Extracting a Common Stochastic Trend:
Theories with Some Applications[1]

Yoosoon Chang [2]

J. Isaac Miller [3]

and

Joon Y. Park [4]

Abstract

This paper investigates the statistical properties of the Kalman filter for state space models including integrated time series. In particular, we derive the full asymptotics of maximum likelihood estimation for some prototypical class of such models, i.e., the models with a single latent common stochastic trend. Indeed, we establish the consistency and asymptotic mixed normality of the maximum likelihood estimator and show that the conventional method of inference is valid for this class of models. The models considered explicitly in the paper comprise a special, yet useful, class of models that we may use to extract the common stochastic trend from multiple integrated time series. As we show in the paper, the models can be very useful to obtain indices that represent fluctuations of various markets or common latent factors that affect a set of economic and financial variables simultaneously. Moreover, our derivation of the asymptotics of this class makes it clear that the asymptotic Gaussianity and the validity of the conventional inference for the maximum likelihood procedure extends to a larger class of more general state space models involving integrated time series. Finally, we demonstrate the utility of the state space model by extracting a common stochastic trend in three empirical analyses: interest rates, return volatility and trading volume, and Dow Jones stock prices.

---

[1]This paper was prepared for the presentations at Symposium on Econometric Theory and Applications (SETA) on May 18-20, 2005, at Academia Sinica, Taipei, Taiwan.

[2]Department of Economics, Rice University.

[3]Department of Economics, University of Missouri.

[4]Department of Economics, Rice University and Sungkyunkwan University.

1. Introduction

The Kalman filter is perhaps one of the most widely used modeling tools, not only in econometrics and finance, but also in such diverse fields as artificial intelligence, aeronautical engineering, and many others. Under linear, Gaussian, and stationary assumptions, the asymptotic properties of the filter are well-known, and the technique works quite well. If linearity is violated, then the extended Kalman filter is a standard alternative. If Gaussianity is violated, then maximum likelihood is instead pseudo-maximum likelihood. As long as the distribution of the state (or transition) equation has a finite second moment, then the filter retains some of its optimal properties. See Caines (1988) and Hamilton (1994) for the statistical properties of the filter and the maximum likelihood procedure. The filter seems to generate reasonable parameter estimates even when the distributions have thick tails, as illustrated in Miller and Park (2004).

In this paper, we focus mainly on a violation of stationarity. Many empirical analyses in the literature use nonstationary data or assume a nonstationary unobservable variable or vector. The reader is referred to Kim and Nelson (1999) for an excellent survey and many concrete examples. The properties of the Kalman filter under such assumptions, however, are not well-known. To the best of our knowledge, no formal theory has yet been developed for the Kalman filter applied to models with nonstationary time series. Moving from stationary to nonstationary processes in any model calls into question rates of convergence of the parameter estimates, if not the parameter estimates themselves. Moreover, the asymptotic theory of the maximum likelihood estimates may diverge from the standard Gaussian framework. Consequently, a solid theoretical analysis of the Kalman filter with nonstationary data is needed.

In this analysis, we focus on an important class of nonstationary models, i.e., models that include integrated time series. More precisely, we consider the state space models with a single latent common integrated stochastic trend, and analyze the properties of the Kalman filter to estimate the parameters in these models. For this class of models, we derive the full asymptotics of maximum likelihood estimation, and establish the consistency and Gaussianity of the maximum likelihood estimator. The limit theories for our models differ from the standard asymptotics for stationary models: The convergence rates are a mixture of $\sqrt{n}$ and n, and the limit distributions are generally mixed normals. However, the Gaussian limit distribution theory makes the conventional method of inference valid also for our models with integrated time series. Though our theories are explicitly developed for the simple prototypical models, it is clear that the main results in the paper extend to more general state space models with integrated time series.

The state space models considered in the paper assume that the included time series share one common stochastic trend. This, of course, implies that there are $(m-1)$-number of independent cointegrating relationships, if we set m to be the number of the time series in the models. We show in the paper how our state space models are related to cointegrated models, especially to their error correction representations. We also discuss decompositions into the permanent and transitory components of a given time series. Our state space models suggest a natural choice for such decompositions. However, we may also use the decomposition based on the error correction representations of our models.

We consider three illustrative examples of uses of the Kalman filter to extract a common stochastic trend. We first take on one of the most common applications in the macroeconomics literature: extracting a common trend from short- and long-term interest rates. Subsequently, we explore an application which seems popular in the finance literature. We look at the relationship between stock return volatility and trading volume by extracting a common stochastic trend from those two series. In the third application, we extract the common trend from 30 series of prices of those stocks comprising the Dow Jones Industrial Average.

The rest of the paper is organized as follows. In Section 2, we introduce the state space model and outline the Kalman filtering technique used to estimate the model. We also present some preliminary results that simplify the theoretical analysis and are useful in estimation. We present the main theoretical findings of our analysis in Section 3. Here we analyze the maximum likelihood procedures and obtain their asymptotics. In particular, we show that the maximum likelihood estimators are consistent and asymptotically mixed normal. Section 4 includes some important results on the relationship between our state space models and the usual error correction representation of cointegrated models. We also discuss the permanent and transitory decompositions based on our models here. In Section 5, we present the three empirical applications, and we conclude with Section 6. Mathematical proofs of our theoretical results are contained in an Appendix.

2. The Model and Preliminary Results

We consider the state space model given by

$$
\begin{aligned}
y_t &= \beta_0 x_t + u_t \\
x_t &= x_{t-1} + v_t,
\end{aligned}
\tag{1}
$$

where we assume that:

(a) β_0 is an m-dimensional vector of unknown parameters,
(b) (x_t) is a scalar latent variable,
(c) (y_t) is an m-dimensional observable time series,
(d) (u_t) and (v_t) are m- and 1-dimensional iid errors that are normal with mean zero and variances Λ_0 and 1, respectively, and independent of each other, and
(e) x_0 is independent of (u_t) and (v_t), and assumed to be given.

The variance of (v_t) is set to be unity for the identification of β, which is identified under this condition up to multiplication by -1. The other conditions are standard and routinely imposed in the type of models we consider here.

Our model can be used to extract a common stochastic trend in the time series (y_t). Note that the latent variable (x_t) is defined as a random walk, and may be regarded as a common stochastic trend in (y_t). Here we do not introduce any dynamics in the measurement equation, and mainly consider the simplest prototypical state space model. This is purely for the purpose of exposition. It will be made clear that our subsequent results are directly

applicable for a more general class of state space models, where we have an arbitrary number of lagged differences of the observed time series in the measurement equation.

The model given in (1) may be estimated by the usual Kalman filter. Let $\mathcal{F}_t$ be the σ-field generated by $y_1, \ldots, y_t$, and for $z_t = x_t$ or y_t, we denote by $z_{t|s}$ the conditional expectation of z_t given $\mathcal{F}_s$, and by $\omega_{t|s}$ and $\Sigma_{t|s}$ the conditional variances of x_t and y_t given $\mathcal{F}_s$, respectively. The Kalman filter consists of the prediction and updating steps. For the prediction step, we utilize the relationships

$$
\begin{aligned}
x_{t|t-1} &= x_{t-1|t-1}, \\
y_{t|t-1} &= \beta x_{t|t-1},
\end{aligned}
$$

and

$$
\begin{aligned}
\omega_{t|t-1} &= \omega_{t-1|t-1} + 1, \\
\Sigma_{t|t-1} &= \omega_{t|t-1}\beta\beta' + \Lambda.
\end{aligned}
$$

On the other hand, the updating step relies on the relationships

$$
\begin{aligned}
x_{t|t} &= x_{t|t-1} + \omega_{t|t-1}\beta'\Sigma_{t|t-1}^{-1}(y_t - y_{t|t-1}), \\
\omega_{t|t} &= \omega_{t|t-1} - \omega_{t|t-1}^2\beta'\Sigma_{t|t-1}^{-1}\beta.
\end{aligned}
$$

The reader is referred to, e.g., Hamilton (1994) or Kim and Nelson (1999) for more details.

For any given values of β and Λ, we may show that there exist the steady state values of $\omega_{t|t-1}$ and $\Sigma_{t|t-1}$, which we denote respectively by ω and Σ. We have

Lemma 2.1 The steady state values ω and Σ exist and are given by

$$
\omega = \frac{1 + \sqrt{1 + 4/(\beta'\Lambda^{-1}\beta)}}{2}, \tag{2}
$$

$$
\Sigma = \frac{1 + \sqrt{1 + 4/(\beta'\Lambda^{-1}\beta)}}{2}\beta\beta' + \Lambda \tag{3}
$$

for any m-dimensional vector β and $m \times m$ matrix Λ.

From now on, we set

$$
\omega_{0|0} = \omega - 1 \tag{4}
$$

so that $\omega_{t|t-1} = \omega$ for all $t \geq 1$ and becomes time invariant (and therefore, $\Sigma_{t|t-1}$ too). This causes no loss of generality in our asymptotic analysis, since $(\omega_{t|t-1})$ converges to its asymptotic steady state value ω as t increases. The following lemma specifies $(x_{t|t-1})$ more explicitly as a function of the observed time series (y_t) and the initial value x_0. Here and elsewhere in the paper, we assume (4). To simplify the presentation, we also make the convention $y_0 = 0$.

Lemma 2.2 We have

$$x_{t|t-1} = \frac{\beta'\Lambda^{-1}}{\beta'\Lambda^{-1}\beta} \left[y_t - \sum_{k=0}^{t-1}(1-1/\omega)^k \triangle y_{t-k} \right] + (1-1/\omega)^{t-1} x_0$$

for all $t \geq 2$.

The result in Lemma 2.2 is given entirely by our prediction and updating steps of the Kalman filter. In particular, it holds regardless of misspecification of our model in (1).

In addition to the initial value of $\omega_{t|t-1}$ given by (4), we let

$$x_0 = 0 \tag{5}$$

for the rest of the paper. It is clearly seen from Lemma 2.2 that relaxing this simplifying assumption would not affect our subsequent asymptotic analyses. Note that $\omega > 1$, and therefore, $0 < 1 - 1/\omega < 1$. Subsequently, the magnitude of the term $(1-1/\omega)^{t-1}x_0$ is geometrically declining as $t \to \infty$, as long as x_0 is fixed and finite a.s.

Let ω_0 be the value of ω defined with the true values β_0 and Λ_0 of β and Λ. If we denote by $x_{t|t-1}^0$ the value of $x_{t|t-1}$ under model (1), we may deduce from from Lemma 2.2 that

Proposition 2.3 We have

$$x_{t|t-1}^0 = x_t + (1/\omega_0)\sum_{k=1}^{t-1}(1-1/\omega_0)^{k-1}\frac{\beta_0'\Lambda_0^{-1}u_{t-k}}{\beta_0'\Lambda_0^{-1}\beta_0} - \sum_{k=0}^{t-1}(1-1/\omega_0)^k v_{t-k}$$

for all $t \geq 2$.

Proposition 2.3 implies in particular that we have

$$x_{t|t-1}^0 - x_t = (1/\omega_0)p_{t-1} - q_t, \tag{6}$$

where

$$p_t = \sum_{k=0}^{\infty}(1-1/\omega_0)^k\frac{\beta_0'\Lambda_0^{-1}u_{t-k}}{\beta_0'\Lambda_0^{-1}\beta_0} \quad \text{and} \quad q_t = \sum_{k=0}^{\infty}(1-1/\omega_0)^k v_{t-k}$$

asymptotically. If we let (u_t) and (v_t) be iid random sequences, then the time series (p_t) and (q_t) introduced in (6) become the stationary first-order autoregressive processes given by

$$\begin{aligned} p_t &= (1-1/\omega_0)p_{t-1} + \frac{\beta_0'\Lambda_0^{-1}}{\beta_0'\Lambda_0^{-1}\beta_0}u_t \\ q_t &= (1-1/\omega_0)q_{t-1} + v_t \end{aligned}$$

respectively. As we noted earlier, we have $0 < 1 - 1/\omega_0 < 1$.

Now it is clear that $(x_{t|t-1}^0)$ is cointegrated with (x_t) with unit cointegrating coefficient, i.e., $(x_{t|t-1}^0)$ and (x_t) have a common stochastic trend. The stochastic trend of (x_t) may

therefore be analyzed by that of $(x_{t|t-1}^0)$. It should be emphasized that the result in Proposition 2.3 only assumes our model specification in (1). In particular, it does not rely upon the iid assumption on (u_t) and (v_t). Our result shows that the Kalman filter extracts the common stochastic trend of (y_t) as long as (u_t) and (v_t) are general stationary processes, i.e., as long as (y_t) is a vector of integrated processes. Unlike (x_t), however, $(x_{t|t-1}^0)$ is not a pure random walk, even if (u_t) and (v_t) are iid random sequences. The latter deviates from the former up to the stationary error that is autoregressive of order one. Our results here assume that the true parameters of the model are known. Of course, the true parameter values are unknown and have to be estimated in most practical applications. Nevertheless, it is rather clear that our conclusions here continue to be valid as long as we use the consistent parameter estimates.

The result in Proposition 2.3 is valid also for the state space model with the measurement equation given by

$$y_t = \beta_0 x_t + \sum_{k=1}^{p} \Phi_k \triangle y_{t-k} + u_t \tag{7}$$

instead of the one in (1). Under this specification, the Kalman filter has exactly the same prediction and updating steps except for

$$y_{t|t-1} = \beta x_{t|t-1} + \sum_{k=1}^{p} \Phi_k \triangle y_{t-k},$$

replacing $y_{t|t-1} = \beta x_{t|t-1}$. Consequently, Lemma 2.1 continues to hold without modification. Moreover, we may easily obtain the result corresponding to Lemma 2.2 by substituting (y_t) by $(y_t - \sum_{k=1}^{p} \Phi_k \triangle y_{t-k})$, and as a result, Proposition 2.3 holds as it is. This can be seen clearly from the proofs of Lemma 2.2 and Proposition 2.3.

