unity model where $\beta = \exp\left(\frac{a}{n}\right) \sim 1 + \frac{a}{n}$, the coefficient is stochastic and may therefore lie in the stationary or the explosive region, depending on the realization of u_t. Deviations from unity are $O_p\left(n^{-1/2}\right)$ in this model rather than deterministic and $O\left(n^{-1}\right)$. The model (25) and (26) is closely related to the STUR model of Granger and Swanson (1997) in which the autoregressive coefficient took the form $\beta_t = e^{\alpha_t}$ where α_t is generated independently of y_t by a stationary autoregression. Some of the properties of that model were studied in Granger and Swanson but no limit theory was provided.

With the localizing coefficient $w_n = \frac{a}{\sqrt{n}}$ in the time-varying representation $\beta_t = \exp(w_n \Delta x_t)$, the behavior of the model is autoregressive in the vicinity of unity and is amenable to functional limit theory, as we show below. This behavior may be directly analyzed as a stochastic alternative to either a unit root model or a constant local to unity model. As we will see, the limit behavior of the system is a nonlinear diffusion process rather than a linear diffusion process.

We shall assume that the moment generating function $M_u(s) = \mathbb{E}\left(e^{su_t}\right)$ of u_t is finite over some interval $s \in (-\delta, \delta)$ of the origin for $\delta > 0$. Solving

(25) we have

$$
\begin{aligned}
Y_t &= \varepsilon_t + \sum_{j=1}^{t-1} \left\{ \prod_{k=0}^{j} \beta_{t-k} \right\} \varepsilon_{t-j} = \varepsilon_t + \sum_{j=1}^{t-1} \left\{ \prod_{k=0}^{j} e^{\frac{a}{\sqrt{n}} u_{t-k}} \right\} \varepsilon_{t-j} \\
&= \varepsilon_t + \sum_{j=1}^{t-1} \left\{ e^{\frac{a}{\sqrt{n}} \sum_{k=0}^{j} u_{t-k}} \right\} \varepsilon_{t-j} = \varepsilon_t + \sum_{s=1}^{t-1} \left\{ e^{\frac{a}{\sqrt{n}} \sum_{\ell=0}^{t-s} u_{s+\ell}} \right\} \varepsilon_s .
\end{aligned}
$$

Observe that the impulse responses in this system are

$$
\frac{\partial Y_t}{\partial \varepsilon_{t-j}} = \prod_{k=0}^{j} \beta_{t-k} = e^{\frac{a}{\sqrt{n}} \sum_{k=0}^{j} u_{t-k}} = e^{a X_{nj}^u}, \tag{27}
$$

where $X_{nj}^u = n^{-1/2} \sum_{k=0}^{j} u_{t-k}$ is stochastic and a normalized partial sum process which wanders over $\mathbb{R}$. Hence, the impulse response function $\frac{\partial Y_t}{\partial \varepsilon_{t-j}} \in \mathbb{R}^+$ and may be arbitrarily small or arbitrarily large, the values being driven by the partial sum process X_{nj}^u.

As in the example following Theorem 6, we assume that partial sums of $\eta_t = (u_t, \varepsilon_t)'$ satisfy the invariance principle $n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} \eta_t \Rightarrow B(r) \equiv \mathrm{BM}(\Sigma)$, where $B = (B_u, B_\varepsilon)'$ is vector Brownian motion (BM) with positive definite variance matrix $\mathrm{vech}(\Sigma) = (\sigma_u^2, \sigma_{u\varepsilon}, \sigma_\varepsilon^2)$. Then, the asymptotic behavior of the time series Y_t in (25) has the following form

$$
\begin{aligned}
n^{-1/2} Y_{\lfloor nr \rfloor} &= n^{-1/2} \sum_{j=1}^{\lfloor nr \rfloor - 1} \left\{ e^{\frac{a}{\sqrt{n}} \sum_{k=0}^{j} u_{\lfloor nr \rfloor - k}} \right\} \varepsilon_{\lfloor nr \rfloor - j} + O_p\left(n^{-1/2}\right) \\
&= n^{-1/2} \sum_{s=1}^{\lfloor nr \rfloor - 1} \left\{ e^{\frac{a}{\sqrt{n}} \sum_{\ell=0}^{\lfloor nr \rfloor - s} u_{s+\ell}} \right\} \varepsilon_s + O_p\left(n^{-1/2}\right) \\
&= n^{-1/2} \sum_{s=1}^{\lfloor nr \rfloor - 1} \left\{ e^{\frac{a}{\sqrt{n}} \sum_{j=s}^{\lfloor nr \rfloor} u_j} \right\} \varepsilon_s + O_p\left(n^{-1/2}\right) \\
&\Rightarrow \int_0^r e^{a \int_p^r dB_u(q)} dB_\varepsilon(p) \\
&= e^{a B_u(r)} \int_0^r e^{-a B_u(p)} dB_\varepsilon(p) =: G_a(r) .
\end{aligned} \tag{28}
$$

The limit process $G_a(r)$ is a nonlinear Itō diffusion and satisfies the stochastic differential equation

$$dG_a(r) = aG_a(r)\,dB_u(r) + \frac{a^2\sigma_u^2}{2}G_a(r)\,dr + dB_\varepsilon(r),$$

showing the form of the drift and conditional volatility in the process that are induced by the similarity function $\beta_t(w_n) = \exp(w_n u_t)$ in (25). The quadratic variation of $G_a(r)$ is

$$\begin{aligned}[G_a]_r &= \sigma_\varepsilon^2 r + a^2\sigma_u^2\int_0^r G_a(s)^2\,ds + 2a\sigma_{u\varepsilon}\int_0^r G_a(s)\,ds \\ &= \int_0^r e_a(s)'\,\Sigma e_a(s)\,ds,\end{aligned}$$

where $e_a(s)' = (aG_a(s), 1)$. Evidently, realized and integrated volatility depend on the past history of the state variable $\{G_a(s), s \le r\}$.

If $\sigma_{u\varepsilon} = 0$, B_u is independent of B_ε and the limit process $G_a(r)$ is mixed Gaussian with the following covariance kernel conditional on $\mathcal{F}_u = \sigma\{B_u(s) : s \in [0,1]\}$

$$\begin{aligned}\gamma_{\mathcal{F}_u}(r,s) &= \mathbb{E}\{G_a(r)\,G_a(s)\,|_{\mathcal{F}_u}\} \\ &= e^{aB_u(r)}e^{aB_u(s)}\int_0^r\int_0^s e^{-aB_u(p_1)}e^{-aB_u(p_2)}\mathbb{E}\{dB_\varepsilon(p_1)\,dB_\varepsilon(p_2)\} \\ &= \sigma_\varepsilon^2 e^{aB_u(r)}e^{aB_u(s)}\int_0^{r\wedge s} e^{-2aB_u(p)}dp,\end{aligned}$$

and unconditional covariance kernel

$$\begin{aligned}
\gamma(r,s) &= \mathbb{E}\left\{\gamma_{\mathcal{F}_u}(r,s)\right\} = \sigma_\varepsilon^2 \mathbb{E}\left\{e^{aB_u(r)+aB_u(s)} \int_0^{r\wedge s} e^{-2aB_u(p)} dp\right\} \\
&= \sigma_\varepsilon^2 \int_0^{r\wedge s} \mathbb{E}\left\{e^{a[B_u(r)-B_u(p)]+a[B_u(s)-B_u(p)]}\right\} dp \\
&= \sigma_\varepsilon^2 \mathbb{E}\left\{e^{a[B_u(r\vee s)-B_u(r\wedge s)]}\right\} \int_0^{r\wedge s} \mathbb{E}\left\{e^{2a[B_u(r\wedge s)-B_u(p)]}\right\} dp \\
&= \sigma_\varepsilon^2 e^{\frac{1}{2}a^2(r\vee s - r\wedge s)\sigma_u^2} \int_0^{r\wedge s} e^{2a^2[r\wedge s - p]\sigma_u^2} dp = \sigma_\varepsilon^2 e^{\frac{1}{2}a^2(r\vee s - r\wedge s)\sigma_u^2}\left[-\frac{e^{2a^2[r\wedge s-p]\sigma_u^2}}{2a^2\sigma_u^2}\right]_0^{r\wedge s} \\
&= \sigma_\varepsilon^2 \frac{e^{2a^2\sigma_u^2 r\wedge s}-1}{2a^2\sigma_u^2} e^{\frac{1}{2}a^2\sigma_u^2(r\vee s - r\wedge s)}.
\end{aligned}$$

Observe that when $a^2\sigma_u^2 \to 0$, $\gamma(r,s) \to \sigma_\varepsilon^2 r \wedge s$, the covariance kernel of the Brownian motion B_ε, corresponding to the case where $\beta_t = 1$, $G_a(r) \to B_\varepsilon(r)$, and (25) is a random walk. Thus, for small $|a|$ or σ_u^2, the model is local to a simple unit root model. When $a \to \pm\infty$, the limit behavior of $Y_{\lfloor nr \rfloor}$ is more complex. In particular, the rates of convergence change, and the limit results become path dependent. This case deserves further study and will be investigated in later work.

