Using Kalman Filter to Extract and Test for Common Stochastic Trends[1]

Yoosoon Chang[2], Bibo Jiang[3] and Joon Y. Park[4]

Abstract

This paper considers a state space model with integrated latent variables. The model provides an effective framework to specify, test and extract common stochastic trends for a set of integrated time series. The model can be readily estimated by the standard Kalman filter, whose asymptotics are fully developed in the paper. In particular, we establish the consistency and asymptotic mixed normality of the maximum likelihood estimator, and therefore, validate the use of conventional methods of inference for our model. Moreover, we construct a trace statistic, which can be used to determine the number of common stochastic trends in a system of integrated time series. It is shown that the limit distribution of the statistic is standard normal. The test is very simple to implement in practical applications. Our simulation study shows that it behaves quite well in finite samples. For an illustration, we apply our methodology to analyze the common stochastic trend in various default-free interest rates with different maturities.

First Draft: Jan 13, 2006
This Version: September 24, 2008

JEL Classification: C22, C51

Key words and phrases: state space model, Kalman filter, common stochastic trends, maximum likelihood estimation, asymptotic theory.

[1]This version is prepared for the presentation at the Midwest Econometrics Conference, October 12-13, St. Louis. Chang and Park gratefully acknowledge the financial supports from the NSF under Grant No. SES-0453069/0730152 and SES-0518619, respectively.

[2]Department of Economics, Texas A&M University
[3]Department of Economics, Rice University
[4]Department of Economics, Texas A&M University and Sungkyunkwan University

1. Introduction

The Kalman filter is the basic tool used in the standard state space models, which typically deals with dynamic time series models that involve unobserved variables. The applications of Kalman filter can be found in many fields including economics and finance. The asymptotic behavior of maximum likelihood (ML) estimators based on the filter is well known under regular conditions, i.e., linearity, Gaussianity, and stationarity. If linearity is violated, the extended Kalman filter is a standard alternative. Moreover, it is well known that the pseudo-ML estimation performs well when Gaussianity does not hold. To the best of our knowledge, however, no research has been done to investigate the properties of the filter for the case that stationarity is violated. Only very recently, Chang, Miller and Park (2007), which will be referred to as CMP hereafter, pioneered in developing a rigorous asymptotic theory for the state space models with one integrated latent variable.

Since CMP allows for only one integrated latent factor, it does not provide any test for the number of distinct latent factors. This would certainly be an important limitation in practical applications. In many empirical analysis, we see some strong evidence that the common stochastic trends in systems consisting of multiple integrated time series cannot be explained by a single factor. The presence of a single common stochastic trend would imply the presence of as many cointegrating relations as only one net of the number of integrated time series included in the system. This is highly unlikely, especially when the underlying system is large and involves many integrated time series, as is often the case in many practical applications. The reader is referred to, e.g., Kim and Nelson (1999) for various models used in practice and previous empirical researches.

In this paper, we extend CMP to allow for multiple latent factors, and develop a test which can be used to formally test for the number of latent factors. Our framework is completely general, except that we require the latent common factors follow random walks in a strict sense. Within this general framework, we show that the ML estimators of the parameters in the model are consistent and asymptotically mixed normal. The standard inference based on the ML procedure is therefore valid. The convergence rate for the ML estimators is $\sqrt{n}$ as in the standard model. However, we have a faster n rate of convergence for the coefficient of latent common stochastic trends along the cointegration space. This is in parallel to the convergence rates in other types of cointegrated models. We also show that a test based on a trace statistic can be applied in our model to test for the number of common stochastic trends, and that it has asymptotically normal distribution. The test appears to be particularly useful for a large system of integrated time series, which shares a relatively small number of common stochastic trends.