3. Asymptotics for Maximum Likelihood Estimation

In this section, we consider the maximum likelihood estimation of our model. In particular, we establish the consistency and asymptotic Gaussianity of the maximum likelihood estimator of the parameter under normality. Note that our model includes an integrated process, and therefore, the usual asymptotic theory for the maximum likelihood estimation for the state space model given by, for instance, Caines (1988), does not apply. We develop our asymptotic theory in a much more general setting than the one given by our model (1) that is explicitly considered in the paper. As will be seen clearly in what follows, our general theory established here would be very useful to obtain the asymptotics for the maximum likelihood estimation in a variety of models including integrated time series, both latent and observed. Below we first develop the general asymptotic theories of maximum likelihood estimation which allow for the presence of nonstationary time series, and then apply them to obtain the asymptotics of the maximum likelihood estimator in our model (1).

We let θ be a κ-dimensional parameter, and define

$$\varepsilon_t = y_t - y_{t|t-1} \tag{8}$$

to be the prediction error with conditional mean zero and covariance matrix Σ. Under normality, the log-likelihood function of $y_1, \ldots, y_n$ is given by

$$\ell_n(\theta) = -\frac{n}{2} \log \det \Sigma - \frac{1}{2} \operatorname{tr} \Sigma^{-1} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t', \tag{9}$$

ignoring the unimportant constant term. Note that Σ and (ε_t) are in general given as functions of θ. If we denote by $s_n(\theta)$ and $H_n(\theta)$ the score vector and Hessian matrix, i.e.,

$$s_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial \theta} \quad \text{and} \quad H_n(\theta) = \frac{\partial^2 \ell_n(\theta)}{\partial \theta \partial \theta'},$$

then it follows directly from (9) that

Lemma 3.1 For the log-likelihood function given in (9), we have

$$s_n(\theta) = -\frac{n}{2} \frac{\partial (vec \, \Sigma)'}{\partial \theta} vec \left(\Sigma^{-1} \right) + \frac{1}{2} \frac{\partial (vec \, \Sigma)'}{\partial \theta} vec \left(\Sigma^{-1} \sum_{t=1}^{n} \varepsilon_t \varepsilon_t' \Sigma^{-1} \right) - \sum_{t=1}^{n} \frac{\partial \varepsilon_t'}{\partial \theta} \Sigma^{-1} \varepsilon_t,$$

and

$$
\begin{aligned}
H_n(\theta) = & -\frac{n}{2} \left[I_\kappa \otimes \left(vec \, \Sigma^{-1} \right)' \right] \left[\frac{\partial^2}{\partial \theta \partial \theta'} \otimes (vec \, \Sigma) \right] \\
& + \frac{1}{2} \left[I_\kappa \otimes \left(vec \, \Sigma^{-1} \left(\sum_{t=1}^{n} \varepsilon_t \varepsilon_t' \right) \Sigma^{-1} \right)' \right] \left[\frac{\partial^2}{\partial \theta \partial \theta'} \otimes (vec \, \Sigma) \right] \\
& + \frac{n}{2} \frac{\partial (vec \, \Sigma)'}{\partial \theta} \left(\Sigma^{-1} \otimes \Sigma^{-1} \right) \frac{\partial (vec \, \Sigma)}{\partial \theta'} \\
& - \frac{1}{2} \frac{\partial (vec \, \Sigma)'}{\partial \theta} \left[\Sigma^{-1} \otimes \Sigma^{-1} \left(\sum_{t=1}^{n} \varepsilon_t \varepsilon_t' \right) \Sigma^{-1} + \Sigma^{-1} \left(\sum_{t=1}^{n} \varepsilon_t \varepsilon_t' \right) \Sigma^{-1} \otimes \Sigma^{-1} \right] \frac{\partial (vec \, \Sigma)}{\partial \theta'} \\
& - \sum_{t=1}^{n} \frac{\partial \varepsilon_t'}{\partial \theta} \Sigma^{-1} \frac{\partial \varepsilon_t}{\partial \theta'} - \sum_{t=1}^{n} \left(I \otimes \varepsilon_t' \Sigma^{-1} \right) \left(\frac{\partial^2}{\partial \theta \partial \theta'} \otimes \varepsilon_t \right) \\
& + \frac{\partial (vec \, \Sigma)'}{\partial \theta} \left(\Sigma^{-1} \otimes \Sigma^{-1} \right) \sum_{t=1}^{n} \left(\frac{\partial \varepsilon_t}{\partial \theta'} \otimes \varepsilon_t \right) + \sum_{t=1}^{n} \left(\frac{\partial \varepsilon_t'}{\partial \theta} \otimes \varepsilon_t' \right) \left(\Sigma^{-1} \otimes \Sigma^{-1} \right) \frac{\partial (vec \, \Sigma)}{\partial \theta'}.
\end{aligned}
$$

In Lemma 3.1 and elsewhere in the paper, $vec \, A$ denotes the column vector obtained by stacking the rows of matrix A.

Denote by $\hat{\theta}_n$ the maximum likelihood estimator of θ, the true value of which is denoted by θ_0. As in the standard stationary model, the asymptotics of $\hat{\theta}_n$ in our model can be obtained from the first order Taylor expansion of the score vector, which is given by

$$s_n(\hat{\theta}_n) = s_n(\theta_0) + H_n(\theta_n) \left(\hat{\theta}_n - \theta_0 \right), \tag{10}$$

where θ_n lies in the line segment connecting $\hat{\theta}_n$ and θ_0. Of course, we have $s_n(\hat{\theta}_n) = 0$ if $\hat{\theta}_n$ is an interior solution. Therefore, it is now clear from (10) that we may write

$$\nu_n' T^{-1}(\hat{\theta}_n - \theta_0) = -\left[\nu_n^{-1} T' H_n(\theta_n) T \nu_n^{-1\prime}\right]^{-1} \left[\nu_n^{-1} T' s_n(\theta_0)\right] \tag{11}$$

for appropriately defined κ-dimensional square matrices ν_n and T, which are introduced here respectively for the necessary normalization and rotation.

Upon appropriate choice of the normalization matrix sequence ν_n and rotation matrix T, we will show that

ML1: $\nu_n^{-1} T' s_n(\theta_0) \to_d N$ as $n \to \infty$,

ML2: $-\nu_n^{-1} T' H_n(\theta_0) T \nu_n^{-1\prime} \to_d M > 0$ a.s. as $n \to \infty$

for some M and N, and

ML3: *There exists a sequence of invertible normalization matrices μ_n such that $\mu_n \nu_n^{-1} \to 0$ a.s., and such that*

$$\sup_{\theta \in \Theta_0} \left\| \mu_n^{-1} T' \left(H_n(\theta) - H_n(\theta_0) \right) T \mu_n^{-1\prime} \right\| \to_p 0,$$

where $\Theta_n = \left\{ \theta \,\big|\, \left\| \mu_n' T^{-1}(\theta - \theta_0) \right\| \leq 1 \right\}$ is a sequence of shrinking neighborhoods of θ_0,

subsequently below.

As shown by Park and Phillips (2001) in their study of the nonlinear regression with integrated time series, conditions ML1 – ML3 above are sufficient to derive the asymptotics for $\hat{\theta}_n$. In fact, under conditions ML1 – ML3, we may deduce from (11) and continuous mapping theorem that

$$\begin{aligned} \nu_n' T^{-1}(\hat{\theta}_n - \theta_0) &= -\left[\nu_n^{-1} T' H_n(\theta_0) T \nu_n^{-1\prime}\right]^{-1} \left[\nu_n^{-1} T' s_n(\theta_0)\right] + o_p(1) \\ &\to_d M^{-1} N \end{aligned} \tag{12}$$

as $n \to \infty$. In particular, ML3 ensures that $s_n(\hat{\theta}_n) = 0$ with probability approaching to one and

$$\nu_n^{-1} T' \left(H_n(\theta_n) - H_n(\theta_0) \right) T \nu_n^{-1\prime} \to_p 0$$

as $n \to \infty$. This was shown earlier by Wooldridge (1994) for the asymptotic analyses of extremum estimators in models including nonstationary time series.

To derive the asymptotics for $s_n(\theta_0)$, let ε_t^0, $(\partial/\partial\theta')\varepsilon_t^0$ and $(\partial/\partial\theta')vec\,\Sigma_0$ be defined respectively as ε_t, $(\partial/\partial\theta')\varepsilon_t$ and $(\partial/\partial\theta')vec\,\Sigma$ evaluated at the true parameter value θ_0 of θ. We have

$$s_n(\theta_0) = \frac{1}{2} \frac{\partial(vec\,\Sigma_0)'}{\partial\theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) vec \left[\sum_{t=1}^{n} \left(\varepsilon_t^0 \varepsilon_t^{0\prime} - \Sigma_0\right)\right] - \sum_{t=1}^{n} \frac{\partial\varepsilon_t^{0\prime}}{\partial\theta} \Sigma_0^{-1} \varepsilon_t^0,$$

for which we note that

Remarks   We have $\varepsilon_t^0 = y_t - y_{t|t-1}^0 = \beta_0(x_t - x_{t|t-1}^0) + u_t$. Then for $s_n(\theta_0)$ we have

(a) $(\varepsilon_t^0, \mathcal{F}_t)$ is an mds by construction and due to (6) with conditional variance Σ_0, and becomes iid $\mathbb{N}(0, \Sigma_0)$ under the assumption of normality. In particular, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\varepsilon_t^0 \varepsilon_t^{0\prime} - \Sigma_0) \to_d \mathbb{N}\Big(0, (I+K)(\Sigma_0 \otimes \Sigma_0)\Big) \tag{13}$$

as $n \to \infty$, where K is the commutation matrix, and

(b) $(\partial/\partial\theta')\varepsilon_t^0$ is $\mathcal{F}_{t-1}$-measurable, and consequently, $((\partial\varepsilon_t^{0\prime}/\partial\theta)\Sigma_0^{-1}\varepsilon_t^0)$ is an mds and

$$\sum_{t=1}^{n} \left(\varepsilon_t^0 \varepsilon_t^{0\prime} - \Sigma_0\right) \quad \text{and} \quad \sum_{t=1}^{n} \frac{\partial\varepsilon_t^{0\prime}}{\partial\theta} \Sigma_0^{-1} \varepsilon_t^0$$

become independent asymptotically.

For the asymptotic result in (13), see, e.g., Muirhead (1982, pp.90-91).

It is now clear that the asymptotics of $s_n(\theta_0)$ can be readily deduced from our remarks above, if our model were stationary. Indeed, if the mds $((\partial\varepsilon_t^{0\prime}/\partial\theta)\Sigma_0^{-1}\varepsilon_t^0)$ admits the standard central limit theory (CLT), then we have as $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial\varepsilon_t^{0\prime}}{\partial\theta} \Sigma_0^{-1} \varepsilon_t^0 \to_d \mathbb{N}(0, \Omega) \tag{14}$$

with the asymptotic variance Ω given by

$$\Omega = \plim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial\varepsilon_t^{0\prime}}{\partial\theta} \Sigma_0^{-1} \frac{\partial\varepsilon_t^0}{\partial\theta'}, \tag{15}$$

as is well known. The reader is referred to, e.g., Hall and Heyde (1980) for the details. Due to the nonstationarity of our model, however, the usual CLT for the mds is not applicable for our model, and therefore, the standard asymptotics given by (14) and (15) do not hold. Our subsequent asymptotic theories focus on the case where $(\partial\varepsilon_t^0/\partial\theta')$ is nonstationary, and given by a mixture of integrated and stationary processes.

We now look at more specifically our model introduced in (1). The parameter θ in the model is given by

$$\theta = (\beta', v(\Lambda)')', \tag{16}$$

with the true value $\theta_0 = (\beta_0', v(\Lambda_0)')'$. Here and elsewhere in the paper we use the notation $v(A)$ to denote the subvector of $vec\, A$ with all subdiagonal elements of A eliminated. It is well known that $v(A)$ and $vec\, A$ are related by $Dv(A) = vec\, A$, where D is the matrix called the duplication matrix. See, e.g., Magnus and Neudecker (1988, pp.48-49). The dimension of θ is given by $\kappa = m + m(m+1)/2$, since in particular there are only $m(m+1)/2$ number of nonredundant elements in Λ.

For our model (1), we may easily deduce from Lemma 2.2 and Proposition 2.3 that

Lemma 3.2 For our model (1), we have

$$\frac{\partial \varepsilon_t^{0\prime}}{\partial \beta} = -\left(I - \frac{\Lambda_0^{-1}\beta_0\beta_0'}{\beta_0'\Lambda_0^{-1}\beta_0}\right)x_t + a_t(u,v) \quad \text{and} \quad \frac{\partial \varepsilon_t^{0\prime}}{\partial vec\Lambda} = b_t(u,v),$$

where $a_t(u,v)$ and $b_t(u,v)$ are stationary linear processes driven by (u_t) and (v_t).

It is now clear from Lemma 3.2 that

$$\frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} = \left(\frac{\partial \varepsilon_t^0}{\partial \beta'}, \frac{\partial \varepsilon_t^0}{\partial v(\Lambda)'}\right)'$$

is a matrix time series consisting of a mixture of integrated and stationary processes.