If $\sigma_{u\varepsilon} \neq 0$, B_u and B_ε are dependent, and the limit process $G_a(r)$ is no longer conditionally Gaussian. Instead, we have

$$\begin{aligned}
G_a(r) &= e^{aB_u(r)} \int_0^r e^{-aB_u(p)} dB_\varepsilon(p) = e^{aB_u(r)} \int_0^r e^{-aB_u(p)} dB_\varepsilon(p) \\
&= e^{aB_u(r)} \int_0^r e^{-aB_u(p)} dB_{\varepsilon.u}(p) + \frac{\sigma_{u\varepsilon}}{\sigma_u^2} e^{aB_u(r)} \int_0^r e^{-aB_u(p)} dB_u(p) \\
&=: G_{a,\varepsilon.u}(r) + \frac{\sigma_{u\varepsilon}}{\sigma_u^2} G_{a,u}(r),
\end{aligned}$$

Here $B_{\varepsilon.u}(r) = B_\varepsilon(r) - \frac{\sigma_{u\varepsilon}}{\sigma_u^2} B_u$ is BM($\sigma_{\varepsilon.u}^2$) with $\sigma_{\varepsilon.u}^2 = \sigma_\varepsilon^2 - \sigma_{u\varepsilon}^2/\sigma_u^2$, and $G_{a,\varepsilon.u}(r)$ is a conditional Gaussian (Itō diffusion) process with conditional covariance kernel

$$\gamma_{\varepsilon.u,\mathcal{F}_u}(r,s) = \mathbb{E}\left\{G_{a,\varepsilon.u}(r)\, G_{a,\varepsilon.u}(s) \,|\mathcal{F}_u\right\} = \sigma_{\varepsilon.u}^2 e^{aB_u(r)} e^{aB_u(s)} \int_0^{r\wedge s} e^{-2aB_u(p)} dp,$$

and unconditional covariance kernel

$$\gamma_{\varepsilon.u}(r,s) = \mathbb{E}\left\{G_{a,\varepsilon.u}(r)\,G_{a,\varepsilon.u}(s)\right\} = \sigma^2_{\varepsilon.u}\frac{e^{2a^2\sigma_u^2 r\wedge s}-1}{2a^2\sigma_u^2}e^{\frac{1}{2}a^2(r\vee s-r\wedge s)\sigma_u^2}.$$

The process $G_{a,u}(r) = e^{aB_u(r)}\int_0^r e^{-aB_u(p)}dB_u(p)$ is a nonlinear stochastic integral of B_u and obviously non Gaussian when $a \neq 0$.

For $t = \lfloor nr \rfloor$ for some $r \in (0,1]$ and large $j = \lfloor n\kappa \rfloor$, $\kappa > 0$, the impulse responses (27) have the form

$$\frac{\partial Y_{\lfloor nr\rfloor}}{\partial \varepsilon_{\lfloor nr\rfloor - \lfloor n\kappa\rfloor}} = \prod_{k=0}^{\lfloor n\kappa\rfloor}\beta_{t-k} = e^{\frac{a}{\sqrt{n}}\sum_{k=0}^{\lfloor n\kappa\rfloor}u_{\lfloor nr\rfloor-k}} \Rightarrow e^{a\int_{r-\kappa}^{r}dB_u(s)} = e^{a[B_u(r)-B_u(r-\kappa)]},$$

which may be arbitrarily small, close to unity, or arbitrarily large depending on the historical trajectory of the process B_u over the past interval $[r-\kappa, r]$.

The functional law (28) enables us to derive the limit behavior of statistics arising from the model (25) and (26). For example, we may consider conventional unit root tests applied to model (25) and (26). Observe that least squares applied to (25) gives

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{t=1}^n Y_tY_{t-1}}{\sum_{t=1}^n Y_{t-1}^2} = \frac{\sum_{t=1}^n e^{\frac{a}{\sqrt{n}}u_t}Y_{t-1}^2}{\sum_{t=1}^n Y_{t-1}^2} + \frac{\sum_{t=1}^n Y_{t-1}\varepsilon_t}{\sum_{t=1}^n Y_{t-1}^2} \\
&= \frac{\sum_{t=1}^n \mathbb{E}\left[e^{\frac{a}{\sqrt{n}}u_t}\right]Y_{t-1}^2}{\sum_{t=1}^n Y_{t-1}^2} + \frac{\sum_{t=1}^n Y_{t-1}\varepsilon_t}{\sum_{t=1}^n Y_{t-1}^2} + \frac{\sum_{t=1}^n\left\{e^{\frac{a}{\sqrt{n}}u_t} - \mathbb{E}\left[e^{\frac{a}{\sqrt{n}}u_t}\right]\right\}Y_{t-1}^2}{\sum_{t=1}^n Y_{t-1}^2} \\
&= M\left(\frac{a}{\sqrt{n}}\right) + \frac{\sum_{t=1}^n Y_{t-1}\varepsilon_t}{\sum_{t=1}^n Y_{t-1}^2} + \frac{\sum_{t=1}^n\left\{e^{\frac{a}{\sqrt{n}}u_t} - M\left(\frac{a}{\sqrt{n}}\right)\right\}Y_{t-1}^2}{\sum_{t=1}^n Y_{t-1}^2}
\end{aligned}$$

so that

$$n\left[\hat{\beta} - M\left(\frac{a}{\sqrt{n}}\right)\right] = \frac{n^{-1}\sum_{t=1}^n Y_{t-1}\varepsilon_t}{n^{-2}\sum_{t=1}^n Y_{t-1}^2} + \frac{n^{-1}\sum_{t=1}^n\left\{e^{\frac{a}{\sqrt{n}}u_t} - M\left(\frac{a}{\sqrt{n}}\right)\right\}Y_{t-1}^2}{n^{-2}\sum_{t=1}^n Y_{t-1}^2}$$

By standard weak convergence arguments we have

$$\frac{n^{-1}\sum_{t=1}^n Y_{t-1}\varepsilon_t}{n^{-2}\sum_{t=1}^n Y_{t-1}^2} \Rightarrow \frac{\int_0^1 G_a(r)\,dB_\varepsilon(r)}{\int_0^1 G_a(r)^2\,dr}.$$

Expanding the moment generating function we have

$$
\begin{aligned}
\eta_{n,t}(a) &= e^{\frac{a}{\sqrt{n}}u_t} - M\left(\frac{a}{\sqrt{n}}\right) \\
&= \left\{1 + \frac{a}{\sqrt{n}}u_t + \frac{1}{2}\left(\frac{a}{\sqrt{n}}\right)^2 u_t^2\right\} - \left\{1 + \frac{1}{2}\left(\frac{a}{\sqrt{n}}\right)^2 \sigma_u^2\right\} + o_p\left(n^{-1}\right) \\
&= \frac{a}{\sqrt{n}}u_t + \frac{a^2}{2n}\left(u_t^2 - \sigma_u^2\right) + o_p\left(n^{-1}\right),
\end{aligned}
$$

so that by standard arguments

$$
\begin{aligned}
\frac{n^{-1}\sum_{t=1}^n Y_{t-1}^2 \eta_{n,t}(a)}{n^{-2}\sum_{t=1}^n Y_{t-1}^2} &\sim a\frac{\sum_{t=1}^n \left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2 \frac{u_t}{\sqrt{n}}}{n^{-1}\sum_{t=1}^n \left(\frac{Y_{t-1}}{\sqrt{n}}\right)^2} + O_p\left(n^{-1/2}\right) \\
&\Rightarrow a\frac{\int_0^1 G_a(r)^2 \, dB_u(r)}{\int_0^1 G_a(r)^2 \, dr},
\end{aligned}
$$

which leads to the limit theory

$$
n\left[\hat{\beta} - M\left(\frac{a}{\sqrt{n}}\right)\right] \Rightarrow \frac{\int_0^1 G_a(r) \, dB_\varepsilon(r)}{\int_0^1 G_a(r)^2 \, dr} + a\frac{\int_0^1 G_a(r)^2 \, dB_u(r)}{\int_0^1 G_a(r)^2 \, dr}.
$$

Correspondingly, with a unit root centering, we have

$$
n\left[\hat{\beta} - 1\right] \Rightarrow \frac{1}{2}a^2\sigma_u^2 + \frac{\int_0^1 G_a(r) \, dB_\varepsilon(r)}{\int_0^1 G_a(r)^2 \, dr} + a\frac{\int_0^1 G_a(r)^2 \, dB_u(r)}{\int_0^1 G_a(r)^2 \, dr} \tag{29}
$$

It follows that standard coefficient based UR tests have only local discriminatory power against a stochastic UR of the form (25). Similar results apply for t-ratio based and other UR tests.

6 Simulations and Empirics

6.1 Simulations

To evaluate the behavior of the estimators, simulations were conducted on the model (1) with $\beta_t = \exp(wZ_t)$ for $n = 250$, 500, 1000, $\mu = 0$, 0.25, and $w = 0.07$, 0.2, with 2000 replications per experiment. In one setting we took $Z_t \sim NID(-w/2, 1)$ and in the other $Z_t \sim U[-1, 1] + b_w$, $b_w = -0.0116648$ if $w = 0.07$ and $b_w = -0.033289$ if $w = 0.2$. With these choices $\mathbb{E}_Z(\beta_t) = 1$, $\forall t$, but for each t, β_t can be greater than- equal to- or less than unity. Means and standard deviations of $\hat{w}_n$, $\hat{\mu}_n$ and $\hat{\beta}_t$ for all the scenarios considered here are summarized in Tables 3 and 4.

Overall, with no noticeable differences between the cases, the means are very close to the true values and the estimated standard deviations decline with n, as is expected, corroborating consistency.

6.2 Financial Data Application

We consider eight country Exchange Traded Funds (ETFs), denoted by P_t, traded in the U.S. and measured in 15-minutes intervals. The model is

$$\begin{aligned} P_t &= \exp\{\Delta NAV_t (w_1 + w_2 D_t) + \Delta SP_t (w_3 + w_4 D_t) \\ &\quad + ECT_{t-1} (w_5 + w_6 D_t)\} P_{t-1} + \varepsilon_t, \end{aligned} \tag{30}$$

where NAV_t is the net asset value, SP_t is the S&P500 index, ECT_t is an error correction term, equal to $P_t - NAV_t$, D_t is a dummy variable, taking the value of unity if t is the U.S. market-open time and zero otherwise, and all variables (apart from D_t) were transformed by a natural logarithm. The data are available for Australia, Hong Kong, Japan, Malaysia, Singapore, Taiwan, South Korea and China. The tickers for these countries are given by EWA, EWH, EWJ, EWM, EWS, EWT, EWY and FXI, respectively. The sample range is 12/15/2000-12/13/2010, apart for China, where it is available for 10/8/2004-12/13/2010. The data is discussed in detail in Levy and Lieberman (2012).