The state space modeling with latent integrated factors provides an alternative way of analyzing cointegrated systems. It is in contrast with the cointegrating regressions considered by, for instance, Phillips (1991) and Park (1992), and also closely related to the error correction formulation used in Johansen (1988, 1991) and Ahn and Reinsel (1990). They all can be used in modeling a system of cointegrated processes which share common stochastic trends. The state space model, however, is unique and distinguishes itself from other competing models in that it may allow for the common stochastic trends to be modeled as pure random walks. As we show in the paper, the state space model with common stochastic trends specified as pure random walks is not compatible with a finite order error correction model (ECM) or vector autoregression (VAR). Therefore, the testing procedure that are based on a finite order ECM or VAR is not applicable for the state space models we consider in the paper.

The rest of the paper is organized as follows. In Section 2, we introduce our state space model and outline the Kalman filtering technique used to estimate the model. Some preliminary results are also included in this section. Section 3 and 4 present the main theoretical findings. In Section 3, we establish the consistency and asymptotic mixed normality of the ML estimators. Theories about the determination of number of common stochastic trends are presented in Section 4. In particular, we introduce and analyze a trace statistic to test for the number of common stochastic trends in Section 4. Section 5 presents some simulation results for the finite sample performance of our estimators and test statistic. An empirical illustration follows in Section 6. Here we use our methodology to investigate a system of interest rates with different maturities. Section 7 concludes the paper. Mathematical proofs are given in Appendix.

2. The Model and Preliminary Results

We consider the state space model given by

$$\begin{aligned} y_t &= A_0 x_t + u_t \\ x_t &= x_{t-1} + v_t \end{aligned} \tag{1}$$

under the following assumptions:

- SSM1: (y_t) is a p-dimensional observable time series,
- SSM2: (x_t) is a q-dimensional vector of latent variable,
- SSM3: A_0 is a $p \times q$ matrix of unknown parameters of rank q, where $q \leq p$,
- SSM4: (u_t) and (v_t) are p- and q-dimensional independent, identically distributed (iid) errors that are normal with mean zero and variance Λ_0 and identity matrix I_q, respectively, and independent of each other, and
- SSM5: x_0 is independent of (u_t) and (v_t), and assumed to be given.

Our model can be used to extract common stochastic trends in time series (y_t). Notice that latent variable (x_t) is defined as a vector of random walks, our model provides a natural way to decompose a cointegrated time series into a permanent and transitory components.

The parameter A_0 and the latent common stochastic trends (x_t) are not globally identified in our model. Obviously, the observable time series (y_t) have the same likelihood under joint transformation

$$A_0 \mapsto A_0 H \qquad \text{and} \qquad x_t \mapsto H' x_t \tag{2}$$

for any q-dimensional orthogonal matrix H. They are identified only up to the equivalence class defined by the transformation in (2). However, both of A_0 and (x_t) are locally identified. Indeed, we may easily see that, for any q-dimensional orthogonal matrix H, $A_0 H$ is not in the neighborhood of any $p \times q$ matrix A_0 of rank q defined by the Euclidean or any equivalent norm in the vector space of $p \times q$ matrices. Of course, (x_t) is identified locally if A_0 is.

In the subsequent development of our theory, we will not impose any extra restrictions to globally identify A_0 and (x_t). This does not seem to be necessary for most potentially useful applications of our model, for which we would be primarily interested in finding out the dimension of common stochastic trends and extracting random walks representing them. All the results in the paper for A_0 and (x_t) should therefore be interpreted as applying to a member of the equivalent class given by the transformation in (2). To ease the exposition of the paper, we first assume that q, i.e., the dimension of (x_t) and rank of A_0, is known to explain how to extract (x_t) and to develop the asymptotic theory for the ML estimation of A_0. A test for q based on a trace statistic will then be introduced and discussed later.