To further analyze the nonstationarity in $(\partial \varepsilon_t^{0\prime}/\partial \theta)$, let $\varepsilon_t = (\varepsilon_{it})_{i=1}^m$ and consider $(\partial \varepsilon_{it}^0/\partial \theta)$ individually for each $i = 1, \ldots, m$. It is easy to see for any $i = 1, \ldots, m$ that $(\partial \varepsilon_{it}^0/\partial \beta)$ is an m-dimensional integrated process with a single common trend. Naturally, there are $(m-1)$-cointegrating relationships in $(\partial \varepsilon_{it}^0/\partial \beta)$ for each $i = 1, \ldots, m$. There is, however, one and only one cointegrating relationship in $(\partial \varepsilon_{it}^0/\partial \beta)$ that is common for all $i = 1, \ldots, m$, which is given by β_0. Notice that

$$P = I - \frac{\Lambda_0^{-1}\beta_0\beta_0'}{\beta_0'\Lambda_0^{-1}\beta_0} \tag{17}$$

is a $(m-1)$-dimensional (non-orthogonal) projection on the space orthogonal to β_0 along $\Lambda_0^{-1}\beta_0$. Consequently, β_0 annihilates the common stochastic trend in $(\partial \varepsilon_{it}^0/\partial \beta)$ for all $i = 1, \ldots, m$, and $(\beta_0'(\partial \varepsilon_{it}^0/\partial \beta))$ becomes stationary for all $i = 1, \ldots, m$. Unlike $(\partial \varepsilon_{it}^0/\partial \beta)$, the process $(\partial \varepsilon_{it}^0/\partial vec\Lambda)$ is purely stationary for all $i = 1, \ldots, m$.

To effectively deal with the singularity of the matrix P in (17), we need to rotate the score vector $s_n(\theta_0)$. To introduce the required rotation, we let Γ_0 be an $m \times (m-1)$ matrix satisfying the conditions

$$\Gamma_0'\Lambda_0^{-1}\beta_0 = 0 \quad \text{and} \quad \Gamma_0'\Lambda_0^{-1}\Gamma_0 = I_{m-1}. \tag{18}$$

It is easy to deduce that

$$P = I - \frac{\Lambda_0^{-1}\beta_0\beta_0'}{\beta_0'\Lambda_0^{-1}\beta_0} = \Lambda_0^{-1}\Gamma_0\Gamma_0', \tag{19}$$

since P is a projection such that $\beta_0'P = P\Lambda_0^{-1}\beta_0 = 0$. Now we define the κ-dimensional rotation matrix

$$T = (T_N, T_S), \tag{20}$$

where T_N and T_S are matrices of dimensions $\kappa \times \kappa_1$ and $\kappa \times \kappa_2$ with $\kappa_1 = m-1$ and $\kappa_2 = 1 + m(m+1)/2$, which are given by

$$T_N = \begin{pmatrix} \Gamma_0 \\ 0 \end{pmatrix} \quad \text{and} \quad T_S = \begin{pmatrix} \dfrac{\beta_0}{(\beta_0'\Lambda_0^{-1}\beta_0)^{1/2}} & 0 \\ 0 & I_{m(m+1)/2} \end{pmatrix}$$

respectively. We have from Lemma 3.2 and (18) and (19)

$$T_N' \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} = -\Gamma_0' x_t + c_t^N(u,v) \quad \text{and} \quad T_S' \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} = c_t^S(u,v) \tag{21}$$

for some stationary linear processes $c_t^N(u,v)$ and $c_t^S(u,v)$ driven by (u_t) and (v_t). Also, we may easily deduce that

$$T^{-1} = \begin{pmatrix} \Gamma_0' \Lambda_0^{-1} & 0 \\ \dfrac{\beta_0' \Lambda_0^{-1}}{(\beta_0' \Lambda_0^{-1} \beta_0)^{1/2}} & 0 \\ 0 & I_{m(m+1)/2} \end{pmatrix} \tag{22}$$

from our definition of T given above in (20).

Lemma 3.3   Under our model (1), the invariance principle holds for the partial sums defined by

$$\Big(U_n(r), V_n(r), W_n(r)\Big) = \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \Sigma_0^{-1} \varepsilon_t^0, \; \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \triangle T_N' \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta}, \; \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} T_S' \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1} \varepsilon_t^0 \right)$$

for $r \in [0,1]$, and we have

$$\Big(U_n, V_n, W_n\Big) \to_d \Big(U, V, W\Big),$$

where U, V and W are (possibly degenerate) Brownian motions such that V and W are independent of U, and such that $\int_0^1 V(r) \Sigma_0^{-1} V(r)' dr$ is of full rank a.s.

The result in Lemma 3.3 enables us to obtain the joint asymptotics of

$$\frac{1}{n} T_N' \sum_{t=1}^{n} \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1} \varepsilon_t^0 \to_d \int_0^1 V(r) \, dU(r), \tag{23}$$

and

$$\frac{1}{\sqrt{n}} T_S' \sum_{t=1}^{n} \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1} \varepsilon_t^0 \to_d W, \tag{24}$$

where we denote $W(1)$ simply by W. This convention will be made for the rest of the paper. Note that the independence of V and U makes the limit distribution in (23) mixed Gaussian. On the other hand, the independence of W and U renders the two limit distributions in (23) and (24) to be independent. Clearly, we have $W =_d \mathbb{N}(0, var(W))$, where

$$var(W) = \plim_{n \to \infty} T_S' \left(\frac{1}{n} \sum_{t=1}^{n} \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1} \frac{\partial \varepsilon_t^0}{\partial \theta'} \right) T_S. \tag{25}$$

Moreover, we may in fact represent the limit distribution in (23) as

$$\left(\int_0^1 B_2(r) \, dB_1(r) \right) I_{\kappa_1}$$

as shown in the proof of Lemma 3.3, where B_1 and B_2 are two independent univariate standard Brownian motions.

We also have

Lemma 3.4 If we let

$$Z_n = \frac{1}{2} T_S' \frac{\partial (vec\, \Sigma_0)'}{\partial \theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) vec \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\varepsilon_t^0 \varepsilon_t^{0\prime} - \Sigma_0 \right) \right],$$

then we have

$$Z_n \to_d Z,$$

where $Z =_d \mathbb{N}(0, var(Z))$ with

$$var(Z) = \frac{1}{2} T_S' \left[\frac{\partial (vec\, \Sigma_0)'}{\partial \theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) \frac{\partial (vec\, \Sigma_0)}{\partial \theta'} \right] T_S, \tag{26}$$

and is independent of U, V and W introduced in Lemma 3.3.

Now we are ready to derive the limit distribution for the ML estimator θ_n of θ defined in (16). It is given by (12) with the rotation matrix T in (20) and the sequence of normalization matrix

$$\nu_n = \text{diag} \left(n I_{\kappa_1}, \sqrt{n} I_{\kappa_2} \right), \tag{27}$$

as we state below as a theorem.

Theorem 3.5 All the conditions in ML1 – ML3 are satisfied for our model (1). In particular, ML1 and ML2 hold, respectively, with

$$N = \begin{pmatrix} -\int_0^1 V(r)\, dU(r) \\ Z - W \end{pmatrix}$$

and

$$M = \begin{pmatrix} \int_0^1 V(r) \Sigma_0^{-1} V(r)' dr & 0 \\ 0 & var(W) + var(Z) \end{pmatrix}$$

in notations introduced in Lemmas 3.3, 3.4, (25) and (26).

Let

$$Q = - \left(\int_0^1 V(r) \Sigma_0^{-1} V(r)' dr \right)^{-1} \int_0^1 V(r)\, dU(r)$$

and

$$\begin{pmatrix} R \\ S \end{pmatrix} = -[var(W) + var(Z)]^{-1}(W - Z),$$

where R and S are 1- and $m(m+1)/2$-dimensional, respectively. Note that Q has a mixed normal distribution, whereas R and S are jointly normal and independent of Q. Now we may readily deduce from Theorem 3.5

$$\sqrt{n}\left(\check{(\hat{\Lambda}_n)} - \check{(\Lambda_0)}\right) \to_d S,$$

and

$$\Gamma_0'\Lambda_0^{-1}\left[n(\hat{\beta}_n - \beta_0)\right] \to_d Q \tag{28}$$

$$\frac{\beta_0'\Lambda_0^{-1}}{(\beta_0'\Lambda_0^{-1}\beta_0)^{1/2}}\left[\sqrt{n}(\hat{\beta}_n - \beta_0)\right] \to_d R. \tag{29}$$

In particular, it follows immediately from (28) and (29) that

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to_d \frac{\beta_0}{(\beta_0'\Lambda_0^{-1}\beta_0)^{1/2}}R, \tag{30}$$

which has degenerate normal distribution. The ML estimators $\hat{\beta}_n$ and $\hat{\Lambda}_n$ converge at the standard rate $\sqrt{n}$, and have limit normal distributions. However, the limit distribution of $\hat{\beta}_n$ is degenerate. In the direction $\Gamma_0'\Lambda_0^{-1}$, it has a faster rate of convergence n and a mixed normal limit distribution. As will be seen in the next section, $\Lambda_0^{-1}\Gamma_0$ is matrix of cointegrating vectors for (y_t). The normal and mixed normal distribution theories of the ML estimators ensure the standard inference to be valid for hypothesis testing in the state space models with an integrated latent common trend. Hence, the usual t-ratios and the asymptotic tests such as Likelihood Ratio, Lagrange Multiplier and Wald tests based on the ML estimates are all valid and can be used in the nonstationary state space models.

It is easy to see that our asymptotic results for the maximum likelihood estimation hold also, at least qualitatively, for the more general state space model introduced in (7). In particular, the convergence rates, degeneracy of the limit distribution and asymptotic Gaussianity that we establish for model (1) are also applicable for model (7). Note that Σ is still a function of only β and Λ for more general measurement equation in (7). This is because Lemma 2.1 is valid for model (7) as well as model (1), as we mentioned earlier in the previous section. The additional parameters $\Phi_1, \ldots, \Phi_p$ appearing in the lagged differenced terms of (y_t) therefore affect the log-likelihood function in (9) only through (ε_t). Moreover, we may observe that the first-order partial derivative of (ε_t) with respect to (Φ_k) yields $(\triangle y_{t-k})$, with all their repeated derivatives vanishing. Consequently, it can be deduced that the score function in Lemma 3.1 with

$$\theta = (\beta', vec(\Phi_1)', \ldots, vec(\Phi_p)', v(\Lambda)')'$$

has now only additional stationary terms involving the products of $(\triangle y_{t-k})$ and (ε_t). In particular, the presence of the additional parameters $\Phi_1, \ldots, \Phi_p$ does not affect the nonstationary component of the score function.

4. Cointegration and Error Correction Representation

Our model (1) implies that (y_t) is cointegrated with the matrix of cointegrating vectors given by an $m \times (m-1)$ matrix $\Lambda_0^{-1}\Gamma_0$. Recall that Γ_0 is the $m \times (m-1)$ matrix satisfying the condition $\Gamma_0'\Lambda_0^{-1}\beta_0 = 0$, as defined in (18). We have $\Gamma_0'\Lambda_0^{-1}y_t = \Gamma_0'\Lambda_0^{-1}u_t$, which is stationary, and therefore $\Lambda_0^{-1}\Gamma_0$ defines the matrix of cointegrating vectors of (y_t). Having $(m-1)$-number of linearly independent cointegrating relationships, (y_t) has one common stochastic trend. Moreover, it follows from Lemma 2.2 that

Proposition 4.1 We have

$$\triangle y_t = -\Gamma_0 A' y_{t-1} - \sum_{k=1}^{t-1} C_k \triangle y_{t-k} + \varepsilon_t^0, \tag{31}$$

where

$$A = \Lambda_0^{-1}\Gamma_0 \quad \text{and} \quad C_k = \frac{\beta_0 \beta_0' \Lambda_0^{-1}}{\beta_0' \Lambda_0^{-1} \beta_0}(1 - 1/\omega_0)^k$$

and we follow our previous convention and use ω_0 and (ε_t^0) to denote ω and (ε_t) evaluated at the true parameter value.

The result in Proposition 4.1 makes clear the relationship between our model (1) and the usual error correction representation of a cointegrated model. Our model (1) given in state space form differs from the conventional error correction model (ECM) in two aspects. First, our ECM model derived from our SS model is not representable as a finite-order vector autoregression (VAR). Here (y_t) is given as VAR(t), where the order increases with time, and therefore it is represented as an infinite-order VAR. Second, our representation implies that we have rank deficiencies in the short-run coefficients (C_k), as well as in the error correction term $\Gamma_0 A'$. Note that (C_k) are of rank one and $A'C_k = 0$ for all $k = 1, 2, \ldots$. In the conventional ECM, on the other hand, there is no such rank restriction imposed on the short-run coefficient matrices.

Our results in the paper for the model (1) can also be used to decompose a cointegrated time series (y_t) into the permanent and transitory components, say (y_t^P) and (y_t^T), such that

$$y_t = y_t^P + y_t^T,$$

where (y_t^P) is I(1) and (y_t^T) is I(0). Of course, the permanent-transitory (PT) decomposition is not unique, and can be done in various ways. The most obvious PT decomposition based on our model is the one given by

$$y_t^P = \beta_0 x_{t|t-1}^0 \quad \text{and} \quad y_t^T = y_t - \beta_0 x_{t|t-1}^0, \tag{32}$$

defining $(x_{t|t-1}^0)$ as the common stochastic trend. The PT decomposition introduced in (32) has the property that (y_t^P) is predictable, while (y_t^T) is a martingale difference sequence. The decomposition introduced in (32) will be referred to as KF-SSM, since we use the

Kalman filter to extract a common trend from a state space model. Obviously, we may estimate the common stochastic trend by $(\hat{x}_{t|t-1})$, i.e., the values of $(x_{t|t-1})$ evaluated at the maximum likelihood estimates of the parameters. The transitory component can also be estimated accordingly.