The exponential similarity function in (30) satisfies Assumptions A0-A4 of the Appendix. Unit root tests for P_t with and without a constant are reported in Table 5. As expected, the p-values in all cases are very high and they are generally higher for the version of the ADF test which does not include a constant. Thus, the (fixed coefficient) unit root null hypothesis cannot be rejected, although the underlying process may well include a volatile lag dependent variable coefficient.

The estimated model results are given in Table 6, with M1 and M2 denoting the unit root with a drift model and model (30), respectively. For the former, the slope coefficient equals unity to the third decimal place throughout, whereas the drift parameter is very small with large standard errors. Thus, a driftless unit root model seems to be a reasonable approximation to the dgp. Nevertheless, the average AIC criterion for M1 equals -7.949, whereas for M2, it is -9.618. To the third decimal place the SC averages are almost identical to the AIC averages and are therefore omitted.

Across all cases, the error correction coefficient estimates $\hat{w}_5$ and $\hat{w}_6$ are negative, with cross country sample averages equal to -0.004 and -0.230, respectively. These results imply that, *ceteris paribus,* when the error correction term is positive so that $P_{t-1} > NAV_{t-1}$, there will be a downward correction to P_t. The model thus has a time varying coefficient together with an embedded error correction mechanism that is a nonlinear driver of the coefficient of the lagged dependent variable.

The graph of $\hat{\beta}_t$ for EWA is displayed in Figure 1, exhibiting volatility and fluctuating around unity, as expected. Other cases look very similar. Finally, the standard errors are based on Eviews' calculation of the outer products of the score and are only indicative for this application. In principle, the suggested test procedure can be applied and compared to bootstrapped p-values. Overall, the random walk model appears to be a reasonable approximation to (30) but it does not reflect any of the period by period variation captured in Figure 1 or potential drivers of that variation.

7 Conclusions

We investigated time varying autoregressions in which variation in the coefficient of the lag dependent variable is driven by a similarity function. A key feature of this model is that the slope coefficient can be equal to, less than, or greater than unity at any point in time, giving the model a high degree of flexibility in the autoregressive response. Consistency of the QMLE of the parameter vector was established together with a complete taxonomy of the required norming rates and standardization of the score and Hessian functions for the different cases. A local to zero similarity-based STUR system was introduced and a new limit theory was established in which the limit process, $G_a(r)$, is a nonlinear Itō diffusion process. The similarity function impacts this model by inducing drift and conditional volatility in the limit process, showing how the flexible autoregressive response in a STUR system can be the source of both drift and volatility.

Our simulations show that the QMLE performs well for sample sizes ranging from 250 to 1000. The model is illustrated empirically in an application to international ETF data. While a unit root model is not rejected, the time varying coefficient characteristics of the autoregressive responses are vividly apparent in the sample data, showing both mildly explosive and mildly integrated realizations.

Appendix

The parameter space is given by $\Theta = \Theta_1 \times \Theta_2$, where Θ_1, Θ_2 are the spaces in which σ^2 and w are assumed to lie, respectively. By K we denote a generic bounding constant, independent of n, which may vary from step to step. In the following we enlist the assumptions used for our model.

Assumption A0: $\{\varepsilon_t\}_{t=1}^n$ *is a sequence of iid continuous random variables, each with a zero mean, variance* σ^2*, cumulants* κ_r*,* $r \geq 3$ *and moment generating function which converges in a narrow strip containing the origin. If* $w \neq w'$*,* $\beta_t(w) \neq \beta_t(w')$*,* $\forall t$*.* The matrix $X = (X_{it})_{1 \leq i \leq m, 1 \leq t \leq n}$ is nonstochastic, real and finite.

Assumption A1: *There exist* σ_L^2*,* σ_H^2*,* w_L *and* w_H*, such that* $\sigma_0^2 \in [\sigma_L^2, \sigma_H^2]$*, with* $0 < \sigma_L^2 < \sigma_H^2 < \infty$ *and for each* $i = 1, ..., m$*,* $w_{i,0} \in [w_L, w_H]$*, with* $-\infty < w_L < w_H < \infty$*. In addition,* $\mu_0 \in R$*.*

Assumption A2: *For all* $1 < t \leq n$*, the function* $\beta_t(w)$ *is non-negative, continuous and is three times continuously differentiable.*

Let $C_0 = C(w_0)$, so that $S_0 = I_n - C_0$. For $r, s, t = 2, ..., m+1$, set

$$\dot{C}_r(w) = \partial C(w) / \partial \theta_r, \; \ddot{C}_{r,s}(w) = \partial^2 C(w) / \partial \theta_r \partial \theta_s$$

and

$$\dddot{C}_{r,s,t}(w) = \partial^3 C(w) / \partial \theta_r \partial \theta_s \partial \theta_t.$$

Assumption A3: *For all* $2 \leq r \leq m+1$*,* $1 \leq i, j \leq n$*,* $w \in \Theta_2 \subset R^m$*,* there exists a $0 < K_L < \infty$, such that

$$[C]_{i,j} \leq K [C_0]_{i,j},$$

$$K_L \left[C_0\right]_{i,j} \leq \left| \left[\dot{C}_r \left(w\right)\right]_{i,j} \right| \leq K \left[C_0\right]_{i,j}.$$

Assumption A4: *For all* $2 \leq r, s, t \leq m+1$, $1 \leq i, j \leq n$, $w \in \Theta_2 \subset R^m$, there exists a $0 < K_L < \infty$, such that

$$K_L \left[C_0\right]_{i,j} \leq \left| \left[\ddot{C}_{r,s} \left(w\right)\right]_{i,j} \right| \leq K \left[C_0\right]_{i,j}$$

and

$$K_L \left[C_0\right]_{i,j} \leq \left| \left[\dddot{C}_{r,s,t} \left(w\right)\right]_{i,j} \right| \leq K \left[C_0\right]_{i,j}.$$

Assumption A0 includes an identification condition. If $\beta_t = \beta$, $\forall t$, then this condition trivially holds. Assumptions A0–A4 are similar to those of Lieberman (2012), the key difference being that μ_0 is allowed to be zero here. It is trivial to verify that all the assumptions hold for the exponential similarity function, because, if $\beta_t \left(w\right) = \exp\left(\sum_{j=1}^{m} w_j \Delta X_{tj}\right)$, then $\partial \beta_t \left(w\right) / \partial w_j^r = \left(\Delta X_{tj}\right)^r \beta_t \left(w\right)$.

The following inequalities will be used throughout (see, among others, Graybill (1983)).

$$\begin{aligned} \left\|A\right\|_2 &\leq \left\|A\right\|_F \leq \sqrt{n} \left\|A\right\|_2, \\ x'Ax &\leq x'x \left\|A\right\|_2, \text{ for } A > 0, \ \left|tr\left(AB\right)\right| \leq \left\|A\right\|_F \left\|B\right\|_F, \\ \left\|AB\right\|_F &\leq \left\|A\right\|_2 \left\|B\right\|_F, \ \left\|AB\right\|_F \leq \left\|A\right\|_F \left\|B\right\|_2. \end{aligned} \tag{31}$$

Proof of Theorem 1: The proof for the case $\mu_0 \neq 0$ was given by Lieberman (2012). The case $\mu_0 = 0$ requires a different normalization and we deal with it here. For any $\delta_1 > 0$, denote by $B_{\delta_1}\left(\theta_0\right)$ the ball $\left\{\theta \in \Theta : \left|\left|\theta - \theta_0\right|\right| \leq \delta_1\right\}$ and by $B_{\delta_1}^c\left(\theta_0\right)$ the complement of $B_{\delta_1}\left(\theta_0\right)$ in Θ. Using Wu's (1981) criterion, it is sufficient to show that $\forall \delta_1 > 0$,

$$\liminf_{n \to \infty} \inf_{B_{\delta_1}^c\left(\theta_0\right)} n^{-1} \left(l_n\left(\sigma_0^2, \theta_{20}'\right) - l_n\left(\sigma^2, \theta_{20}'\right)\right), \tag{32}$$

and

$$\liminf_{n \to \infty} \inf_{B_{\delta_1}^c\left(\theta_0\right)} \left\|S_0^{-1}\right\|_F^{-2} \left(l_n\left(\sigma_0^2, \theta_{20}'\right) - l_n\left(\sigma_0^2, \theta_2'\right)\right) \tag{33}$$

are strictly positive in probability. Now,

$$\mathbb{E}_{\theta_0}\left(n^{-1}l_n\left(\sigma^2,\theta_{20}'\right)\right)=-\frac{1}{2}\log\left(2\pi\sigma^2\right)-\frac{\sigma_0^2}{2\sigma^2},$$

and

$$Var_{\theta_0}\left(n^{-1}l_n\left(\sigma^2,\theta_{20}'\right)\right)=\frac{1}{2\sigma^4 n}\left(\sigma_0^4+\frac{\kappa_4}{2}\right).$$

Hence,

$$n^{-1}\left(l_n\left(\sigma_0^2,\theta_{20}\right)-l_n\left(\sigma^2,\theta_{20}'\right)\right)\to_p\frac{1}{2}\left(\frac{\sigma_0^2}{\sigma^2}-1-\log\left(\frac{\sigma_0^2}{\sigma^2}\right)\right)\geq 0,$$

with equality iff $\sigma^2=\sigma_0^2$.