Throughout the paper, we will mainly look at the simple model given by (1). This is purely for expositional convenience. Our subsequent results extend trivially to a more general class of state space models with measurement equation given by

$$y_t = A_0 x_t + \sum_{k=1}^{m} \Pi_k \triangle y_{t-k} + u_t, \tag{3}$$

in place of the one in (1). The inclusion of the lagged differences of (y_t) in (3) only introduces more parameters associated with the observable stationary components of the model, and would not affect our asymptotic theory in any important manner. In our subsequent development of the theory, we will mention explicitly what modifications are needed to accommodate the general model in (3). In all cases, the necessary modifications are obvious and straightforward.

The model defined in (1) can 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

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

and

$$\begin{aligned} \Omega_{t|t-1} &= \Omega_{t-1|t-1} + I_q, \\ \Sigma_{t|t-1} &= A \Omega_{t|t-1} A' + \Lambda. \end{aligned}$$

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

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

The ML estimation method is used in estimating the unknown parameters.

For many uses of Kalman filter, the primary goal is to calculate a forecast and also the conditional variance of the observed time series (y_t) as a function of previous observations. However, in the case that the value of the unobserved variable is of interest for its own sake, smoothing technique is often used, denoted $x_{t|n} = \mathbb{E}(x_t|\mathcal{F}_n)$. The smoothed series $(x_{t|n})$ is estimated conditionally on all of the information in the sample - not just the information up to time t. The following is the key equation for smoothing:

$$x_{t|n} = x_{t|t} + \Omega_{t|t}\Omega_{t+1|t}^{-1}(x_{t+1|n} - x_{t+1|t}).$$

This procedure works recursively by starting from $t = n-1$. Starting value $x_{n|n}$ together with series $(x_{t|t})$, $(x_{t+1|t})$, $(\Omega_{t|t})$ and $(\Omega_{t+1|t})$ are achieved in the estimation procedure. The reader is referred to Hamilton (1994) or Kim and Nelson (1999) for more details of this technique. One thing is clear that smoothing is implemented after the model parameters are estimated, therefore this procedure has no effect on the parameter estimates.

For any given values of A and Λ, there exist steady state values of $\Omega_{t|t-1}$ and $\Sigma_{t|t-1}$, which we denote by Ω and Σ.

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

$$\begin{aligned}\Omega &= \frac{1}{2}(I_q + (I_q + 4(A'\Lambda^{-1}A)^{-1})^{1/2}),\\ \Sigma &= \frac{1}{2}A(I_q + (I_q + 4(A'\Lambda^{-1}A)^{-1})^{1/2})A' + \Lambda\end{aligned}$$

for $p \times q$ matrix A and $p \times p$ matrix Λ.

We will set

$$\Omega_{0|0} = \Omega - I_q \tag{4}$$

for the rest of the paper, so that $\Omega_{t|t-1}$ takes its steady state value Ω for all $t \geq 1$. Of course, $\Sigma_{t|t-1}$ also becomes time invariant and takes its steady state value Σ under this convention.[5] 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. To simplify the exposition, we let $y_0 = 0$.

Lemma 2.2 We have

$$x_{t|t-1} = (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t - \sum_{k=0}^{t-1}(I_q - \Omega^{-1})^k(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\triangle y_{t-k} + (I_q - \Omega^{-1})^{t-1}x_0$$

for all $t \geq 2$.

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

[5] Though we do not show explicitly in the paper, $(\Omega_{t|t-1})$ always converges in our experiments to the steady state value Ω as t increases, regardless of the starting values.

It follows from Lemma 2.1 that $\Omega > I_q$, and therefore, $0 < \Omega^{-1} < I_q$. As a consequence, we have $0 < I_q - \Omega^{-1} < I_q$, and therefore, the magnitude of the term $(I_q - \Omega^{-1})^{t-1}x_0$ is geometrically declining as $t \to \infty$. It implies that the effect of x_0 on $x_{t|t-1}$ dilutes out as $t \to \infty$, as long as x_0 is fixed and finite a.s. Therefore, we may set

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

without affecting our asymptotic results.