The PT decomposition proposed by Park (1990) and Gonzalo and Granger (1995) is particularly appealing in our context. The decomposition is based on the error correction representation of a cointegrated system, and is given by

$$y_t^P = \frac{\beta_0 \beta_0' \Lambda_0^{-1}}{\beta_0' \Lambda_0^{-1} \beta_0} y_t = \beta_0 x_t + \frac{\beta_0 \beta_0' \Lambda_0^{-1} u_t}{\beta_0' \Lambda_0^{-1} \beta_0} \tag{33}$$

and

$$y_t^T = \Gamma_0 \Gamma_0' \Lambda_0^{-1} y_t = \Gamma_0 \Gamma_0' \Lambda_0^{-1} u_t. \tag{34}$$

Clearly, (y_t^P) and (y_t^T) defined in (33) and (34) are I(1) and I(0), respectively. They decompose (y_t) into two directions, i.e., β_0 and Γ_0. Note that $\beta_0 \beta_0' \Lambda_0^{-1}/\beta_0' \Lambda_0^{-1} \beta_0$ is the projection on β_0 along Γ_0, and that $\Gamma_0 \Gamma_0' \Lambda_0^{-1}$ is the projection on Γ_0 along β_0. In particular, we have

$$\frac{\beta_0 \beta_0' \Lambda_0^{-1}}{\beta_0' \Lambda_0^{-1} \beta_0} + \Gamma_0 \Gamma_0' \Lambda_0^{-1} = I,$$

and $y_t = y_t^P + y_t^T$.

The directions that are orthogonal to the matrix of cointegrating vectors characterize the long-run equilibrium path of (y_t). In fact β_0 defines the direction orthogonal to the matrix of cointegrating vectors $\Lambda_0^{-1} \Gamma_0$ since $\Gamma_0' \Lambda_0^{-1} \beta_0 = 0$, and therefore, the shocks in the direction of β_0 lie in the equilibrium path of (y_t). This in turn implies that such shocks do not disturb the long-run equilibrium relationships in (y_t). On the other hand, Γ_0 is the matrix of error correction coefficients as shown in Proposition 4.1, and as a result, the shocks in the direction of Γ_0 only have a transient effect. The shocks in every other direction than the direction given by Γ_0 have a permanent effect that may interfere with the long-run equilibrium path of (y_t), thereby distorting the long-run relationships between the components of y_t. The only permanent shocks that do not disturb the equilibrium relationships at the outset are those in the direction of β_0, as discussed above. The reader is referred to Park (1990) for more details. Moreover, the decomposition given in (33) and (34) has an important desirable property that is not present in the usual ECM: The permanent and transitory components are independent of each other. This is because

$$\text{cov}(\beta_0' \Lambda_0^{-1} u_t, \Gamma_0' \Lambda_0^{-1} u_t) = 0,$$

as one may easily check.

In our subsequent empirical applications, we also obtain the decomposition introduced in (33) and (34). The common stochastic trend and stationary component in the decomposition are given by

$$\beta_0' \Lambda_0^{-1} y_t \quad \text{and} \quad \Gamma_0' \Lambda_0^{-1} y_t$$

respectively. These can be readily estimated using the maximum likelihood estimates of the parameters in our model. However, to be more consistent with the methodology in

Park (1990) and Gonzalo and Granger (1995), we rely on the method by Johansen (1988) and compute them using the maximum likelihood estimates of the parameters based on a finite-order ECM. Note that the estimates of the parameters $\Lambda_0^{-1}\beta_0$ and $\Lambda_0^{-1}\Gamma_0$ are easily obtained from the estimates $\hat{\Gamma}$ and $\hat{A}$ respectively of Γ_0 and A in ECM (31): $\Lambda_0^{-1}\beta_0$ is a vector orthogonal to Γ_0 and $\Lambda_0^{-1}\Gamma_0$ is given by A. In what follows, the decomposition will be referred to as ML-ECM, in contrast to KF-SSM.

For the more general state space model with measurement equation in (7), the error correction representation (31) in Proposition 4.1 is not valid. However, we may readily obtain a valid representation from the result in Proposition 4.1 simply by replacing (y_t) by $(y_t - \sum_{k=1}^p \Phi_k \triangle y_{t-k})$. This is obvious from the proof of Proposition 4.1 and the discussions in Section 2. Consequently, for t sufficiently large, we have the error correction representation of (y_t) generated by (7) that is identical to (31) with newly defined coefficients (C_k). The coefficient A would not change. In sum, the general model (7) would have the same error correction representation as in (31) with no rank restriction in (C_k). The decomposition in (32) and that in (33) and (34) can be defined accordingly.

5. Empirical Applications

Selection of $x_{0|0}$ in applications is not as obvious as the selection of $\omega_{0|0}$ discussed in Section 2. The initial value is asymptotically negligible by Lemma 2.2, but it may still have a substantial impact on the likelihood for values of t close to zero and finite n. To this end, Kim and Nelson (1999) and others suggest dropping some of the initial observations when evaluating the likelihood function. We follow a two-step methodology that avoids dropping observations in the final step. We drop some initial observations in the first step, so that we get reasonable parameter estimates regardless of $x_{0|0}$. Once we do this, we re-run the estimation with all observations and with a value of $x_{0|0}$ that is "close" to where the series $(x_{t|t})$ appears to begin.[5] In our applications, $x_{0|0} \neq 0$. By dropping initial observations, we are dropping terms in the summation of conditional likelihoods in which t is small. The effect of the initial condition on the remaining terms is slight. In order to avoid identification problems, we restrict $\sigma_v^2 = 1$ as mentioned in Section 2. To ensure positive semi-definiteness of Λ, we estimate the Cholesky decomposition $\Pi = (\pi_{ij})$ of Λ with the restrictions $\pi_{ii} > 0$ on its diagonal elements.

In order to illustrate the uses of the KF in extracting a common stochastic trend, we examine three well-known applications from the literature. The first two feature bivariate (y_t), and the third features thirty-dimensional (y_t). We use a series that is smoothed according to the usual Kalman smoothing technique, rather than the unsmoothed series analyzed above. The smoothed series utilizes information through the end of the sample. Although this presents an advantage in empirical analysis, it necessarily complicates theoretical analysis, as $(x_{t|n})$ is defined in terms of both past *and* future information. Nevertheless, these

---

[5]We do not have any rigorous methodology for ascertaining "closeness". However, comparing the magnitudes of $(y_t - y_{t|t-1})^2$ for t close to zero with those of $(y_t - y_{t|t-1})^2$ for t close to the end of the sample provides a rough notion. If the magnitudes are similar at the end of the sample and at the beginning of the sample, then $x_{0|0}$ is close enough to x_0 so that differences may be attributed to measurement error.
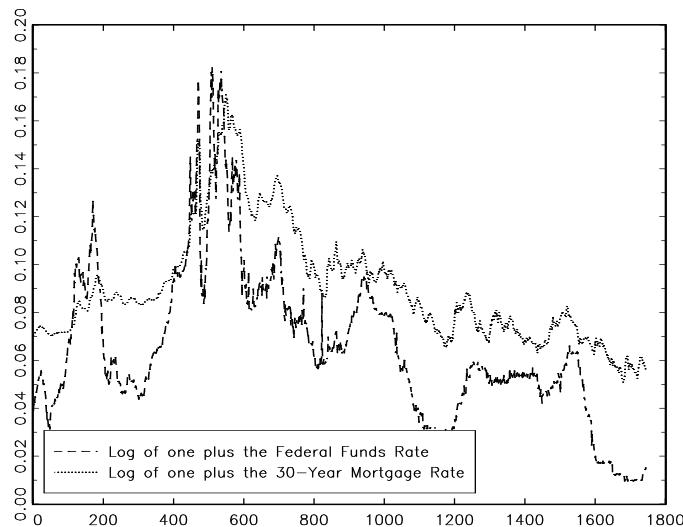
Figure 5.1: Log of one plus the federal funds rate and the 30-year mortgage rate (April 9, 1971 – September 10, 2004).

series should not have noticeably different empirical characteristics. We present the PT decompositions based on both KF-SSM and ML-ECM.

5.1 Short- and Long-Term Interest Rates

There are a number of analyses in the literature aimed at the linkage between short- and long-term interest rates. If economic agents have rational expectations, they buy or sell assets based on expected future interest rates. This means that the rate of return on an asset with a longer term (long rate) is correlated with the rate of return on that of a shorter term (short rate), since investors may choose to purchase a long-term asset or a sequence of short-term assets. We therefore expect a stochastic trend common to rates on assets of different terms. A more thorough discussion may be found in Modigliani and Shiller (1973) or Sargent (1979), for example. Early theoretical analyses of cointegration and estimation using an error correction model, such as Engle and Granger (1987), Campbell and Shiller (1987), and Stock and Watson (1988), use this application to test the relationship between short and long rates. More recent applied work, such as Bauwens et al. (1997) and Hafer et al. (1997), analyze common trends among interest rates in more general international contexts.

We extract the common stochastic trend from the federal funds rate (short rate) and the 30-year conventional fixed mortgage rate (long rate), obtained from the Federal Reserve Bank of St. Louis (http://research.stlouisfed.org/fred2/). These series are sampled at weekly intervals over the period April 9, 1971 through September 10, 2004,[6] and we trans-

---

[6]The federal funds rate is measured on Wednesday of each week, whereas the mortgage rate is measured on Friday of each week.
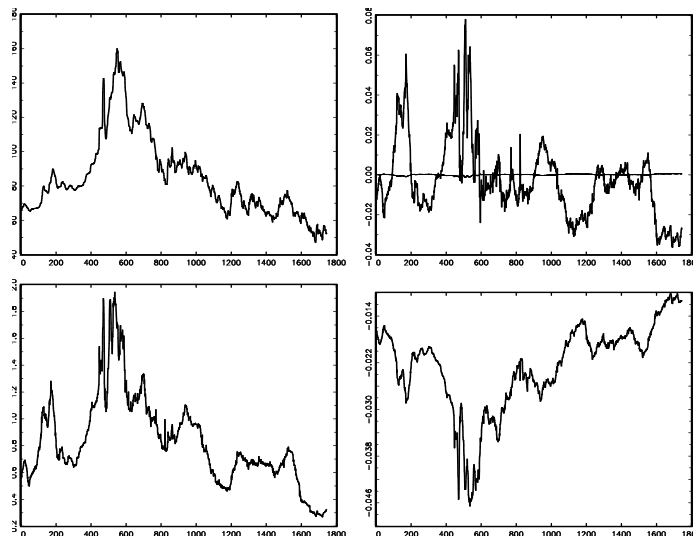
Figure 5.2: (a) Common stochastic trend extracted using KF-SSM; (b) Stationary components extracted using KF-SSM; (c) Common stochastic trend extracted using ML-ECM; (d) Stationary component extracted using ML-ECM.

form them by taking the natural logarithm of one plus the interest rate. The transformed series are illustrated in Figure 5.1. The following table shows parameter estimates using our technique.

Table 5.1: Parameter Estimates from KF-SSM

Parameter	Estimate	Std. Error
β_1	0.0008	1.1×10^{-5}
β_2	0.0011	1.2×10^{-5}
π_{11}	0.0191	0.0003
π_{12}	-0.0003	0.0002
π_{22}	6.0×10^{-11}	3.0×10^{-5}

All parameter estimates are significant except the last one. This possibly degenerate variance implies that the common trend is very similar to the long rate. This implication is clearly visible from the trend extracted using our technique, which is represented in Figure 5.2(a), top left panel. Also, note that the more dominant of the two transitory components [Figure 5.2(b), top right panel] more closely resembles the short rate – although it naturally appears more stationary.

In contrast to KF-SSM, the trend extracted using ML-ECM [Figure 5.2(c), bottom left panel] seems to more closely follow the short rate, as does the transitory component [Figure 5.2(d), bottom right panel] – at least up to a negative scale transformation. This transitory component does not appear to be stationary.
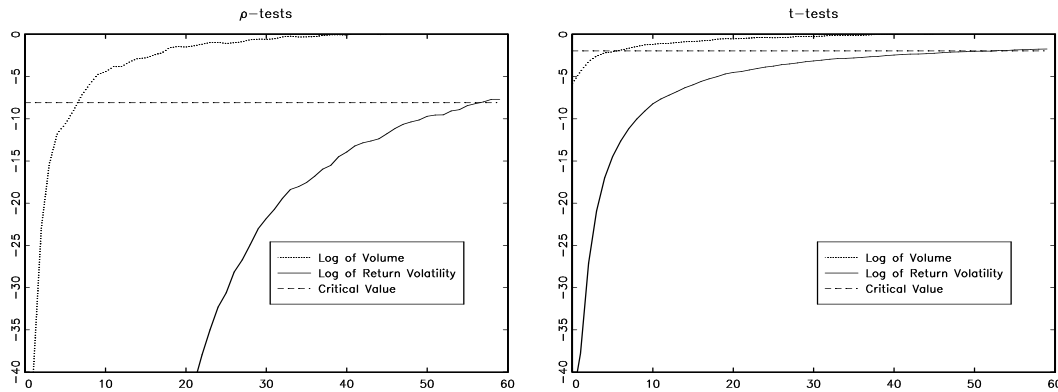
Figure 5.3: (a) ρ-tests and (b) t-tests on the log of return volatility and log of trading volume with increasing p.