To establish (33), we use the decomposition

$$\left(l_n\left(\sigma_0^2,\theta_{20}'\right)-l_n\left(\sigma_0^2,\theta_2'\right)\right)=\frac{1}{2\sigma_0^2}\left(y'S'Sy-y'S_0S_0y\right)$$

$$\begin{aligned}
&= \frac{1}{2\sigma_0^2}y'S_0'\left(S_0^{-1\prime}G'+GS_0^{-1}\right)S_0y \\
&\quad +\frac{1}{2\sigma_0^2}y'G'Gy, \\
&= Q_{1n}+Q_{2n}, \qquad (34)
\end{aligned}$$

say, where $G=S-S_0$. We have,

$$Q_{1n}=O_p\left(\left\|S_0^{-1}\right\|_F\right). \qquad (35)$$

Considering Q_{2n}, we have

$$\mathbb{E}_{\theta_0}\left(y'G'Gy\right)=\mathbb{E}_{\theta_0}\left(\varepsilon'S_0^{-1\prime}G'GS_0^{-1}\varepsilon\right)\leq K\left\|S_0^{-1}\right\|_F^2,$$

and

$$Var_{\theta_0}\left(y'G'Gy\right)\leq K\left\|S_0^{-1}\right\|_F^4$$

so that

$$Q_{2n}=O_p\left(\left\|S_0^{-1}\right\|_F^2\right).$$

To complete the proof, we see that $Q_{2n}\geq 0$, because $G'G$ is positive semidefinite. Now, with $g_{\min} = \min_{2\leq i\leq n} [G]_{i,i-1}$,

$$\begin{aligned}
\frac{\mathbb{E}_{\theta_0}\left(\varepsilon' S_0^{-1\prime} G' G S_0^{-1}\varepsilon\right)}{\left\|S_0^{-1}\right\|_F^2} &= \frac{\sigma_0^2 tr\left(S_0^{-1\prime} G' G S_0^{-1}\right)}{\left\|S_0^{-1}\right\|_F^2} \\
&= \frac{\sigma_0^2 \left\|G S_0^{-1}\right\|_F^2}{\left\|S_0^{-1}\right\|_F^2}
\end{aligned}$$

$$\begin{aligned}
&\geq \sigma_0^2 g_{\min}^2 \frac{\sum_{i,j=1}^{n-1}\left[S_0^{-1}\right]_{i,j}^2}{\sum_{i,j=1}^{n}\left[S_0^{-1}\right]_{i,j}^2} \\
&\geq K_L,
\end{aligned}$$

for some $K_L > 0$ which is independent of n. As $\left\|S_0^{-1}\right\|_F^2 \geq n$, Q_{2n} strictly dominates Q_{1n} and because y is continuous, $Q_{2n} \geq 0$ and $\mathbb{E}_{\theta_0}\left(Q_{2n}/\left\|S_0^{-1}\right\|_F^2\right) > 0$, $Q_{2n}/\left\|S_0^{-1}\right\|_F^2$ is strictly positive in probability uniformly in $B_{\delta_1}^c(\theta_0)$, as required. ■

Proof of Theorem 2:

Case 1: $\mu_0 = 0$. The score with respect to σ^2 is given by

$$z_{n1}(\theta_0) = -\frac{\sqrt{n}}{2\sigma_0^2} + \frac{y' S_0' S_0 y}{2\sigma_0^4\sqrt{n}}. \tag{36}$$

Hence,

$$\mathbb{E}_{\theta_0}(z_{n1}(\theta_0)) = 0$$

and

$$Var_{\theta_0}(z_{n1}(\theta_0)) = \frac{1}{2\sigma_0^4}\left(1 + \frac{\kappa_4}{2\sigma_0^4}\right). \tag{37}$$

Higher order cumulants of $z_{n1}(\theta_0)$ tend to zero, so part 1 of the Theorem is done. For $r = 2, ..., m+1$,

$$\begin{aligned}\frac{\partial l_n(\theta_0)}{\partial \theta_r} &= -\frac{y'\left(\dot{S}_{0r}'S_0 + S_0'\dot{S}_{0r}\right)y}{2\sigma_0^2} \\ &= -\frac{\varepsilon'\left(\dot{S}_{0r}S_0^{-1} + S_0^{-1\prime}\dot{S}_{0r}'\right)\varepsilon}{2\sigma_0^2}. \end{aligned} \tag{38}$$

It follows that

$$\varepsilon'\left(\dot{S}_{0r}S_0^{-1} + S_0^{-1\prime}\dot{S}_{0r}'\right)\varepsilon = O_p\left(\left\|S_0^{-1}\right\|_F\right). \tag{39}$$

By Assumption A3,

$$\left\|\dot{C}_{0r}S_0^{-1}\right\|_F \geq K_L\left\|S_0^{-1} - I_n\right\|_F.$$

Hence,

$$\begin{aligned}\frac{\left\|S_0^{-1}\right\|_F^2}{tr\left(S_0^{-1\prime}\dot{S}_{0r}'\dot{S}_{0r}S_0^{-1}\right)} &= \frac{\left\|S_0^{-1}\right\|_F^2}{\left\|\dot{S}_{0r}S_0^{-1}\right\|_F^2} \\ &\leq K_L\frac{\left\|S_0^{-1}\right\|_F^2}{\left\|S_0^{-1} - I\right\|_F^2}\end{aligned}$$

$$\begin{aligned}&= K_L\frac{\sum_{i,j=1}^n\left[C_0 + \cdots + C_0^{m-1}\right]_{i,j}^2 + n}{\sum_{i,j=1}^n\left[C_0 + \cdots + C_0^{m-1}\right]_{i,j}^2} \\ &< K_L\left(1 + \frac{n}{\sum_{i,j=1}^n\left[C_0\right]_{i,j}^2}\right) \\ &\leq K.\end{aligned}$$

Therefore, there exists a c, such that,

$$\frac{1}{\left\|S_0^{-1}\right\|_F^2}Var_{\theta_0}\left(\varepsilon'\left(\dot{S}_{0r}S_0^{-1} + S_0^{-1\prime}\dot{S}_{0r}'\right)\varepsilon\right) > c > 0,$$

implying that the required normalization for $\partial l_n(\theta_0)/\partial\theta_r$, $r=2,...,m+1$, is $\left\|S_0^{-1}\right\|_F^{-1}$.

Case 2: $\mu_0 \neq 0$. The score wrt σ^2 has been dealt with by Lieberman (2012). The concentrated score wrt θ_r, $r=2,...,m+1$, is given by

$$\frac{\partial l_n^c(\theta_0)}{\partial\theta_r} = -\frac{1}{2\sigma_0^2}\left(QF_n^c + 2\mu_0 LF_n^c\right), \tag{40}$$

where QF_n^c and LF_n^c are given by (14) and (15). We know from Lieberman (2012) that

$$QF_{nr}^c = O_p\left(\left\|S_0^{-1}\right\|_F\right) \tag{41}$$

and that

$$LF_{nr}^c = O_p\left(\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}\right). \tag{42}$$

Therefore, we need to distinguish between two subcases.

Subcase 2(i): $\rho_n = O_e(1)$. In this subcase, we can normalize the score by $\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{-1/2}$ to obtain

$$z_{nr}^c(\theta_0) = O_p(1) + O_p\left(\sqrt{\rho_n}\right).$$

But in this case, $O_{p,e}\left(\sqrt{\rho_n}\right) = O_{p,e}(1)$, and because $\left\|S_0^{-1}\right\|_F \geq K_L\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}$, for some $1 > K_L > 0$, obtaining the lower bound on the variance of $QF_n^c/\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}$ is similar to the derivation of the lower bound in the $\mu_0 = 0$ case and therefore, this subcase is done.

Subcase 2(ii): $\rho_n = o(1)$. By (40)-(42), in this subcase

$$z_{nr}^c(\theta_0) = \frac{LF_n^c}{\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}} + o_p(1).$$

The lower bound on $LF_n^c/\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2}$ was established in Lieberman (2012) and therefore the proof of Theorem 2 is completed.■

Proof of Theorem 3:

Case 1: $\mu_0 = 0$. It is straightforward to verify part (1) of the Theorem for this case. Part (2) is established on the observation that

$$E_{\theta_0}\left(\varepsilon'\left(\dot{S}_{0r}S_0^{-1} + S_0^{-1\prime}\dot{S}_{0r}'\right)\varepsilon\right) = 0$$

and

$$
\begin{aligned}
Var_{\theta_0}\left(\varepsilon'\left(\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'\right)\varepsilon\right) &= 4\sigma_0^4 tr\left(S_0^{-1\prime}\dot{S}_{0r}'\dot{S}_{0r}S_0^{-1}\right) \\
&\quad +\kappa_4\sum\left[\dot{S}_{0r}S_0^{-1}+S_0^{-1\prime}\dot{S}_{0r}'\right]_{i,i}^2 \\
&\leq 4\sigma_0^4\left\|S_0^{-1}\right\|_F^2\sup_t\dot{\beta}_{0t}^2 \\
&\leq K\left\|S_0^{-1}\right\|_F^2.
\end{aligned}
$$

because $\dot{S}_{0r}S_0^{-1}$ has zero diagonal elements. For $2\leq r,s\leq m+1$,

$$
\begin{aligned}
H_{nr,s}(\theta_0) &= -\frac{y'\left(\ddot{S}_{0r,s}'S_0+\dot{S}_{0r}'\dot{S}_{0s}+\dot{S}_{0s}'\dot{S}_{0r}+S_0'\ddot{S}_{0r,s}\right)y}{2\sigma_0^2\left\|S_0^{-1}\right\|_F^2} \\
&= -\frac{1}{2\sigma_0^2\left\|S_0^{-1}\right\|_F^2}\{\varepsilon'\left(S_0^{-1\prime}\ddot{S}_{0r,s}'+\ddot{S}_{0r,s}S_0^{-1}\right)\varepsilon \\
&\quad +2\varepsilon'S_0^{-1\prime}\dot{S}_{0s}'\dot{S}_{0r}S_0^{-1}\varepsilon\}.
\end{aligned} \tag{43}
$$