Let Ω_0 be the value of Ω defined with the true values A_0 and Λ_0 of A and Λ. If we denote by $x^0_{t|t-1}$ the value of $x_{t|t-1}$ under model (1), we may deduce from Lemma 2.2 and smoothing technique that

Proposition 2.3 We have

$$x^0_{t|t-1} = x_t + \Omega_0^{-1}\sum_{k=1}^{t-1}(I_q - \Omega_0^{-1})^{k-1}(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_{t-k} - \sum_{k=0}^{t-1}(I_q - \Omega_0^{-1})^k v_{t-k}$$

for all $t \geq 2$, and

$$x^0_{t|n} = x^0_{t|t} + \sum_{k=1}^{n-t}(I_q - \Omega_0^{-1})^k \Delta x^0_{t+k|t+k}$$

for all $t \leq n-1$.

Proposition 2.3 implies in particular that

$$x^0_{t|t-1} - x_t = \Omega_0^{-1}a_{t-1} - b_{t-1},$$

where

$$a_{t-1} = \sum_{k=1}^{t-1}(I_q - \Omega_0^{-1})^{k-1}(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_{t-k} \quad \text{and} \quad b_{t-1} = \sum_{k=0}^{t-1}(I_q - \Omega_0^{-1})^k v_{t-k}.$$

Under the assumption that (u_t) and (v_t) are iid random sequences, the time series (a_t) and (b_t) become the stationary first-order VAR processes given by

$$\begin{aligned} a_t &= (I_q - \Omega_0^{-1})a_{t-1} + (A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_t, \\ b_t &= (I_q - \Omega_0^{-1})b_{t-1} + v_t \end{aligned}$$

respectively, since $0 < I_q - \Omega_0^{-1} < I_q$.

Clearly, every component of $(x^0_{t|t-1})$ or $(x^0_{t|n})$ is cointegrated with the corresponding component of (x_t) with unit cointegrating coefficient. The stochastic trends in (x_t) may therefore be identified and represented by those in $(x^0_{t|t-1})$ or $(x^0_{t|n})$. It seems worth noting that the results in Proposition 2.3 do not rely on the iid assumption of (u_t) and (v_t). In particular, our results here imply that we may extract the common stochastic trend in (y_t) using the predicting and smoothing steps of Kalman filter, as long as (u_t) and (v_t) are general stationary processes. Apparently, we need to know the true parameter values to obtain $(x^0_{t|t-1})$ or $(x^0_{t|n})$. The true parameter values are typically unknown and have to be estimated. In most practical applications, we should therefore use the parameter estimates to compute $(x^0_{t|t-1})$ or $(x^0_{t|n})$. It is rather clear that the estimates of $(x^0_{t|t-1})$ and $(x^0_{t|n})$ based on the estimated parameter values are close to $(x^0_{t|t-1})$ and $(x^0_{t|n})$, respectively, if we use the consistent parameter estimates.

Once we obtain $(x^0_{t|t-1})$, we may decompose time series (y_t) into the permanent and transitory (PT) components. If we denote them as (y^P_t) and (y^T_t), respectively, they are given by

$$y^P_t = A_0 x^0_{t|t-1} \quad \text{and} \quad y^T_t = y_t - A_0 x^0_{t|t-1}. \tag{6}$$

The permanent component (y^P_t) is I(1), whereas the transitory component (y^T_t) is I(0). Note that the permanent component (y^P_t) is predictable, while the transitory component (y^T_t) is a martingale difference sequence (mds) and unpredictable.

The Kalman filter has exactly the same prediction and updating steps for the measurement equation (3), if we let

$$y_{t|t-1} = Ax_{t|t-1} + \sum_{k=1}^{m} \Pi_k \triangle y_{t-k}.$$

in place of $y_{t|t-1} = Ax_{t|t-1}$. Therefore, it is clear that Lemma 2.1 and Proposition 2.3 hold for this general model without any modification. Moreover, Lemma 2.2 continues to be valid if we only replace (y_t) with $(y_t - \sum_{k=1}^m \Pi_k \triangle y_{t-k})$. The theory of Kalman filter for the general model is thus followed immediately.