5.2 Stock Return Volatility and Trading Volume

There is a large literature on the relationship between stock price volatility and trading volume. Some notable papers among those from the 1980's and early 1990's include Tauchen and Pitts (1983), Karpoff (1987), and Foster and Viswanathan (1995). These analyses generally feature structural models intended to model some of the market microstructure issues involved. More recent analyses, such as Fleming et al. (2004) focus on reduced form models with one or more unobserved common factors. In light of our theoretical results in this paper, these approaches are justified when the common factor is nonstationary. One could conjecture that the approach is also valid with more than one common factor, but the theoretical justification is beyond the scope of the present analysis.

In order to abstract from market microstructure issues, we look at a stock market index. To the extent that nonlinearity enters into the model, this approach may introduce aggregation bias. Specifically, squared returns of the index may understate the actual volatility of the individual stocks, due essentially to Schwartz's inequality. We obtain qualitatively similar results using KF-SSM on individual stock data, but we do not report these results.

In contrast to the interest rates considered in the previous application and to the stock prices considered in the final application, it is not clear that stock return volatility and transaction volume are nonstationary. As these are very noisy time series, we require very many observations to detect nonstationarity. In this application, we examine these measurements with respect to the Dow Jones Industrial Average (DJIA) over the period October 2, 1928 through October 29, 2004. With weekends and holidays omitted, this daily series obtained from Yahoo! Finance (http://finance.yahoo.com/) has 19,017 observations. In order to detect the nonstationarity that is visibly apparent from the data, many lags must be included in an ADF test. The short-term volatility of each series clouds the long-term trend when the order of the ADF test is small. Figures 5.3(a) and (b) illustrate the values of the two ADF tests, ρ-test (test based on normalized unit root coefficient) and t-test (test based on usual t-ratio on unit root coefficient), for these two time series as the number of
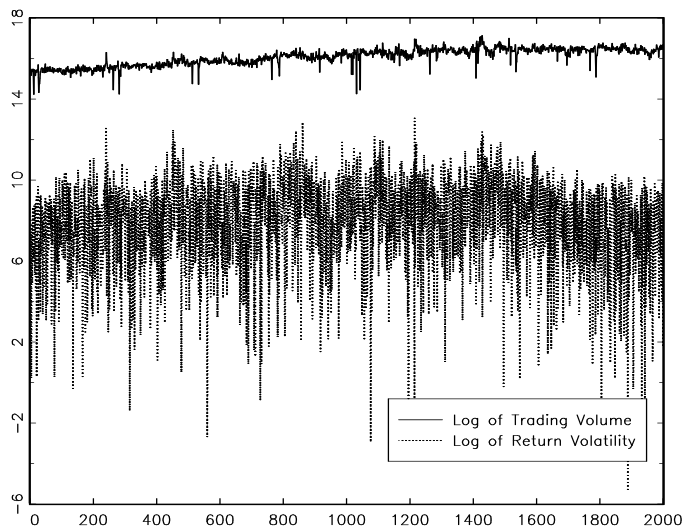
Figure 5.4: Log of the DJIA return volatility and trading volume (November 13, 1996 – October 29, 2004).

lags increases. Nonstationarity is more apparent at higher lags.

Figure 5.4 illustrates the last $2,000$ observations of the time series (November 13, 1996 through October 29, 2004). In order to extract a common stochastic trend, we obtain the following parameter estimates using KF-SSM.

Table 5.2: Parameter Estimates from KF-SSM

Parameter	Estimate	Std. Error
β_1	0.1011	0.0013
β_2	0.0231	0.0003
π_{11}	0.1655	0.0014
π_{12}	1.1579	0.0356
π_{22}	2.9606	0.0184

Our trend [Figure 5.5(a), top left panel] seems to be more closely correlated with trading volume. Since the variance parameter π_{11} and coefficient β_1 associated with trading volume are relatively small and large, respectively, the stronger correlation with this series is natural. The transitory components are illustrated in Figure 5.5(b), top right panel.

We obtain a similar stochastic trend using ML-ECM [Figure 5.5(c), bottom left panel], although the ML-ECM trend appears more volatile. The transitory component extracted using ML-ECM [Figure 5.5(d), bottom right panel] seems to resemble return volatility, which we argue is not stationary.
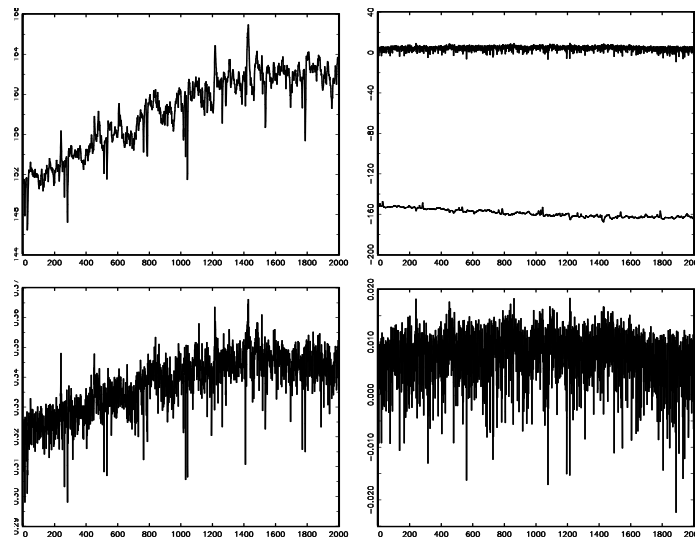
Figure 5.5: (a) Common stochastic trend extracted using KF-SSM; (b) Stationary components extracted using KF-SSM; (c) Common stochastic trend extracted using ML-ECM; (d) Stationary component extracted using ML-ECM.

5.3 Stock Market Index

In the final application, we consider a stock market index. We wish to extract the common stochastic trend embedded in 30 series of prices of the stocks that comprise the DJIA. We may thus compare the common trend with the index itself. This methodology could easily be used to extract the common stochastic trend from the prices of any set of stocks. The novelty of this approach is that it may easily be generalized to incorporate *any* group of assets. KF-SSM provides an easy way to extract a common stochastic trend or a customized index, with weights that are optimally chosen by the algorithm. This avoids the need to weight stocks by market capitalization or by trading volume, as some indices do.

We expect ML-ECM to fail in this application, because the parameters are estimated using Johansen's method. Essentially, ML-ECM is designed to extract the best trends, with the number of trends unrestricted. KF-SSM extracts the best single trend. In two-dimensional applications, such as the previous two, both approaches must extract only one trend. In higher-dimensional applications, only one trend may be desired. In this case, KF-SSM provides a better approach.

Gonzalo and Granger (1995) recommend dividing such a high-dimensional cointegrated system into cohorts when analyzing with ML-ECM.[7] Essentially, those authors propose a two-step methodology to extract a single common trend. A common trend is extracted

---

[7]To put their recommendation in proper context, it should be noted that Gonzalo and Granger (1995) propose the cohort approach more to justify extracting common trends from cohorts of cointegrated systems than to actually estimate the parameters of a high-dimensional cointegrated system. They do *not* propose the cohort approach as an *alternative* to one-step estimation.
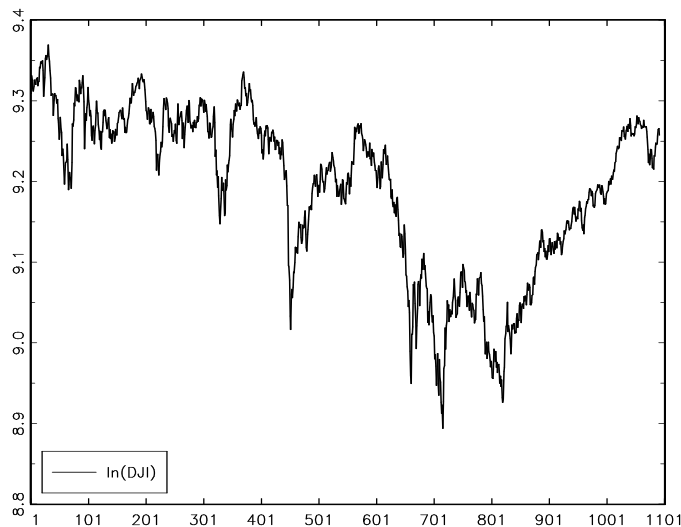
Figure 5.6: Log of the DJIA (December 2, 1999 – April 7, 2004).

from the common trends that are first extracted from each of these cohorts. This reduces the dimensionality involved in each step. Natural cohorts may be difficult to identify. For example, in this application, it would be reasonable to group American Express, Citigroup, and JP Morgan Chase as financial companies. But other companies such as Exxon Mobil cannot easily be grouped with others. Moreover, with current computing power, this is unnecessary. Calculations with a 30×30 matrix can be accomplished with GAUSS on a desktop computer in a reasonable amount of time.

In this application, we use prices from Yahoo! Finance (http://finance.yahoo.com/). These daily closing prices are adjusted to take into account stock splits and dividends using the methodology developed by the Center for Research in Security Prices. Figure 5.6 illustrates the DJIA observed from December 2, 1999 through April 7, 2004, which is the longest recent stretch during which the companies comprising the DJIA did not change (we have $1,092$ observations, excluding weekends and holidays). Figure 5.7 shows the 30 series. We include this figure to illustrate the behavioral diversity of these 30 stocks comprising the index. It is not obvious from casual observation of these 30 series how the common stochastic trend should look.

The implementation of KF-SSM is not as straightforward as that of ML-ECM, since numerical optimization is required. In order to reduce the complexity and necessary computing time created by the large number of parameters, we restrict Λ to be diagonal in this application ($\pi_{ij} = 0$ for $i \neq j$). This reduces the number of parameters to be estimated using MLE from 495 to 60. Parameter estimates and standard errors for β using KF-SSM are given in the following table.
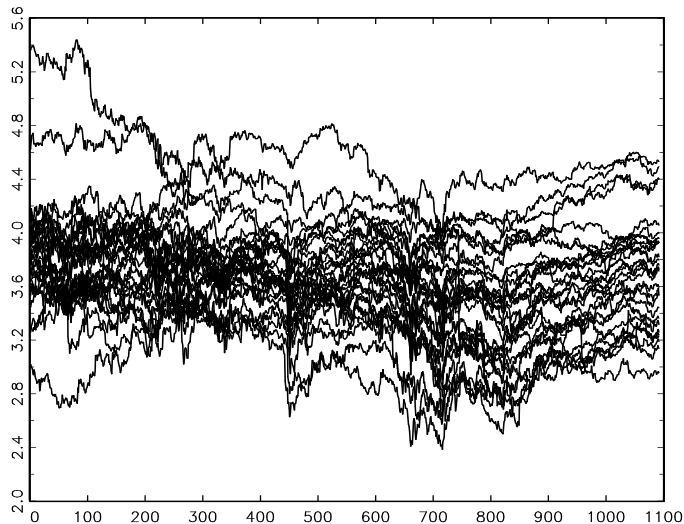
Figure 5.7: Log of the 30 stocks comprising the DJIA (December 2, 1999 – April 7, 2004).

Table 5.3: Parameter Estimates from KF-SSM

Parameter	Estimate	Std. Error	Parameter	Estimate	Std. Error
β_1	0.0093	5.7×10^{-5}	β_{16}	0.0087	5.1×10^{-5}
β_2	0.0080	4.6×10^{-5}	β_{17}	0.0082	4.8×10^{-5}
β_3	0.0084	5.4×10^{-5}	β_{18}	0.0080	5.2×10^{-5}
β_4	0.0088	5.1×10^{-5}	β_{19}	0.0107	6.1×10^{-5}
β_5	0.0094	7.5×10^{-5}	β_{20}	0.0084	4.9×10^{-5}
β_6	0.0087	5.1×10^{-5}	β_{21}	0.0084	5.0×10^{-5}
β_7	0.0090	5.5×10^{-5}	β_{22}	0.0091	5.4×10^{-5}
β_8	0.0087	5.0×10^{-5}	β_{23}	0.0076	4.5×10^{-5}
β_9	0.0091	5.2×10^{-5}	β_{24}	0.0094	5.4×10^{-5}
β_{10}	0.0087	5.0×10^{-5}	β_{25}	0.0080	4.7×10^{-5}
β_{11}	0.0083	4.9×10^{-5}	β_{26}	0.0084	5.1×10^{-5}
β_{12}	0.0085	4.8×10^{-5}	β_{27}	0.0080	4.8×10^{-5}
β_{13}	0.0084	4.9×10^{-5}	β_{28}	0.0098	5.7×10^{-5}
β_{14}	0.0090	5.2×10^{-5}	β_{29}	0.0093	5.4×10^{-5}
β_{15}	0.0074	4.8×10^{-5}	β_{30}	0.0075	4.5×10^{-5}

In the interest of brevity, we do not report parameter estimates or standard errors for the variance parameters.