The first term in (43) has

$$
E_{\theta_0}\left(\varepsilon\left(S_0^{-1\prime}\ddot{S}_{0r,s}'+\ddot{S}_{0r,s}S_0^{-1}\right)\varepsilon\right)=0
$$

and

$$
\begin{aligned}
Var_{\theta_0}\left(\varepsilon\left(S_0^{-1\prime}\ddot{S}_{0r,s}'+\ddot{S}_{0r,s}S_0^{-1}\right)\varepsilon\right) &= 2\sigma_0^4 tr\left(S_0^{-1\prime}\ddot{S}_{0r,s}'+\ddot{S}_{0r,s}S_0^{-1}\right)^2 \\
&\quad +\kappa_4\sum_{i=1}^n\left[S_0^{-1\prime}\ddot{S}_{0r,s}'+\ddot{S}_{0r,s}S_0^{-1}\right]_{i,i}^2.
\end{aligned} \tag{44}
$$

The matrix $\ddot{S}_{0r,s}S_0^{-1}$ is lower triangular with zero diagonal elements and therefore the lhs of (44) is bounded by

$$
4\sigma_0^4 tr\left(S_0^{-1\prime}\ddot{S}_{0r,s}'\ddot{S}_{0r,s}S_0^{-1}\right)=4\sigma_0^4\left\|\ddot{S}_{0r,s}S_0^{-1}\right\|_F^2\leq K\left\|\ddot{S}_{0r,s}\right\|_2^2\left\|S_0^{-1}\right\|_F^2.
$$

Under Assumption A4,

$$\left\|\ddot{S}_{0r,s}\right\|_2^2 = \sup_{|x|=1} x'\ddot{S}'_{0r,s}\ddot{S}_{0r,s}Ax = \sup_{|x|=1} x'\ddot{C}'_{0r,s}\ddot{C}_{0r,s}Ax \leq \sup_t \ddot{\beta}_{tr,s} \leq K \quad (45)$$

and therefore

$$\frac{\varepsilon'\left(S_0^{-1\prime}\ddot{S}'_{0r,s} + \ddot{S}_{0r,s}S_0^{-1}\right)\varepsilon}{2\sigma_0^2\left\|S_0^{-1}\right\|_F^2} = O_p\left(\left\|S_0^{-1}\right\|_F^{-1}\right). \quad (46)$$

The second term on the rhs of (43) has

$$\begin{aligned}\left|E_{\theta_0}\left(2\varepsilon' S_0^{-1\prime}\dot{S}'_{0s}\dot{S}_{0r}S_0^{-1}\varepsilon\right)\right| &= \sigma_0^2\left|tr\left(S_0^{-1\prime}\left(\dot{S}'_{0s}\dot{S}_{0r} + \dot{S}'_{0r}\dot{S}_{0s}\right)S_0^{-1}\right)\right| \\ &= \sigma_0^2\left\|\left(\dot{S}'_{0s}\dot{S}_{0r} + \dot{S}'_{0r}\dot{S}_{0s}\right)^{1/2}S_0^{-1}\right\|_F^2 \\ &\leq \sigma_0^2\left\|\left(\dot{S}'_{0s}\dot{S}_{0r} + \dot{S}'_{0r}\dot{S}_{0s}\right)^{1/2}\right\|_2^2\left\|S_0^{-1}\right\|_F^2.\end{aligned}$$

Under Assumption A3,

$$\left\|\left(\dot{S}'_{0s}\dot{S}_{0r} + \dot{S}'_{0r}\dot{S}_{0s}\right)^{1/2}\right\|_2^2 = \sup_{|x|=1} x'\left(\dot{S}'_{0s}\dot{S}_{0r} + \dot{S}'_{0r}\dot{S}_{0s}\right)x \leq \sup_t\left|\dot{\beta}_{0t,r}\dot{\beta}_{0t,s}\right| \leq K$$

by similar reasoning to (45). On the other hand,

$$\begin{aligned}E_{\theta_0}\left(2\varepsilon' S_0^{-1\prime}\dot{S}'_{0s}\dot{S}_{0r}S_0^{-1}\varepsilon\right) &\geq K_L tr\left(S_0^{-1\prime}S_0^{-1}\right)\inf_t\left(\dot{\beta}_{0t,r}\dot{\beta}_{0t,s}\right) \\ &= K_L\left\|S_0^{-1}\right\|_F^2\inf_t\left(\dot{\beta}_{0t,r}\dot{\beta}_{0t,s}\right). \quad (47)\end{aligned}$$

Under Assumption A3, $\left|\inf_t\left(\dot{\beta}_{0t,r}\dot{\beta}_{0t,s}\right)\right| < \infty$ and we conclude that

$$-\frac{\varepsilon' S_0^{-1\prime}\dot{S}'_{0s}\dot{S}_{0r}S_0^{-1}\varepsilon}{\sigma_0^2\left\|S_0^{-1}\right\|_F^2} = O_{p,e}(1), \quad (48)$$

establishing part (1) of the theorem.

Case 2: $\mu_0 \neq 0$. Showing parts (1) and (2) of the Theorem is very similar to the case $\mu_0 = 0$. For $2 \leq r, s \leq m+1$,

$$\begin{aligned} H^c_{nr,s}(\theta_0) &= -\frac{y'\left(\ddot{S}'_{0r,s}MS_0 + \dot{S}'_{0r}M\dot{S}_{0s} + \dot{S}'_{0s}M\dot{S}_{0r} + S'_0M\ddot{S}_{0r,s}\right)y}{2\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1} \\ &= -\frac{1}{2\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}\{\varepsilon'\left(S_0^{-1\prime}\ddot{S}'_{0r,s}M + M\ddot{S}_{0r,s}S_0^{-1}\right)\varepsilon \qquad (49) \\ &\quad +2\mu_0 1'S_0^{-1\prime}\ddot{S}'_{0r,s}M\varepsilon + 2\varepsilon'S_0^{-1\prime}\dot{S}'_{0s}M\dot{S}_{0r}S_0^{-1}\varepsilon \\ &\quad +4\mu_0 1'S_0^{-1\prime}\dot{S}'_{0s}M\dot{S}_{0r}S_0^{-1}\varepsilon + 2\mu_0^2 1'S_0^{-1\prime}\dot{S}'_{0s}M\dot{S}_{0r}S_0^{-1}1\}. \end{aligned}$$

The first term on the rhs of (49) has

$$\begin{aligned} \left|E_{\theta_0}\left(\varepsilon'\left(S_0^{-1\prime}\ddot{S}'_{0r,s}M + M\ddot{S}_{0r,s}S_0^{-1}\right)\varepsilon\right)\right| &= 2\sigma_0^2\left|tr\left(P\ddot{S}_{0r,s}S_0^{-1}\right)\right| \\ &\leq 2\sigma_0^2\left\|P\right\|_F\left\|\ddot{S}_{0r,s}S_0^{-1}\right\|_F \\ &\leq 2\sigma_0^2\left\|\ddot{S}_{0r,s}\right\|_2\left\|S_0^{-1}\right\|_F \\ &\leq K\left\|S_0^{-1}\right\|_F, \end{aligned}$$

because of (45). With similar calculations, we see that

$$Var_{\theta_0}\left(\varepsilon'\left(S_0^{-1\prime}\ddot{S}'_{0r,s}M + M\ddot{S}_{0r,s}S_0^{-1}\right)\varepsilon\right) \leq K\left\|S_0^{-1}\right\|_F^2$$

and therefore

$$\frac{\varepsilon'\left(S_0^{-1\prime}\ddot{S}'_{0r,s}M + M\ddot{S}_{0r,s}S_0^{-1}\right)\varepsilon}{\left\|S_0^{-1\prime}S_0^{-1}\right\|_1} = O_p\left(\rho_n\left\|S_0^{-1}\right\|_F^{-1}\right),$$

which is asymptotically negligible.

The term $2\mu_0 1'S_0^{-1\prime}\ddot{S}'_{0r,s}M\varepsilon$ in (49) has zero expectation and

$$\begin{aligned} Var_{\theta_0}\left(2\mu_0 1'S_0^{-1\prime}\ddot{S}'_{0r,s}M\varepsilon\right) &= 4\mu_0^2\sigma_0^2 1'S_0^{-1\prime}\ddot{S}'_{0r,s}M\ddot{S}_{0r,s}S_0 1 \\ &\leq 4\mu_0^2\sigma_0^2\left\|S_0^{-1\prime}S_0^{-1}\right\|_1\left\|\ddot{S}'_{0r,s}M\ddot{S}_{0r,s}\right\|_2 \\ &\leq K\left\|S_0^{-1\prime}S_0^{-1}\right\|_1, \end{aligned}$$

implying that

$$\frac{2\mu_0 1' S_0^{-1\prime} \ddot{S}_{0r,s}' M \varepsilon}{\left\| S_0^{-1\prime} S_0^{-1} \right\|_1} = O_p\left(\left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{-1/2} \right),$$

which is also asymptotically negligible.