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 under normality. Because the integrated process is involved, the usual asymptotic theory for ML estimation of state space models given by, for instance, Caines (1988), does not apply. CMP develops a general asymptotic theory of ML estimation, which allows for the presence of nonstationary time series. They obtain the asymptotics of ML estimators of the parameters in their model, where the number of latent variable is restricted to one. In this paper, we derive the asymptotic properties of the ML estimators of the parameters in the state space model that has multiple stochastic latent variables. In developing our asymptotic theory, we will frequently refer to the results obtained previously in CMP.

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

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

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

$$\ell_n(\theta) = -\frac{n}{2} \log\det\Sigma - \frac{1}{2} \mathrm{tr}\,\Sigma^{-1} \sum_{\mathrm{t}=1}^{\mathrm{n}} \varepsilon_{\mathrm{t}} \varepsilon_{\mathrm{t}}'$$

ignoring the unimportant constant term. Here, Σ and (ε_t) are in general given as functions of θ. Let $s_n(\theta)$ and $H_n(\theta)$ be the score vector and Hessian matrix, i.e.,

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

After applying some algegra, we may deduce that

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

and

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

as given in CMP. Here and elsewhere in the paper, *vec* A denotes the column vector obtained by stacking the rows of matrix A.

Denoted by $\hat\theta_n$ the maximum likelihood estimator of θ, the true value of which is set as 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

$$s_n(\hat\theta_n) = s_n(\theta_0) + H_n(\theta_n)(\hat\theta_n - \theta_0), \tag{7}$$

where θ_n lies in the line segment connecting $\hat\theta_n$ and θ_0. Assuming that $\hat\theta_n$ is an interior solution, we have $s_n(\hat\theta_n) = 0$ immediately. Therefore, it is now clear from (7) that we may write

$$\nu_n' T^{-1}(\hat\theta_n - \theta_0) = -[\nu_n^{-1}T'H_n(\theta_n)T\nu_n^{-1'}]^{-1}[\nu_n^{-1}T's_n(\theta_0)] \tag{8}$$

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 $n \to \infty$ for some N,

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

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

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

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 (8) and continuous mapping theorem that

$$\nu_n' T^{-1}(\hat{\theta}_n - \theta_0) = -[\nu_n^{-1}T'H_n(\theta_0)T\nu_n^{-1'}]^{-1}[\nu_n^{-1}T's_n(\theta_0)] + o_p(1) \to_d M^{-1}N \qquad (9)$$

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

$$\nu_n^{-1}T' \left(H_n(\theta_n) - H_n(\theta_0)\right) T\nu_n^{-1\prime} \to_p 0 \qquad (10)$$

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

To obtain the limit distribution of $s_n(\theta_0)$, we first 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. Then we have

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

As shown in CMP,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varepsilon_t^0\varepsilon_t^{0'} - \Sigma_0) \to_d \mathbb{N}\left(0, (I+K)(\Sigma_0 \otimes \Sigma_0)\right) \qquad (11)$$

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

$$\sum_{t=1}^{n}(\varepsilon_t^0\varepsilon_t^{0\prime} - \Sigma_0) \;\text{ and }\; \sum_{t=1}^{n}\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\Sigma_0^{-1}\varepsilon_t^0 \;\text{ are asymptotically independent.} \qquad (12)$$

Note in particular that

$$\varepsilon_t^0 = y_t - y_{t|t-1}^0 = A_0(x_t - x_{t|t-1}^0) + u_t,$$

and as a consequence $(\varepsilon_t^0, \mathcal{F}_t)$ is a martingale difference sequence and $((\partial/\partial\theta')\varepsilon_t^0)$ is a predictable sequence with respect to the filtration $(\mathcal{F}_t)$.