We cannot reproduce the DJIA exactly, because we restrict $\sigma_v^2 = 1$ for the identification of β.[8] In light of this restriction, we expect to extract a common trend resembling the

[8]Alternatively, we could set σ_v^2 equal to the variance of the differenced DJIA, which might reasonably
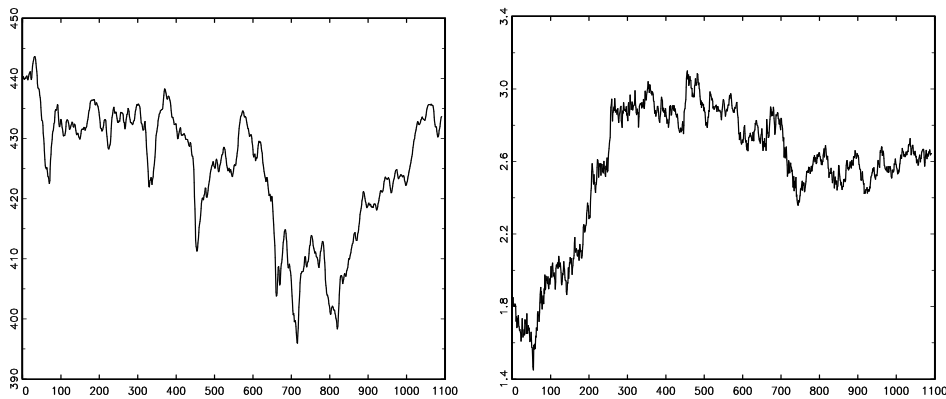
Figure 5.8: (a) Common stochastic trend extracted using KF-SSM; (b) Common stochastic trend extracted using ML-ECM.

DJIA only up to an affine transformation. Figure 5.8(a) shows the common trend extracted using KF-SSM, and may be directly compared with the actual DJIA shown in Figure 5.6. The similarity suggests that KF-SSM works quite well. (We do not illustrate the transitory components in this application, since the thirty-dimensional series generated by KF-SSM does not contribute anything substantial to our exposition.)

Figure 5.8(b) shows the trend extracted using ML-ECM. The failure of ML-ECM to replicate anything resembling the actual Dow Jones Industrial Average supports our intuition about the dangers of restricting the number of trends to be estimated when using that technique. Essentially, it restricts the number of common trends *after* multiple trends are extracted. Whereas, KF-SSM imposes the restriction *before* the trend is extracted. If the restriction is desired – as it is in this application – KF-SSM is clearly a more appealing approach.

6. Conclusions

The chief aim of this paper from a theoretical point of view is to justify the use of the Kalman filter when the underlying state space model contains integrated time series. Specifically, this class of models is useful when a single stochastic trend is common to a vector of observable cointegrated time series. Our technique is certainly not novel. The literature contains many applications that employ the Kalman filter when the assumption of stationarity cannot be maintained. The filter seems to work reasonably well in such applications, but there is no well-known theory to support its use in the nonstationary case. Therein lies the raison d'être of our theoretical analysis. Our empirical applications demonstrate how our models and theories are useful in practice.

yield a common trend more closely resembling the DJIA. However, in more general applications, no obvious choice for σ_v^2 exists.

Our research suggests many avenues for future efforts along these lines. We limit our focus to extracting a single stochastic trend, but certainly the Kalman filter could be applied with an unobservable vector of trends. More formal testing procedures could perhaps be developed to test for the number of trends, as has been done for error correction models. An advantage of the approach based on the state space models is that it does not require nonstationarity. It is easily conceivable that a common trend with a near-unit root or an unobservable vector with a combination of stationary and nonstationary components, for example, could be extracted using such an approach. Although we find evidence of a unit root in the common trend extracted from stock return volatility and trading volume, we do not present a formal test. We leave this and other challenges for future research.

Mathematical Proofs

Proof of Lemma 2.1   We have

$$\omega_{t+1|t} = 1 + \omega_{t|t-1} - \omega_{t|t-1}^2 \beta' \Sigma_{t|t-1}^{-1} \beta, \tag{35}$$

which follows directly from the prediction and updating steps of the Kalman filter. Moreover, we may easily deduce that

$$\Sigma_{t|t-1}^{-1} = \Lambda^{-1} - \frac{\omega_{t|t-1}}{1 + \omega_{t|t-1}\beta' \Lambda^{-1}\beta} \Lambda^{-1} \beta \beta' \Lambda^{-1}, \tag{36}$$

and therefore,

$$\beta' \Sigma_{t|t-1}^{-1} \beta = \beta' \Lambda^{-1} \beta - \frac{\omega_{t|t-1}(\beta' \Lambda^{-1}\beta)^2}{1 + \omega_{t|t-1}\beta' \Lambda^{-1}\beta} = \frac{\beta' \Lambda^{-1}\beta}{1 + \omega_{t|t-1}\beta' \Lambda^{-1}\beta}. \tag{37}$$

Therefore, it follows from (35) and (37) that

$$\omega_{t+1|t} = 1 + \frac{\omega_{t|t-1}}{1 + \omega_{t|t-1}\beta' \Lambda^{-1}\beta}, \tag{38}$$

which defines a first order difference equation for $\omega_{t|t-1}$.

Now we may readily see that the first order difference equation in (38) has the unique asymptotic steady state solution given in Lemma 1. For this, consider the function

$$f(x) = 1 + \frac{x}{1 + \tau x}$$

for $x \geq 1$, with any $\tau \geq 0$. The function has an intersection with the identity function at

$$x = \frac{1 + \sqrt{1 + 4/\tau}}{2}$$

over its domain $x \geq 1$. Moreover, the function is monotone increasing with

$$f'(x) = \frac{1}{(1 + \tau x)^2} < 1$$

for all $x \geq 1$. This completes the proof for the existence of the stable value of $\omega_{t|t-1}$. The proof of $\Sigma_{t|t-1}$ follows immediately from that of $\omega_{t|t-1}$. $\square$

Proof of Lemma 2.2   From the prediction and updating steps of the Kalman filter, we have

$$\begin{aligned}
x_{t+1|t} &= x_{t|t-1} + \omega_{t|t-1}\beta' \Sigma_{t|t-1}^{-1}(y_t - y_{t|t-1}) \\
&= x_{t|t-1} + \omega_{t|t-1}\beta' \Sigma_{t|t-1}^{-1}(y_t - \beta x_{t|t-1}).
\end{aligned} \tag{39}$$

However, it follows from (36) that

$$\Sigma^{-1} = \Lambda^{-1} - \frac{\omega}{1 + \omega\,\beta'\Lambda^{-1}\beta}\,\Lambda^{-1}\beta\beta'\Lambda^{-1}, \tag{40}$$

where Σ is the asymptotic steady state value of $\Sigma_{t|t-1}$ given in (3). Therefore, we have from (36)

$$
\begin{aligned}
\Sigma^{-1}\beta &= \Lambda^{-1}\beta - \frac{\omega}{1 + \omega\,\beta'\Lambda^{-1}\beta}\,\Lambda^{-1}\beta\beta'\Lambda^{-1}\beta \\
&= \frac{1}{1 + \omega\,\beta'\Lambda^{-1}\beta}\,\Lambda^{-1}\beta, \\
&= \frac{\beta'\Lambda^{-1}\beta}{1 + \omega\,\beta'\Lambda^{-1}\beta}\,\frac{\Lambda^{-1}\beta}{\beta'\Lambda^{-1}\beta}
\end{aligned} \tag{41}
$$

and

$$
\begin{aligned}
\beta'\Sigma^{-1}\beta &= \beta'\Lambda^{-1}\beta - \frac{\omega}{1 + \omega\,\beta'\Lambda^{-1}\beta}\,\beta'\Lambda^{-1}\beta\beta'\Lambda^{-1}\beta \\
&= \frac{\beta'\Lambda^{-1}\beta}{1 + \omega\,\beta'\Lambda^{-1}\beta}.
\end{aligned} \tag{42}
$$

Moreover, we may deduce that

$$\omega\,\beta'\Sigma^{-1}\beta = \frac{\omega\,\beta'\Lambda^{-1}\beta}{1 + \omega\,\beta'\Lambda^{-1}\beta} = \frac{1}{\omega} \tag{43}$$

due to (35).

Now we have from (39), (41), (42) and (43) that

$$
\begin{aligned}
x_{t+1|t} &= x_{t|t-1} + \frac{1}{\omega}\frac{\beta'\Lambda^{-1}}{\beta'\Lambda^{-1}\beta}y_t - \frac{1}{\omega}x_{t|t-1} \\
&= \left(1 - \frac{1}{\omega}\right)x_{t|t-1} + \frac{1}{\omega}\frac{\beta'\Lambda^{-1}}{\beta'\Lambda^{-1}\beta}\,y_t,
\end{aligned}
$$

and consequently,

$$x_{t|t-1} = \frac{1}{\omega}\sum_{k=1}^{t-1}\left(1 - \frac{1}{\omega}\right)^{k-1}\frac{\beta'\Lambda^{-1}}{\beta'\Lambda^{-1}\beta}\,y_{t-k} + \left(1 - \frac{1}{\omega}\right)^{t-1}x_{1|0}. \tag{44}$$

Moreover,

$$
\begin{aligned}
\frac{1}{\omega}\sum_{k=1}^{t-1}\left(1 - \frac{1}{\omega}\right)^{k-1}y_{t-k} &= \left[1 - \left(1 - \frac{1}{\omega}\right)\right]\sum_{k=1}^{t-1}\left(1 - \frac{1}{\omega}\right)^{k-1}y_{t-k} \\
&= y_t - \sum_{k=0}^{t-2}\left(1 - \frac{1}{\omega}\right)^{k}\triangle y_{t-k} - \left(1 - \frac{1}{\omega}\right)^{t-1}y_1.
\end{aligned} \tag{45}
$$

The stated result now follows from (44) and (45) in a straightforward manner. Note that $x_{1|0} = x_{0|0} = x_0$ and $y_0 = 0$. The proof is therefore complete. $\qquad\square$

Proof of Proposition 2.3 It follows from Lemma 2.2 that

$$
\begin{aligned}
x_{t|t-1}^0 &= \frac{\beta_0'\Lambda_0^{-1}}{\beta_0'\Lambda_0^{-1}\beta_0}\left[(\beta_0 x_t + u_t) - \sum_{k=0}^{t-1}\left(1 - \frac{1}{\omega_0}\right)^k(\beta_0 v_{t-k} + (u_{t-k} - u_{t-k-1}))\right] \\
&= x_t + \frac{\beta_0'\Lambda_0^{-1}}{\beta_0'\Lambda_0^{-1}\beta_0}\left[u_t - \sum_{k=0}^{t-1}\left(1 - \frac{1}{\omega_0}\right)^k(u_{t-k} - u_{t-k-1})\right] - \sum_{k=0}^{t-1}\left(1 - \frac{1}{\omega_0}\right)^k v_{t-k} \quad (46)
\end{aligned}
$$

However, we may easily deduce that

$$
\sum_{k=0}^{t-1}\left(1 - \frac{1}{\omega_0}\right)^k(u_{t-k} - u_{t-k-1}) = u_t - \frac{1}{\omega_0}\sum_{k=1}^{t-1}\left(1 - \frac{1}{\omega_0}\right)^{k-1}u_{t-k}. \quad (47)
$$

The stated result now follows readily from (46) and (47). □

Proof of Lemma 3.1 The proof just requires the standard rules of differentiation, and the details are therefore omitted. □

Proof of Lemma 3.2 In the proof, we use the generic notation w_t to signify any stationary linear process driven by (u_t) and (v_t). In particular, the definition of w_t is different from line to line. Due to Lemma 2.2, we may write

$$
x_{t|t-1} = \frac{\beta'\Lambda^{-1}}{\beta'\Lambda^{-1}\beta}y_t + w_t. \quad (48)
$$

for all values of β and Λ. Moreover, we may readily deduce that

$$
\frac{\partial x_{t|t-1}}{\partial \beta} = \frac{\Lambda^{-1}}{\beta'\Lambda^{-1}\beta}\left(I - 2\frac{\beta\beta'\Lambda^{-1}}{\beta'\Lambda^{-1}\beta}\right)y_t + w_t \quad (49)
$$

$$
\frac{\partial x_{t|t-1}}{\partial vec\Lambda} = -\frac{\Lambda^{-1}\beta \otimes \Lambda^{-1}}{\beta'\Lambda^{-1}\beta}\left(I - \frac{\beta\beta'\Lambda^{-1}}{\beta'\Lambda^{-1}\beta}\right)y_t + w_t \quad (50)
$$

for all values of β and Λ. As a consequence, if we use the superscript "0" to denote the derivative $\partial x_{t|t-1}/\partial\theta$ evaluated at the true parameter values consistently with our earlier notations, then we have

$$
\frac{\partial x_{t|t-1}^0}{\partial \beta} = -\frac{\Lambda_0^{-1}\beta_0}{\beta_0'\Lambda_0^{-1}\beta_0}x_t + w_t \quad \text{and} \quad \frac{\partial x_{t|t-1}^0}{\partial vec\Lambda} = w_t. \quad (51)
$$

We now note that

$$
\frac{\partial \varepsilon_t^{0\prime}}{\partial \beta} = -x_{t|t-1}^0 I - \frac{\partial x_{t|t-1}^0}{\partial \beta}\beta' \quad \text{and} \quad \frac{\partial \varepsilon_t^{0\prime}}{\partial vec\Lambda} = -\frac{\partial x_{t|t-1}^0}{\partial vec\Lambda}\beta',
$$

from which the stated result follows immediately. □

Proof of Lemma 3.3 It follows follow immediately from (21) that

$$V_n(r) = -\Gamma'_0 \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} v_t + o_p(1).$$

Moreover, due to (21), $T'_S(\partial \varepsilon_t^{0\prime}/\partial \theta)$ is a stationary linear process. Moreover, it is $\mathcal{F}_{t-1}$-measurable. Consequently, W_n is a partial sum process of the martingale difference sequence $T'_S(\partial \varepsilon_t^{0\prime}/\partial \theta)\Sigma_0^{-1}\varepsilon_t^0$. The stated results can therefore be readily deduced from the invariance principle for the martingale difference sequence. □

Proof of Lemma 3.4 To deduce the stated result, we simply note that

$$\frac{\partial(vec\,\Sigma_0)'}{\partial \theta} vec \left[\Sigma_0^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\varepsilon_t^0 \varepsilon_t^{0\prime} - \Sigma_0 \right) \Sigma_0^{-1} \right]$$

$$= \frac{\partial v(\Sigma_0)'}{\partial \theta} D' \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) vec \left[\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(\varepsilon_t^0 \varepsilon_t^{0\prime} - \Sigma_0 \right) \right]$$

$$\to_d \frac{\partial v(\Sigma_0)'}{\partial \theta} D' \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) \mathbb{N} \Big(0, (I + K)(\Sigma_0 \otimes \Sigma_0) \Big)$$

$$=_d \frac{\partial v(\Sigma_0)'}{\partial \theta} \mathbb{N} \left(0, 2D' \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) D \right)$$

$$=_d \frac{\partial(vec\,\Sigma_0)'}{\partial \theta} \mathbb{N} \left(0, 2 \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) \right).$$

Here we use the fact

$$KD = D,$$

as shown in, e.g., Magnus and Neudecker (1988, p.49). □

Proof of Theorem 3.5 The proof will be done in three steps, for each of ML1 – ML3. In the proof, we use the following notational convention:

(a) (w_t) denotes a linear process driven by $(u_s)_{s=1}^t$ and $(v_s)_{s=1}^t$ that has geometrically decaying coefficients, and

(b) $(\overline{w}_t)$ is such a process that is $\mathcal{F}_t$-measurable.