For the third term on the rhs of (49),

$$\varepsilon' S_0^{-1\prime} \dot{S}_{0s}' M \dot{S}_{0r} S_0^{-1} \varepsilon = \varepsilon' S_0^{-1\prime} \dot{S}_{0s}' \dot{S}_{0r} S_0^{-1} \varepsilon - \varepsilon' S_0^{-1\prime} \dot{S}_{0s}' P \dot{S}_{0r} S_0^{-1} \varepsilon. \quad (50)$$

By (48), the first term on the rhs of (50) is $O_{p,e}\left(\left\| S_0^{-1} \right\|_F^2 \right)$ and for the second term, we have

$$\begin{aligned} \left| E_{\theta_0}\left(\varepsilon' S_0^{-1\prime} \dot{S}_{0s}' P \dot{S}_{0r} S_0^{-1} \varepsilon \right) \right| &= \frac{\sigma_0^2}{n} \left| 1' \dot{S}_{0r} S_0^{-1} S_0^{-1\prime} \dot{S}_{0s}' 1 \right| \quad (51) \\ &\leq \frac{K}{n} 1' \left(S_0^{-1} - I_n \right) \left(S_0^{-1} - I_n \right)' 1 \\ &\leq \frac{K}{n} \left(1' S_0^{-1} S_0^{-1\prime} 1 + 2 1' S_0^{-1} 1 + n \right). \end{aligned}$$

Now,

$$1' S_0^{-1} S_0^{-1\prime} 1 \leq n \left\| S_0^{-1} S_0^{-1\prime} \right\|_2 \leq n \left\| S_0^{-1} \right\|_2^2 \leq n \left\| S_0^{-1} \right\|_F^2$$

and $1' S_0^{-1} 1 \leq \sqrt{n} \left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{1/2}$. Therefore, (51) is less than or equal to

$$K \left(\left\| S_0^{-1} \right\|_F^2 + \frac{2}{\sqrt{n}} \left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{1/2} + 1 \right). \quad (52)$$

With a $\left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{-1}$-normalization, the dominant term in (52) is bounded by $K\rho_n$. Similar calculations reveal the variance to be of $\varepsilon' S_0^{-1\prime} \dot{S}_{0s}' M \dot{S}_{0r} S_0^{-1} \varepsilon$ to be of the order $O\left(\left\| S_0^{-1} \right\|_F^4 \right)$. Together with the exact order of the first term on the rhs of (50), it follows that

$$-\frac{\varepsilon' S_0^{-1\prime} \dot{S}_{0s}' M \dot{S}_{0r} S_0^{-1} \varepsilon}{\sigma_0^2 \left\| S_0^{-1\prime} S_0^{-1} \right\|_1} = O_{p,e}\left(\rho_n \right).$$

The fourth term on the rhs of (49) equals

$$4\mu_0 1' S_0^{-1\prime} \dot{S}_{0s}' \left(I_n - P\right) \dot{S}_{0r} S_0^{-1} \varepsilon, \tag{53}$$

which has zero expectation. Also,

$$\begin{aligned} Var_{\theta_0}\left(4\mu_0 1' S_0^{-1\prime} \dot{S}_{0s}' \dot{S}_{0r} S_0^{-1} \varepsilon\right) &= 16\mu_0^2 \sigma_0^2 1' S_0^{-1\prime} \dot{S}_{0s}' \dot{S}_{0r} S_0^{-1} S_0^{-1\prime} \dot{S}_{0r}' \dot{S}_{0s} S_0^{-1} 1 \\ &\leq K 1' \left(S_0^{-1\prime} S_0^{-1}\right)^2 1 \\ &\leq \left\| S_0^{-1\prime} S_0^{-1} \right\|_1 \left\| S_0^{-1} S_0^{-1\prime} \right\|_2 \\ &\leq \left\| S_0^{-1\prime} S_0^{-1} \right\|_1 \left\| S_0^{-1} \right\|_F^2 . \end{aligned} \tag{54}$$

Under Assumption A3, it is also true that

$$Var_{\theta_0}\left(4\mu_0 1' S_0^{-1\prime} \dot{S}_{0s}' \dot{S}_{0r} S_0^{-1} \varepsilon\right) \geq K_L 1' \left(S_0^{-1\prime} S_0^{-1}\right)^2 1.$$

With tedious albeit straightforward calculations we obtain

$$\frac{\left(1' \left(S_0^{-1\prime} S_0^{-1}\right)^2 1\right)^{1/2}}{\left\| S_0^{-1\prime} S_0^{-1} \right\|_1} = O_e\left(1\right), \text{ if } \beta_t = \beta > 1, \forall t$$

and

$$\frac{\left(1' \left(S_0^{-1\prime} S_0^{-1}\right)^2 1\right)^{1/2}}{\left\| S_0^{-1\prime} S_0^{-1} \right\|_1} = o\left(1\right), \text{ if } 0 \leq \beta_t = \beta < 1, \forall t.$$

In general though, the bound in (54) implies that

$$\frac{4\mu_0 1' S_0^{-1\prime} \dot{S}_{0s}' \dot{S}_{0r} S_0^{-1} \varepsilon}{\left\| S_0^{-1\prime} S_0^{-1} \right\|_1} = O_p\left(\sqrt{\rho_n}\right).$$

The second term in (53) has zero expectation and

$$\begin{aligned} Var_{\theta_0}\left(4\mu_0 1' S_0^{-1\prime} \dot{S}_{0s}' P \dot{S}_{0r} S_0^{-1} \varepsilon\right) &= 16\mu_0^2 \sigma_0^2 1' S_0^{-1\prime} \dot{S}_{0s}' P \dot{S}_{0r} S_0^{-1} S_0^{-1\prime} \dot{S}_{0r}' P \dot{S}_{0s} S_0^{-1} 1 \\ &\leq K \left\| S_0^{-1\prime} S_0^{-1} \right\|_1 \left\| \dot{S}_{0s}' P \dot{S}_{0r} S_0^{-1} S_0^{-1\prime} \dot{S}_{0r}' P \dot{S}_{0s} \right\|_2 \\ &\leq \left\| S_0^{-1\prime} S_0^{-1} \right\|_1 \left\| S_0^{-1} \right\|_F^2 . \end{aligned}$$

We recall that in both of the cases $\beta_t = \beta > 1$, $\forall t$ and $0 \leq \beta_t = \beta < 1$, $\forall t$ $\rho_n = O_e(1)$. The implication is that

$$-\frac{2\mu_0 1' S_0^{-1\prime} \dot{S}_{0s}' M \dot{S}_{0r} S_0^{-1} \varepsilon}{\sigma_0^2 \left\| S_0^{-1\prime} S_0^{-1} \right\|_1} = O_p\left(\sqrt{\rho_n}\right) \tag{55}$$

and it is emphasized that the bound is an upper one and not an exact one. Specifically, in the fixed coefficient explosive case the bound is also exact whereas in the fixed coefficient stationary case the upper bound is not exact and in fact it holds that in this case (55) is $o_p(1)$.

The last term in (49) was shown to be $O_e(1)$ in Lieberman (2012).■

Proof of Lemma 4:

Case 1: $\mu_0 = 0$. The terms $[A_n]_{1,1}$ and $[A_n]_{1,r}$, $r = 2, ..., m+1$, are given in (37) and in the proof of Theorem 3, respectively. The term $[A_n]_{2\leq r,s\leq m+1}$ follows from (11) and (38) and because the diagonal of Λ_{0r} is equal to zero.

Case 2: $\mu_0 \neq 0$. The treatment of the terms $[A_n^c]_{1,1}$ and to $[A_n^c]_{1,r}$ is similar to the previous case, the additional $o(1)$ arising from the fact that $tr(M) = n - 1$. For $r, s = 2, ..., m+1$, we use the facts:

$$Cov_{\theta_0}\left(QF_{nr}^c, QF_{ns}^c\right) = 2\sigma_0^4 tr\left(\Gamma_{0r}\Gamma_{0s}\right) + \kappa_4 \sum_{i=1}^{n} \left[\Gamma_{0r}\right]_{i,i} \left[\Gamma_{0s}\right]_{i,i}$$

$$Cov_{\theta_0}\left(LF_{nr}^c, LF_{ns}^c\right) = \sigma_0^2 b_{0r}' b_{0s},$$

and

$$Cov_{\theta_0}\left(QF_{nr}^c, LF_{ns}^c\right) = \kappa_3 \sum_{i=1}^{n} \left[\Gamma_{0r}\right]_{i,i} \left[b_{0s}\right]_i .$$

Moreover, $E_{\theta_0}\left(LF_{nr}^c\right) = 0$ and

$$\left|E_{\theta_0}\left(QF_{nr}^c\right)\right| = 2\sigma_0^2 \left|tr\left(P\dot{S}_{0r}S_0^{-1}\right)\right| \leq \frac{K}{n} 1' \left(S_0^{-1} - I_n\right) 1.$$

But $1' S_0^{-1} 1 \leq \sqrt{n} \left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{1/2}$, implying that

$$\frac{\left|E_{\theta_0}\left(QF_{nr}^c\right)\right|}{\left\| S_0^{-1\prime} S_0^{-1} \right\|_1^{1/2}} \leq \frac{K}{\sqrt{n}}$$

and therefore,

$$\frac{1}{\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}Cov_{\theta_0}\left(QF_{nr}^c, QF_{ns}^c\right) = \frac{1}{\left\|S_0^{-1\prime}S_0^{-1}\right\|_1}E_{\theta_0}\left(QF_{nr}^c QF_{ns}^c\right) + O\left(\frac{1}{n}\right).$$

The proof is completed on the observation of (17) and (18).■

Lemma 7 *Under Assumptions A0-A4, for all* $j,k,l = 1,...,m+1$ *and uniformly in* Θ*,*

$$D_n \frac{\partial l_n(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l}D_n = O_p(1) \text{ and } D_n^c \frac{\partial l_n^c(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l}D_n^c = O_p(1).$$

Proof of Lemma 7: It will be sufficient to consider derivatives wrt the θ_2 components. For the case $\mu_0 = 0$, for $j,k,l = 2,...,m+1$,

$$\begin{aligned}\frac{\partial l_n(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} &= -\frac{1}{2\sigma^2}y'(\dddot{S}'_{j,k,l}S + \ddot{S}'_{j,k}\dot{S}_l + \ddot{S}'_{j,l}\dot{S}_k + \dot{S}'_j\ddot{S}_{k,l} \\ &\quad +\ddot{S}'_{k,l}\dot{S}_j + \dot{S}'_k\ddot{S}_{j,l} + \dot{S}'_l\ddot{S}_{j,k} + S'\dddot{S}_{j,k,l})y\end{aligned} \tag{56}$$

and therefore, we need to show that for $p = 1, 2$,

$$\begin{aligned}&|tr(S_0^{-1\prime}(\dddot{S}'_{j,k,l}S + \ddot{S}'_{j,k}\dot{S}_l + \ddot{S}'_{j,l}\dot{S}_k + \dot{S}'_j\ddot{S}_{k,l} \\ &+\ddot{S}'_{k,l}\dot{S}_j + \dot{S}'_k\ddot{S}_{j,l} + \dot{S}'_l\ddot{S}_{j,k} + S'\dddot{S}_{j,k,l})S_0^{-1})^p| \\ \leq\;& K\left\|S_0^{-1}\right\|^p,\end{aligned} \tag{57}$$

uniformly in Θ. Denote by S^* be a derivative of S wrt any θ_2 component. We observe that the terms in (57) are either of the form

$$\left|tr\left(\left(S_0^{-1\prime}\left(S^{*\prime}S^*\right)S_0^{-1}\right)^p\right)\right|$$

or

$$\left|tr\left(\left(S_0^{-1\prime}\left(S^{*\prime}S + S'S^*\right)S_0^{-1}\right)^p\right)\right|.$$