If our model were stationary, the limit distribution would therefore be easily derivable from (11), (12) and

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\Sigma_0^{-1}\varepsilon_t^0 \to_d \mathbb{N}\left(0, \operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\Sigma_0^{-1}\frac{\partial\varepsilon_t^0}{\partial\theta'}\right), \tag{13}$$

which can be readily obtained by employing the standard martingale CLT. Of course, asymptotics in (13) does not hold for our nonstationary model with integrated latent variables. As we will show below in Lemma 3.1, the multivariate process $((\partial/\partial\theta')\varepsilon_t^0)$ is given by a mixture of stationary and nonstationary processes. Our subsequent asymptotic analysis will therefore be focused on solving the complexity caused by this mixture of stationarity and nonstationarity.

Now we look at our model more specifically. The parameter θ for our model is given by

$$\theta = ((vecA)', v(\Lambda)')', \tag{14}$$

with the true value $\theta_0 = ((vecA_0)', v(\Lambda_0)')'$. Here and elsewhere in the paper, $v(A)$ denotes the subvector of $vecA$ with all subdiagonal elements of A eliminated. Therefore, $v(A)$ vectorizes only the nonredundant elements of A. We may relate $vec(A)$ and $v(A)$ by $Dv(A) = vecA$, where D is the duplication matrix. See, e.g., Magnus and Neudecker (1988, pp.48-49). The dimension of θ is given by $\kappa = pq + p(p+1)/2$, since in particular there are only $p(p+1)/2$ number of nonredundant elements in Λ.

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

Lemma 3.1 We have

$$\frac{\partial\varepsilon_t^{0\prime}}{\partial vecA} = -\left(I_p - \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\right)\otimes x_t + a_t(u,v) \quad \text{and} \quad \frac{\partial\varepsilon_t^{0\prime}}{\partial vec\Lambda} = b_t(u,v),$$

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

According to Lemma 3.1,

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

is a matrix time series consisting of a mixture of integrated and stationary processes since $a_t(u,v)$ and $b_t(u,v)$ are stationary linear processes driven by (u_t) and (v_t). Notice that

$$P = I_p - \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}A_0' \tag{15}$$

is a $(p-q)$-dimensional (non-orthogonal) projection on the space orthogonal to A_0 along $\Lambda_0^{-1}A_0$. Naturally, we have $A_0'P = 0$. Consequently, $A_0 \otimes I_q$ annihilates the common stochastic trends in $(\partial\varepsilon_t^{0\prime}/\partial vecA)$, and therefore $((A_0 \otimes I_q)'(\partial\varepsilon_t^{0\prime}/\partial vecA))$ becomes stationary. Unlike $(\partial\varepsilon_t^{0\prime}/\partial vecA)$, it is rather clear from Lemma 3.1 that $(\partial\varepsilon_t^{0\prime}/\partial vec\Lambda)$ is entirely stationary.

In order to effectively deal with the singularity of the matrix P in (15), we follow CMP and introduce the necessary rotation. Let B_0 be an $p \times (p-q)$ matrix satisfying the conditions

$$B_0'\Lambda_0^{-1}A_0 = 0 \quad \text{and} \quad B_0'\Lambda_0^{-1}B_0 = I_{p-q}. \tag{16}$$

Note that if $\text{rank}(A_0) = q = p$, such a B_0 does not exist. In the following discussion we will focus on the case where $q < p$. It is easy to deduce that

$$P = I_p - \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}A_0' = \Lambda_0^{-1}B_0B_0', \tag{17}$$

since P is a projection matrix such that $A_0'P = P\Lambda_0^{-1}A_0 = 0$.