The notations w_t and $\overline{w}_t$ are *generic* and signify any processes satisfying the conditions specified above. In general, w_t and $\overline{w}_t$ appearing in different lines represent different processes.

First Step   We have

$$
\begin{aligned}
\frac{1}{n}T'_N s_n(\theta_0) &= \frac{1}{2\sqrt{n}}T'_N \frac{\partial(vec\,\Sigma_0)'}{\partial\theta}\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\right)vec\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\varepsilon_t^0\varepsilon_t^{0\prime}-\Sigma_0\right)\right] \\
&\quad -\frac{1}{n}\sum_{t=1}^{n}T'_N\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\Sigma_0^{-1}\varepsilon_t^0 \\
&= -\frac{1}{n}\sum_{t=1}^{n}T'_N\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\Sigma_0^{-1}\varepsilon_t^0 + O_p(n^{-1/2}) \\
&= -\int_0^1 V_n(r)\,dU_n(r) + o_p(1) \\
&\to_d -\int_0^1 V(r)\,dU(r) \tag{52}
\end{aligned}
$$

as $n\to\infty$. On the other hand, we may deduce that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}T'_S s_n(\theta_0) &= \frac{1}{2}T'_S \frac{\partial(vec\,\Sigma_0)'}{\partial\theta}\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\right)vec\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\varepsilon_t^0\varepsilon_t^{0\prime}-\Sigma_0\right)\right] \\
&\quad -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}T'_S\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\Sigma_0^{-1}\varepsilon_t^0 \\
&= Z_n - W_n \\
&\to_d Z - W \tag{53}
\end{aligned}
$$

as $n\to\infty$. Consequently, it follows from (52) and (53) that ML1 holds with N given in the theorem.

Second Step   Next we establish ML2. First, we note that

$$
\begin{aligned}
\frac{1}{n^2}T'_N H_n(\theta_0)T_N &= -\frac{1}{n^2}T'_N\left(\sum_{t=1}^{n}\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\Sigma_0^{-1}\frac{\partial\varepsilon_t^0}{\partial\theta'}\right)T_N + O_p(n^{-1}) \\
&= -\int_0^1 V_n(r)\Sigma_0^{-1}V_n(r)'dr + o_p(1) \\
&\to_d -\int_0^1 V(r)\Sigma_0^{-1}V(r)'dr \tag{54}
\end{aligned}
$$

as $n\to\infty$, and that

$$
\frac{1}{n^{3/2}}T'_N H_n(\theta_0)T_S = O_p(n^{-1/2}) \tag{55}
$$

for large n, which are in particular due to

$$\sum_{t=1}^{n} \left(I \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1}\right) \left(\frac{\partial^2}{\partial\theta\partial\theta'} \otimes \varepsilon_t^0\right) = O_p(n)$$

$$\frac{\partial(vec\,\Sigma_0)'}{\partial\theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \sum_{t=1}^{n} \left(\frac{\partial\varepsilon_t^0}{\partial\theta'} \otimes \varepsilon_t^0\right) = O_p(n)$$

$$\sum_{t=1}^{n} \left(\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta} \otimes \varepsilon_t^{0\prime}\right) \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \frac{\partial(vec\,\Sigma_0)}{\partial\theta'} = O_p(n)$$

for large n.

Secondly, we show that

$$\frac{1}{n} T_S' H_n(\theta_0) T_S \to_p var(W) + var(Z), \tag{56}$$

which establishes ML2, together with (54) and (55). To derive (56), we first write

$$\frac{1}{n} T_S' H_n(\theta_0) T_S = A + B_n + C_n + (D_n + D_n'), \tag{57}$$

where

$$A = -\frac{1}{2} T_S' \left[\frac{\partial(vec\,\Sigma_0)'}{\partial\theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \frac{\partial(vec\,\Sigma_0)}{\partial\theta'}\right] T_S$$

$$B_n = -\frac{1}{n} \sum_{t=1}^{n} T_S' \left(\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta} \Sigma_0^{-1} \frac{\partial\varepsilon_t^0}{\partial\theta'}\right) T_S$$

$$C_n = -\frac{1}{n} \sum_{t=1}^{n} T_S' \left[\left(I \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1}\right) \left(\frac{\partial^2}{\partial\theta\partial\theta'} \otimes \varepsilon_t^0\right)\right] T_S$$

$$D_n = T_S' \left[\frac{\partial(vec\,\Sigma_0)'}{\partial\theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial\varepsilon_t^0}{\partial\theta'} \otimes \varepsilon_t^0\right)\right] T_S.$$

As shown earlier, we have

$$A = var(Z) \quad \text{and} \quad B_n = var(W) + o_p(1). \tag{58}$$

Moreover, since

$$T_S' \frac{\partial\varepsilon_t^{0\prime}}{\partial\theta} = \overline{w}_{t-1},$$

we have

$$D_n = T_S' \left[\frac{\partial(vec\,\Sigma_0)'}{\partial\theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1}\right) \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\partial\varepsilon_t^0}{\partial\theta'} T_S \otimes \varepsilon_t^0\right)\right] = O_p(n^{-1/2}) \tag{59}$$

for large n.

Now it suffices to show that

$$C_n = \begin{pmatrix} C_n(\beta, \beta) & C_n(\beta, \Lambda) \\ C_n(\Lambda, \beta) & C_n(\Lambda, \Lambda) \end{pmatrix} = O_p(n^{-1/2}), \tag{60}$$

since (56) follows immediately from (57) – (60). For the subsequent proof, it will be very useful to note that

$$x_{t|t-1} + \beta' \frac{\partial x_{t|t-1}}{\partial \beta} = \overline{w}_{t-1}$$

for all β and Λ. Therefore, we have upon differentiating with respect to β and Λ

$$2\frac{\partial x_{t|t-1}}{\partial \beta'} + \beta' \frac{\partial^2 x_{t|t-1}}{\partial\beta\partial\beta'} = \overline{w}_{t-1} \tag{61}$$

$$\frac{\partial x_{t|t-1}}{\partial(vec\,\Lambda)'} + \beta' \frac{\partial^2 x_{t|t-1}}{\partial\beta\partial(vec\,\Lambda)'} = \overline{w}_{t-1}, \tag{62}$$

which hold for all β and Λ.

For (60), we first prove

$$C_n(\beta, \beta) = \frac{1}{n}\sum_{t=1}^{n} \beta_0' \left[(I \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}) \left(\frac{\partial^2}{\partial\beta\partial\beta'} \otimes \varepsilon_t^0 \right) \right] \beta_0 = O_p(n^{-1/2}). \tag{63}$$

This can be easily derived, since we have

$$\begin{aligned}
\left(\frac{\partial^2}{\partial\beta\partial\beta'} \otimes \varepsilon_t^0 \right) \beta_0 &= \left(\frac{\partial}{\partial\beta'} vec\, \frac{\partial\varepsilon_t^{0\prime}}{\partial\beta} \right) \beta_0 \\
&= \left[-(vec\, I)\frac{\partial x_{t|t-1}^0}{\partial\beta'} - \frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial\beta'} \otimes \beta_0 - \frac{\partial x_{t|t-1}^0}{\partial\beta} \otimes I \right] \beta_0 \\
&= -vec \left[\left(\frac{\partial x_{t|t-1}^0}{\partial\beta'}\beta_0 \right) I + \frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial\beta'}\beta_0\beta_0' + \frac{\partial x_{t|t-1}^0}{\partial\beta}\beta_0' \right] \\
&= -vec \left[\left(\frac{\partial x_{t|t-1}^0}{\partial\beta'}\beta_0 \right) I - \frac{\partial x_{t|t-1}^0}{\partial\beta}\beta_0' \right] + \overline{w}_{t-1},
\end{aligned}$$

and it follows that

$$\begin{aligned}
&\beta_0' \left[(I \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}) \left(\frac{\partial^2}{\partial\beta\partial\beta'} \otimes \varepsilon_t^0 \right) \right] \beta_0 \\
&= -\beta_0' \left[\left(\frac{\partial x_{t|t-1}^0}{\partial\beta'}\beta_0 \right) I - \frac{\partial x_{t|t-1}^0}{\partial\beta}\beta_0' \right] \Sigma_0^{-1}\varepsilon_t^0 + \overline{w}_{t-1}\varepsilon_t^0 = \overline{w}_{t-1}\varepsilon_t^0,
\end{aligned}$$

from which we may deduce (63), upon noticing

$$\sum_{t=1}^{n} \overline{w}_{t-1}\varepsilon_t^0 = O_p(n^{1/2})$$

for large n.

Similarly, we have for any vector λ of conformable dimension

$$
\begin{aligned}
\left(\frac{\partial^2}{\partial\beta\partial(vec\,\Lambda)'} \otimes \varepsilon_t^0 \right) \lambda &= \left(\frac{\partial}{\partial(vec\,\Lambda)'} vec\, \frac{\partial\varepsilon_t^{0\prime}}{\partial\beta} \right) \lambda \\
&= \left[-(vec\,I)\frac{\partial x_{t|t-1}^0}{\partial(vec\,\Lambda)'} - \frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial(vec\,\Lambda)'} \otimes \beta_0 \right] \lambda \\
&= -vec \left[\left(\frac{\partial x_{t|t-1}^0}{\partial(vec\,\Lambda)'}\lambda \right) I + \frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial(vec\,\Lambda)'}\lambda\beta_0' \right],
\end{aligned}
$$

and it follows that

$$
\begin{aligned}
\beta_0' &\left[(I \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}) \left(\frac{\partial^2}{\partial\beta\partial(vec\,\Lambda)'} \otimes \varepsilon_t^0 \right) \right] \lambda \\
&= -\beta_0' \left[\left(\frac{\partial x_{t|t-1}^0}{\partial(vec\,\Lambda)'}\lambda \right) I + \frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial(vec\,\Lambda)'}\lambda\beta_0' \right] \Sigma_0^{-1}\varepsilon_t^0 = \overline{w}_{t-1}\varepsilon_t^0,
\end{aligned}
$$

as was to be shown. The proof for $C_n(\Lambda, \Lambda)$ is straightforward, since

$$
\frac{\partial x_{t|t-1}^0}{\partial vec\,\Lambda} = \overline{w}_{t-1} \quad \text{and} \quad \frac{\partial^2 x_{t|t-1}^0}{\partial vec\,\Lambda\partial(vec\,\Lambda)'} = \overline{w}_{t-1}. \tag{64}
$$

This can be easily deduced after some tedious but straightforward algebra. The proof for ML2 with given M is therefore complete.

Third Step   To establish ML3, we let

$$
\mu_n = \nu_n^{1-\delta}
$$

for some $\delta > 0$ small, and let $\theta \in \Theta_n$ be arbitrarily chosen. Since

$$
\Gamma_0'\Lambda_0^{-1}(\beta - \beta_0) = O(n^{-1+\delta})
$$
$$
\frac{\beta_0'\Lambda_0^{-1}}{(\beta_0'\Lambda_0^{-1}\beta_0)^{1/2}}(\beta - \beta_0) = O(n^{-1/2+\delta}),
$$

we may set

$$
\beta = \beta_0 + n^{-1/2+\delta}\beta_0 + n^{-1+\delta}\Gamma_0 \tag{65}
$$
$$
\Lambda = \Lambda_0 + n^{-1/2+\delta}I \tag{66}
$$

without loss of generality.