Under Assumptions A3-A4, both terms are uniformly bounded by $K\left\|S_0^{-1}\right\|^p$ by very similar arguments used in the proofs of Theorems 2 and 3. For the $\mu_0 \neq 0$ case, instead of replacing y by $S_0^{-1}\varepsilon$ in (56), we replace it by $\mu_0 S_0^{-1}1 + S_0^{-1}\varepsilon$ and again use similar reasoning to obtain a uniform $O_p\left(\left\|S_0^{-1\prime}S_0^{-1}\right\|_1\right)$ bound.■

Proof of Theorem 5: We shall deal with the $\mu_0 \neq 0$ case only - the complementary case is very similar. For part (2), we write

$$\text{vech}\left(H_n\left(\bar{\theta}_n\right)\right) = \text{vech}\left(H_n\left(\theta_0\right)\right) + \frac{\partial \text{vech}\left(H_n\left(\tilde{\theta}_n\right)\right)}{\partial \theta'}\left(\bar{\theta}_n - \theta_0\right),$$

where $\left|\left|\tilde{\theta}_n - \theta_0\right|\right| \leq \left|\left|\bar{\theta}_n - \theta_0\right|\right|$ and vech$(\cdot)$ is the operator which vectorizes the lower half, including the main diagonal, of a symmetric matrix. By virtue of Theorem 3 vech$\left(H_n\left(\theta_0\right)\right) = O_p(1)$. By Theorem 1, the mean value $\tilde{\theta}_n$ satisfies $\tilde{\theta}_n - \theta_0 = o_p(1)$. Under Assumption A4, with very similar calculations to the proof of Theorem 3, which we omit for brevity, $\partial \text{vech}\left(H_n\left(\tilde{\theta}_n\right)\right)/\partial\theta'$ is uniformly bounded. Therefore,

$$\text{vech}\left(H_n\left(\bar{\theta}_n\right)\right) = \text{vech}\left(H_n\left(\theta_0\right)\right) + o_p(1).$$

The Theorem is established by an application of Lemma 2.4(a) of Hayashi (2000) and using Lemma 7.■

Proof of Theorem 6: The model is $S_0 y = \mu_0 1 + \varepsilon$ and therefore

$$\mu_0 = \frac{1' S_0 y}{n} - \bar{\varepsilon}_n,$$

where $\bar{\varepsilon}_n = \sum_{t=1}^{n} \varepsilon_t$. Hence,

$$\begin{aligned}
\hat{\mu}_n - \mu_0 &= \frac{1'\left(S\left(\hat{\theta}_n\right) - S\left(\theta_0\right)\right)y}{n} + \bar{\varepsilon}_n \\
&= \sum_{r=2}^{m+1}\left(\hat{\theta}_{nr} - \theta_{0r}\right)\frac{1'\dot{S}_r\left(\bar{\theta}_n\right)y}{n} + \bar{\varepsilon}_n,
\end{aligned}$$

where $\bar{\theta}_n$ satisfies $\left\|\bar{\theta}_n - \hat{\theta}_n\right\| \leq \left\|\hat{\theta}_n - \theta_0\right\|$. Under Assumptions A2-A3,

$$\begin{aligned}
\left|1'\dot{S}_r\left(\bar{\theta}_n\right)y\right| &= \left|1'\dot{S}_r\left(\bar{\theta}_n\right)\left(\mu_0 S_0^{-1}1 + S_0^{-1}\varepsilon\right)\right| \\
&\leq \sqrt{1'\dot{C}_r\left(\bar{\theta}_n\right)'\dot{C}_r\left(\bar{\theta}_n\right)1}\left(\mu_0\sqrt{1'S_0^{-1\prime}S_0^{-1}1} + \sqrt{\varepsilon' S_0^{-1\prime}S_0^{-1}\varepsilon}\right) \\
&\leq K\sqrt{n}\left(\mu_0\left\|S_0^{-1\prime}S_0^{-1}\right\|_1^{1/2} + O_p\left(\left\|S_0^{-1}\right\|_F\right)\right).
\end{aligned}$$

It follows that,

$$\left|\hat{\mu}_n - \mu_0\right| \leq \frac{K}{\sqrt{n}} \left(\mu_0 \left\|S_0^{-1} S_0^{-1}\right\|_1^{1/2} + O_p\left(\left\|S_0^{-1}\right\|_F\right)\right) \sum_{r=2}^{m+1} \left|\hat{\theta}_{nr} - \theta_{0r}\right| + \left|\bar{\varepsilon}_n\right|.$$

By (19), Theorem 2 and Theorem 3, $D_n^{-1}\left(\hat{\theta}_n - \theta_0\right) = O_{p,e}(1)$, if $\mu_0 = 0$ and $\left(D_n^c\right)^{-1}\left(\hat{\theta}_n - \theta_0\right) = O_{p,e}(1)$, if $\mu_0 \neq 0$. This implies that

$$\begin{aligned}
\left|\hat{\mu}_n - \mu_0\right| &\leq \frac{K}{\sqrt{n}} + \left|\bar{\varepsilon}_n\right|, \text{ if } \mu_0 = 0, \\
\left|\hat{\mu}_n - \mu_0\right| &\leq \frac{K}{\sqrt{n}}\left(\mu_0 + \rho_n^{1/2}\right) + \left|\bar{\varepsilon}_n\right|, \text{ if } \mu_0 \neq 0
\end{aligned}$$

and the proof is completed.■

Table 1. Normalization factors for the consistency proof.

	$\mu_0 = 0$	$\mu_0 \neq 0$
	Normalization	Normalization
$\rho_n = O_e(1)$	$\left\Vert S_0^{-1}\right\Vert_F^{-2}$ or $\left\Vert S_0^{-1\prime} S_0^{-1}\right\Vert_1^{-1}$	$\left\Vert S_0^{-1\prime} S_0^{-1}\right\Vert_1^{-1}$
$\rho_n = o(1)$	$\left\Vert S_0^{-1}\right\Vert_F^{-2}$	$\left\Vert S_0^{-1\prime} S_0^{-1}\right\Vert_1^{-1}$

Table 2. Normalization factors and dominant terms for the score.

	$\mu_0 = 0$	$\mu_0 \neq 0$	
	Normalization	Normalization	Dominant Term
$\rho_n = O_e(1)$	$\left\Vert S_0^{-1}\right\Vert_F^{-1}$	$\Vert S^{-1\prime} S^{-1}\Vert_1^{-1/2}$	$QF_n + LF_n$
$\rho_n = o(1)$	$\left\Vert S_0^{-1}\right\Vert_F^{-1}$	$\Vert S^{-1\prime} S^{-1}\Vert_1^{-1/2}$	LF_n

Table 3. Performance of the Model Estimators

$\mu = 0$, $w = 0.07$

	$\hat{w}$		$\hat{\mu}$		$\hat{\beta}_t$	
n	Mean	SD	Mean	SD	Mean	SD
250	.0696	.0164	−.0010	.0624	1.0000	.0027
500	.0702	.0082	.0001	.0448	1.0000	.0018
1000	.0699	.0042	−.0006	.0318	1.0000	.0013

$\mu = 0$, $w = 0.2$

	$\hat{w}$		$\hat{\mu}$		$\hat{\beta}_t$	
n	Mean	SD	Mean	SD	Mean	SD
250	.1998	.0170	.0023	.0632	.9996	.0074
500	.1998	.0093	.0007	.0448	.9997	.0052
1000	.2001	.0048	.0000	.0309	.9999	.0037

$\mu = 0.25$, $w = 0.07$

	$\hat{w}$		$\hat{\mu}$		$\hat{\beta}_t$	
n	Mean	SD	Mean	SD	Mean	SD
250	.0701	.0038	.2500	.0626	1.0000	.0026
500	.0700	.0013	.2496	.0450	1.0000	.0018
1000	.0700	.0005	.2501	.0326	1.0000	.0013

$\mu = 0.25$, $w = 0.2$

	$\hat{w}$		$\hat{\mu}$		$\hat{\beta}_t$	
n	Mean	SD	Mean	SD	Mean	SD
250	.1998	.0053	.2513	.0630	.9999	.0074
500	.2000	.0021	.2523	.0447	1.0001	.0052
1000	.2000	.0010	.2501	.0316	1.0001	.0036

$Z_t \sim U[-1, 1] + b_w$, where $b_w = -0.0116648$ if $w = 0.07$ and $b_w = -0.033289$ if $w = 0.2$.