We will show that

$$\frac{1}{n^{2(1-\delta)}}T_N' \left[\sum_{t=1}^{n} \left(\frac{\partial \varepsilon_t'}{\partial \theta} - \frac{\partial \varepsilon_t^{0'}}{\partial \theta} \right) \Sigma_0^{-1} \frac{\partial \varepsilon_t^0}{\partial \theta'} \right] T_N \to_p 0 \tag{67}$$

$$\frac{1}{n^{2(1-\delta)}}T_N' \left[\sum_{t=1}^{n} \left(\frac{\partial \varepsilon_t'}{\partial \theta} - \frac{\partial \varepsilon_t^{0'}}{\partial \theta} \right) \Sigma_0^{-1} \left(\frac{\partial \varepsilon_t}{\partial \theta'} - \frac{\partial \varepsilon_t^0}{\partial \theta'} \right) \right] T_N \to_p 0 \tag{68}$$

$$\frac{1}{n^{1-\delta}}T_S' \left[\sum_{t=1}^{n} \left(\frac{\partial \varepsilon_t'}{\partial \theta} - \frac{\partial \varepsilon_t^{0'}}{\partial \theta} \right) \Sigma_0^{-1} \frac{\partial \varepsilon_t^0}{\partial \theta'} \right] T_S \to_p 0 \tag{69}$$

$$\frac{1}{n^{1-\delta}} \sum_{t=1}^{n} T_S' \left[\left(I \otimes (\varepsilon_t' - \varepsilon_t^{0'}) \Sigma_0^{-1} \right) \left(\frac{\partial^2}{\partial \theta \partial \theta'} \otimes \varepsilon_t^0 \right) \right] T_S \to_p 0 \tag{70}$$

$$\frac{1}{n^{1-\delta}} \sum_{t=1}^{n} T_S' \left[\left(I \otimes \varepsilon_t^{0'} \Sigma_0^{-1} \right) \left(\frac{\partial^2}{\partial \theta \partial \theta'} \otimes (\varepsilon_t - \varepsilon_t^0) \right) \right] T_S \to_p 0 \tag{71}$$

$$T_S' \left[\frac{\partial (vec\, \Sigma_0)'}{\partial \theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) \frac{1}{n^{1-\delta}} \sum_{t=1}^{n} \left(\left(\frac{\partial \varepsilon_t}{\partial \theta'} - \frac{\partial \varepsilon_t^0}{\partial \theta'} \right) \otimes \varepsilon_t^0 \right) \right] T_S \to_p 0 \tag{72}$$

$$T_S' \left[\frac{\partial (vec\, \Sigma_0)'}{\partial \theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) \frac{1}{n^{1-\delta}} \sum_{t=1}^{n} \left(\frac{\partial \varepsilon_t^0}{\partial \theta'} \otimes (\varepsilon_t - \varepsilon_t^0) \right) \right] T_S \to_p 0 \tag{73}$$

and

$$\frac{1}{n^{1-\delta}}T_S' \left[\sum_{t=1}^{n} \left(\frac{\partial \varepsilon_t'}{\partial \theta} - \frac{\partial \varepsilon_t^{0'}}{\partial \theta} \right) \Sigma_0^{-1} \left(\frac{\partial \varepsilon_t}{\partial \theta'} - \frac{\partial \varepsilon_t^0}{\partial \theta'} \right) \right] T_S \to_p 0 \tag{74}$$

$$\frac{1}{n^{1-\delta}} \sum_{t=1}^{n} T_S' \left[\left(I \otimes (\varepsilon_t' - \varepsilon_t^{0'}) \Sigma_0^{-1} \right) \left(\frac{\partial^2}{\partial \theta \partial \theta'} \otimes (\varepsilon_t - \varepsilon_t^0) \right) \right] T_S \to_p 0 \tag{75}$$

$$T_S' \left[\frac{\partial (vec\, \Sigma_0)'}{\partial \theta} \left(\Sigma_0^{-1} \otimes \Sigma_0^{-1} \right) \frac{1}{n^{1-\delta}} \sum_{t=1}^{n} \left(\left(\frac{\partial \varepsilon_t}{\partial \theta'} - \frac{\partial \varepsilon_t^0}{\partial \theta'} \right) \otimes (\varepsilon_t - \varepsilon_t^0) \right) \right] T_S \to_p 0 \tag{76}$$

for all β and Λ satisfying (65) and (66).

Here we only prove that the nonstationary components in (67) – (76) satisfy the required conditions. It is rather obvious that the required conditions hold for the stationary components. In what follows, we use the generic notation $\Delta(n^\kappa x_t)$ to denote the terms which include n^κ (or of a lower order) times (x_t). Clearly, we have

$$\varepsilon_t - \varepsilon_t^0, \quad \frac{\partial \varepsilon_t'}{\partial \theta} - \frac{\partial \varepsilon_t^{0'}}{\partial \theta}, \quad \frac{\partial^2}{\partial \theta \partial \theta'} \otimes \left(\varepsilon_t - \varepsilon_t^0 \right) = \Delta(n^{-1/2+\delta} x_t) + w_t, \tag{77}$$

since both $\beta = \beta_0 + O(n^{-1/2+\delta})$ and $\Lambda = \Lambda_0 + O(n^{-1/2+\delta})$. The results in (67) – (73) follow immediately from (77).

The proofs for (74) – (76) are more involved. For (75) and (76), we need to show

$$x_{t|t-1}^0 = x_t + w_t \tag{78}$$

$$x_{t|t-1} - x_{t|t-1}^0 = -n^{-1/2+\delta} x_t \Delta(n^{-1+2\delta} x_t) + w_t. \tag{79}$$

The result in (78) follows directly from (48). To establish (79), note that

$$x_{t|t-1} - x_{t|t-1}^0 = \frac{\partial x_{t|t-1}^0}{\partial \beta'}(\beta - \beta_0) + \frac{\partial x_{t|t-1}^0}{\partial(vec\,\Lambda)'}(vec\,\Lambda - vec\,\Lambda_0) + \Delta(n^{-1+2\delta}x_t) + w_t$$

$$= -n^{-1/2+\delta}x_t + \Delta(n^{-1+2\delta}x_t) + w_t$$

due to (51). Now it follows immediately from (78) and (79) that

$$\varepsilon_t - \varepsilon_t^0 = -(\beta - \beta_0)x_{t-1}^0 - \beta\left(x_{t|t-1} - x_{t|t-1}^0\right) = \Delta(n^{-1+2\delta}x_t) + w_t, \tag{80}$$

from which, together with (77), we may easily deduce (75) and (76).

Finally, we prove (74). To do so, we first show that

$$\beta_0' \frac{\partial x_{t|t-1}^0}{\partial \beta} = -x_t + w_t \tag{81}$$

$$\beta_0' \left(\frac{\partial x_{t|t-1}}{\partial \beta} - \frac{\partial x_{t|t-1}^0}{\partial \beta}\right) = 2n^{-1/2+\delta}x_t + \Delta(n^{-1+2\delta}x_t) + w_t. \tag{82}$$

The result in (81) follows immediately from (51). To derive (82), we note that

$$\frac{\partial x_{t|t-1}}{\partial \beta} - \frac{\partial x_{t|t-1}^0}{\partial \beta} = \frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial\beta'}(\beta - \beta_0) + \frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial(vec\,\Lambda)'}(vec\,\Lambda - vec\,\Lambda_0)$$

$$+ \Delta(n^{-1+2\delta}x_t) + w_t,$$

and that

$$\beta_0' \frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial\beta'} = -2\beta_0'\frac{\partial x_{t|t-1}^0}{\partial\beta'}(\beta - \beta_0) + w_t \quad \text{and} \quad \beta_0'\frac{\partial^2 x_{t|t-1}^0}{\partial\beta\partial(vec\,\Lambda)'} = w_t, \tag{83}$$

which follow from (51), (61) and (62). Consequently, we have

$$\beta_0' \left(\frac{\partial\varepsilon_t'}{\partial\beta} - \frac{\partial\varepsilon_t^{0\prime}}{\partial\beta}\right) = -\left(x_{t|t-1} - x_{t|t-1}^0\right)\beta_0' - \beta_0'\frac{\partial x_{t|t-1}^0}{\partial\beta}(\beta - \beta_0)'$$

$$- \left(\frac{\partial x_{t|t-1}}{\partial\beta} - \beta_0'\frac{\partial x_{t|t-1}^0}{\partial\beta}\right)\beta'$$

$$= n^{-1+2\delta}x_t + w_t, \tag{84}$$

due to (79), (81) and (82).

Moreover, we have

$$\frac{\partial x_{t|t-1}}{\partial vec\,\Lambda} - \frac{\partial x_{t|t-1}^0}{\partial vec\,\Lambda}$$

$$= \frac{\partial^2 x_{t|t-1}^0}{\partial vec\,\Lambda\partial\beta'}(\beta - \beta_0) + \frac{\partial^2 x_{t|t-1}^0}{\partial vec\,\Lambda\partial(vec\,\Lambda)'}(vec\,\Lambda - vec\,\Lambda_0) + \Delta(n^{-1+2\delta}x_t) + w_t$$

$$= \Delta(n^{-1+2\delta}x_t) + w_t, \tag{85}$$

due to (64) and (83). Consequently, it follows directly from (85) that

$$
\frac{\partial \varepsilon_t'}{\partial vec\,\Lambda} - \frac{\partial \varepsilon_t^{0'}}{\partial vec\,\Lambda} = -\left(\frac{\partial x_{t|t-1}}{\partial vec\,\Lambda} - \frac{\partial x_{t|t-1}^0}{\partial vec\,\Lambda}\right)\beta' - \frac{\partial x_{t|t-1}^0}{\partial vec\,\Lambda}(\beta - \beta_0)'
$$
$$
= \Delta(n^{-1+2\delta}x_t) + w_t. \tag{86}
$$

Now (84) and (86) yield (74), and the proof is complete. $\qquad\square$

Proof of Proposition 4.1 The stated result follows immediately from Lemma 2.2 and the result in (19). Note that we have from Lemma 2.2

$$
\beta_0 x_{t|t-1}^0 = \frac{\beta_0 \beta_0' \Lambda_0^{-1}}{\beta_0' \Lambda_0^{-1} \beta_0}\left[y_t - \sum_{k=0}^{t-1}(1 - 1/\omega_0)^k \triangle y_{t-k}\right]
$$

under the convention (5), and the stated result may now be easily derived using (19) and

$$
\beta_0 x_{t|t-1}^0 = y_t - \varepsilon_t^0,
$$

which is due to the definition of (ε_t^0). $\qquad\square$

References

Anderson, B.D.O. and J.B. Moore (1979). *Optimal Filtering.* Englewood Cliffs, N.J.: Prentice Hall.

Bauwens, L., D. Deprins, and J.-P. Vandeuren (1997). "Modelling Interest Rates with a Cointegrated VAR-GARCH Model," Discussion Paper 9780, CORE.

Campbell, J.Y. and R.J. Shiller (1987). "Cointegration and Tests of Present Value Models," *The Journal of Political Economy*, 95, 1062-88.

Caines, P.E. (1988). *Linear Stochastic Systems.* New York, N.Y.: John Wiley & Sons.

Engle, R.F. and C.W.J. Granger (1987). "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251-76.

Fleming, J., C. Kirby, and B. Ostdiek (2004). "Stochastic Volatility, Trading Volume, and the Daily Flow of Information," *Journal of Business*, forthcoming.

Foster, F.D. and S. Viswanathan (1995). "Can Speculative Trading Explain the Volume-Volatility Relation?" *Journal of Business and Economic Statistics*, 13, 379-96.

Gonzalo, J. and C.W.J. Granger (1995). "Estimation of Common Long-Memory Components in Cointegrated Systems," *Journal of Business & Economic Statistics*, 13, 27-35.

Hafer, R.W., A.M. Kutan, and S. Zhou (1997). "Linkage in EMS Term Structures: Evidence from Common Trend and Transitory Components," *Journal of International Money and Finance*, 16, 595-607.

Hai, W., N.C. Mark and Y. Wu (1997). "Understanding spot and forward exchange rate regressions," *Journal of Applied Econometrics*, 12, 715-734.

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application*. New York, N.Y.: Academic Press.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton, N.J.: Princeton University Press.

Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231-54.

Karpoff, J.M. (1987). "The Relation Between Price Changes and Trading Volume: A Survey," *The Journal of Financial and Quantitative Analysis*, 22, 109-26.

Kim, C.-J. and C.R. Nelson (1999). *State-Space Models with Regime Switching*. Cambridge, M.A.: MIT Press.

Magnus, J.R. and H. Neudecker (1988). *Matrix Differential Calculus*. Chichester, U.K.: John Wiley & Sons.

Miller, J.I. and J.Y. Park (2004). "Nonlinearity, Nonstationarity, and Thick Tails: How They Interact to Generate Persistency in Memory," mimeographed, Department of Economics, Rice University.

Modigliani, F. and R. J. Shiller (1973). "Inflation, Rational Expectations and the Term Structure of Interest Rates," *Economica*, 40, 12-43.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York, N.Y.: John Wiley & Sons.

Park, J.Y. (1990). "Disequilibrium Imputs Analysis," mimeographed, Department of Economics, Cornell University.

Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69, 117-61.

Quah (1992). "The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds," *Econometrica*, 60, 107-18.

Sargent, T.J. (1979). "A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure," *Journal of Monetary Economics*, 5, 133-43.

Stock, J.H. and M.W. Watson (1988). "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097-107.

Stock, J.H. and M.W. Watson (1991). "A probability model of the coincidence economic indicators," in *Leading Economic Indicators: New Approaches and Forecasting Records*, ed. K. Lahiri and G.H. Moore. Cambridge, U.K.: Cambridge University Press, 63-89.

Tauchen, G.E., and M. Pitts (1983). "The Price Variability-Volume Relationship on Speculative Markets," *Econometrica*, 51, 485-506.

Wooldridge, J.M. (1994). "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, Vol. IV, R.F. Engle and D.L. McFadden, ed. Amsterdam: Elsevier, 2639-738.