Table 4. Performance of the Model's Estimators

$\mu = 0$, $w = 0.07$

	$\hat{w}$		$\hat{\mu}$		$\hat{\beta}_t$	
n	Mean	SD	Mean	SD	Mean	SD
250	.0700	.0094	−.0010	.0630	.9999	.0044
500	.0698	.0050	−.0009	.0457	.9999	.0031
1000	.0700	.0023	−.0009	.0324	1.0000	.0022

$\mu = 0$, $w = 0.2$

	$\hat{w}$		$\hat{\mu}$		$\hat{\beta}_t$	
n	Mean	SD	Mean	SD	Mean	SD
250	.1996	.0106	.0026	.0629	.9998	.0128
500	.1997	.0052	−.0004	.0454	.9998	.0091
1000	.1999	.0027	−.0002	.0321	1.0000	.0064

$\mu = 0.25$, $w = 0.07$

	$\hat{w}$		$\hat{\mu}$		$\hat{\beta}_t$	
n	Mean	SD	Mean	SD	Mean	SD
250	.0700	.0026	.2485	.0638	1.0000	.0045
500	.0700	.0010	.2507	.0460	.9999	.0032
1000	.0700	.0004	.2494	.0315	1.0000	.0022

$\mu = 0.25$, $w = 0.2$

	$\hat{w}$		$\hat{\mu}$		$\hat{\beta}_t$	
n	Mean	SD	Mean	SD	Mean	SD
250	.2000	.0043	.2514	.0636	.9999	.0128
500	.1999	.0019	.2501	.0436	.9998	.0093
1000	.1999	.0009	.2499	.0322	1.0001	.0063

$Z_t \sim NID\,(\mu_Z, 1)$ with $\mu_Z = -w/2$.

Table 5. p-values of the ADF tests for the ETF data

Ticker	**EWA** (Australia)	**EWH** (Hong Kong)	**EWJ** (Japan)	**EWM** (Malaysia)
Intercept	0.1800	0.7910	0.6403	0.7002
None	0.5913	0.6190	0.6109	0.8491
Ticker	**EWS** (Singapore)	**EWT** (Taiwan)	**EWY** (South Korea)	**FXI** (China)
Intercept	0.6105	0.6142	0.0675	0.3243
None	0.8078	0.6400	0.8034	0.8596

Note: Intercept-the ADF test with an intercept; None-the ADF test without a trend or an intercept.

Table 6. Similarity-based model estimation of the ETF data

Ticker		$\hat{\mu}$	$\hat{w}_1$	$\hat{w}_2$	$\hat{w}_3$	$\hat{w}_4$	$\hat{w}_5$	$\hat{w}_6$	**AIC**
EWA	M1	0.001	1.000						-7.625
(Australia)		(0.0004)	(0.0001)						
	M2		0.269	0.001	0.283	0.017	-0.003	-0.251	-9.903
			(0.0033)	(0.0036)	(0.0018)	(0.0032)	(0.0004)	(0.0019)	
EWH	M1	0.000	1.000						-8.010
(Hong Kong)		(0.003)	(0.0001)						
	M2		0.040	0.244	0.374	-0.053	-0.003	-0.241	-9.668
			(0.0026)	(0.0026)	(0.0019)	(0.0035)	(0.0004)	(0.0018)	
EWJ	M1	0.000	1.000						-8.779
(Japan)		(0.0002)	(0.0000)						
	M2		0.213	0.072	0.336	0.054	-0.002	-0.237	-10.180
			(0.0037)	(0.0040)	(0.0017)	(0.0028)	(0.0003)	(0.0016)	
EWM	M1	0.000	1.000						-8.144
(Malaysia)		(0.0003)	(0.0001)						
	M2		0.031	0.329	0.339	-0.082	-0.008	-0.236	-9.196
			(0.0109)	(0.0114)	(0.0032)	(0.0055)	(0.0007)	(0.0035)	
EWS	M1	0.001	1.000						-7.583
(Singapore)		(0.0005)	(0.0002)						
	M2		0.104	0.299	0.424	-0.165	-0.007	-0.348	-9.575
			(0.0141)	(0.0144)	(0.0030)	(0.0059)	(0.0008)	(0.0035)	
EWT	M1	0.000	1.000						-7.958
(Taiwan)		(0.0003)	(0.0001)						
	M2		0.0258	0.275	0.399	0.042	-0.004	-0.204	-9.335
			(0.0082)	(0.0084)	(0.0025)	(0.0044)	(0.0004)	(0.0020)	
EWY	M1	0.001	1.000						-7.773
(South Korea)		(0.0004)	(0.0001)						
	M2		0.004	0.202	0.281	0.008	-0.001	-0.159	-9.370
			(0.0023)	(0.0026)	(0.0016)	(0.0030)	(0.0003)	(0.0013)	
FXI	M1	0.002	1.000						-7.722
(China)		(0.0008)	(0.0002)						
	M2		0.010	0.189	0.332	-0.061	-0.001	-0.166	-9.715
			(0.0170)	(0.0170)	(0.0017)	(0.0034)	(0.0003)	(0.0017)	

Note: M1-the random walk with a drift model; For M1, $\hat{w}_1 \equiv \hat{\beta}$; M2-the similarity model (30); standard errors are in brackets.

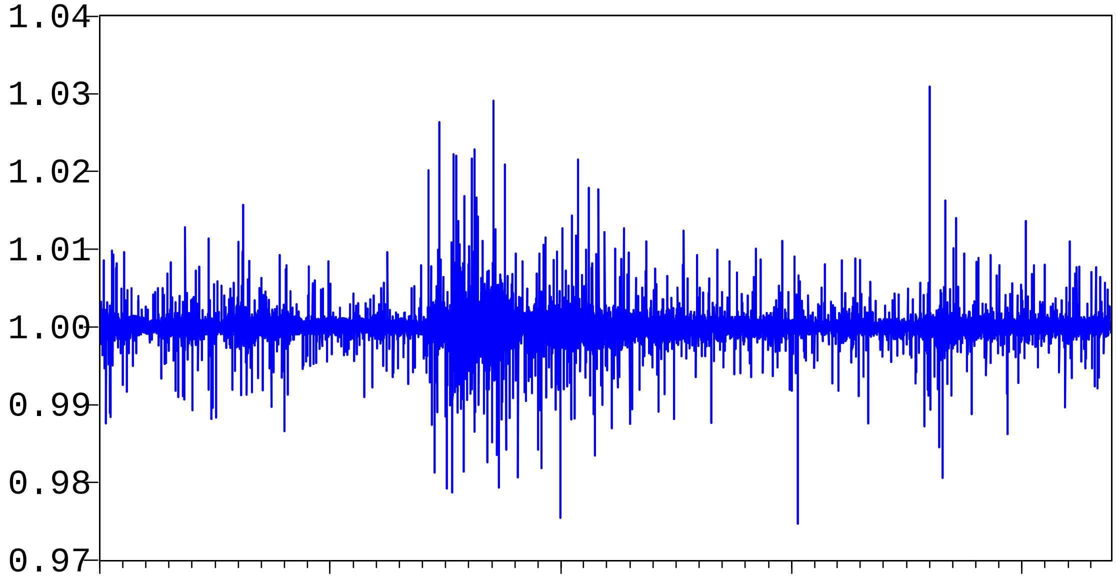

Fig. 1: Recursive values of $\hat{\beta}_t$ based on the fitted version of (30)

References

Chen, R. & R. S. Tsay (1993) Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88, 298–308.

Dahlhaus, R. (2000) A likelihood approximation for locally stationary processes. *The Annals of Statistics* 28, 1762–1794.

Dahlhaus, R., R.H. Neumann & R. Von Sachs (1999) Nonlinear wavelet estimation of time-varying autoregressive process. *Bernoulli* 5, 873–906.

Evans, G.B.A. & N.E. Savin (1984) Testing for Unit Roots: 2. *Econometrica* 52, 1241–1269.

Feigin, P.D. (1976) Maximum likelihood estimation for continuous time stochastic processes. *Advances in Applied Probability* 8, 712–736.

Granger, C.W.J. & N.R. Swanson (1997) Introduction to stochastic unit root processes. *Journal of Econometrics* 80, 35–62.

Graybill, F.A. (1983) *Matrices with Applications in Statistics*, 2nd Edition. Wadsworth, Belmont, California.

Heyde, C.C. (1975) Remarks in efficiency of estimation for branching processes. *Biometrika* 62, 49–55.

Hwang, S. Y. & I. V. Basawa (2005). Explosive random coefficient AR(1) processes and related asymptotics for least squares estimation, *Journal of Time Series Analysis*, 26, 807-824.

Kadane, J. B. (1971). Comparison of k-class estimators when the disturbances are small. *Econometrica*, 39, 723-737

Levy, A. & O. Lieberman (2012) Overreaction of Country ETFs to US Market Returns: Intraday vs. Daily Horizons and the Role of Synchronized Trading. Forthcoming in *Journal of Banking and Finance.*

Lieberman, O. (2010) Asymptotic Theory for Empirical Similarity models. *Econometric Theory* 26, 1032–1059.

Lieberman, O. (2012) A similarity-based approach to time-varying coefficient nonstationary autoregression. *Journal of Time Series Analysis* 33, 484–502.

Lundbergh, S., T. Terăsvirta, & D. van Dijk (2003) Time-varying smooth transition autoregressive models. *Journal of Business and Economic Statistics* 21, 104-121.

McCabe, B. P. and A. R. Tremayne (1995). Testing a time series for difference stationarity. *Annals of Statistics*, 23, 1015-1028.

Nicholls, D.F. & B.G. Quinn (1980) The estimation of random coefficient autoregressive models. I. *Journal of Time Series Analysis* 1, 37– 46.

Nicholls, D.F. and B.G. Quinn (1981) The estimation of random coefficient autoregressive models. II. *Journal of Time Series Analysis* 2, 185–203.

Nicholls, D.F. & B.G. Quinn (1982) *Random Coefficient Autoregressive Models: An Introduction*. Lecture Notes in Statistics 11. Springer-Verlag.

Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Phillips P. C. B., Y. Wu and J. Yu (2011). "Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?", *International Economic Review,* 52, 201-226.

Phillips P. C. B. and J. Yu (2011). "Dating the Timeline of Financial Bubbles during the Subprime Crisis," *Quantitative Economics,* 2, 455-491.

Samuelson, P. A. (197). The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances and Higher Moments, *Review of Economic Studie*s, 37, 537-542.

Tong, H. (2010). Threshold models in time-series analysis–30 years on. *Statistics and Its Interface*, 4, 107-136.

Wu, C.F. (1981) Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics* 9, 501–513.