Now the κ-dimensional rotation matrix T is defined as

$$T = (T_N,\, T_S), \tag{18}$$

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

$$T_N = \begin{pmatrix} B_0 \otimes I_q \\ 0 \end{pmatrix} \qquad \text{and} \qquad T_S = \begin{pmatrix} A_0(A_0'\Lambda_0^{-1}A_0)^{-1/2} \otimes I_q & 0 \\ 0 & I_{p(p+1)/2} \end{pmatrix}$$

respectively. It follows immediately from Lemma 3.1, (16) and (17) that

$$T_N'\frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} = \left(B_0' \otimes I_q\right)\left(\frac{\partial \varepsilon_t^{0\prime}}{\partial vecA}\right) = -B_0' \otimes x_t + c_t^N(u,v) \tag{19}$$

and

$$T_S'\frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} = \begin{pmatrix} \left[(A_0'\Lambda_0^{-1}A_0)^{-1/2}A_0' \otimes I_q\right]\dfrac{\partial \varepsilon_t^{0\prime}}{\partial vecA} \\ \dfrac{\partial \varepsilon_t^{0\prime}}{\partial v(\Lambda)} \end{pmatrix} = c_t^S(u,v) \tag{20}$$

for some stationary linear processes $c_t^N(u,v)$ and $c_t^S(u,v)$ driven by (u_t) and (v_t). Moreover, we can easily get the inverse of the rotation matrix T as

$$T^{-1} = \begin{pmatrix} B_0'\Lambda_0^{-1} \otimes I_q & 0 \\ (A_0'\Lambda_0^{-1}A_0)^{-1/2}A_0'\Lambda_0^{-1} \otimes I_q & 0 \\ 0 & I_{p(p+1)/2} \end{pmatrix} \tag{21}$$

from our definition of T given above in (18).

Before deriving the main asymptotic results for the ML estimator $\hat{\theta}_n$ of θ, we need to establish two lemmas, which will be presented subsequently. They are straightforward extensions of Lemmas 3.3 and 3.4 in CMP.

Lemma 3.2 If we let

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

for $r \in [0,1]$, then it follows that

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

as $n \to \infty$, 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.

We may readily establish from Lemma 3.2 the joint asymptotics of

$$\frac{1}{n} T_N' \sum_{t=1}^{n} \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1} \varepsilon_t^0 \to_d \int_0^1 V(r) dU(r), \tag{22}$$

and

$$\frac{1}{\sqrt{n}} T_S' \sum_{t=1}^{n} \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1} \varepsilon_t^0 \to_d W, \tag{23}$$

where we denote $W(1)$ simply as W. This convention will be made for the rest of the paper. Because of the independence of V and U, the limiting distribution in (22) is mixed normal. On the other hand, the independence of W and U renders the two limit distributions in (22) and (23) to be independent. Clearly, we have $W =_d \mathbb{N}(0, var(W))$, where

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

Moreover, if we define

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

then it follows that $Z_n \to Z$, where $Z =_d \mathbb{N}(0, var(Z))$ with

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

As noted earlier, Z is also independent of U, V and W introduced in Lemma 3.2.

Now we are ready to derive the limit distribution for the ML estimator $\hat{\theta}_n$ of θ_0 defined in (14). They are given by (9) with the rotation matrix T in (18) and the sequence of normalization matrix

$$\nu_n = \operatorname{diag}(nI_{\kappa_1}, \sqrt{n} I_{\kappa_2}),$$

as we state below as a theorem.

Theorem 3.3 All three conditions in ML1-ML3 are satisfied for our model. 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 before.

Theorem 3.3 is completely analogous to Theorem 3.5 in CMP. In particular, Theorem 3.3 shows that the results in Theorem 3.5 of CMP extends well to the multi-dimensional case, though the proof is much more involved to deal with the multi-dimensionality of the common stochastic trend.

As in CMP, we let

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

and

$$\begin{pmatrix} R \\ S \end{pmatrix} = -[var(W) + var(Z)]^{-1}(W - Z), \tag{24}$$

where R and S are κ^2-, and $p(p+1)/2$-dimensional, respectively. Note that Q has a mixed normal distribution, whereas R and S are jointly normal and independent of Q. Now we may easily deduce from Theorem 3.3 that

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

and

$$n\left(B_0'\Lambda_0^{-1} \otimes I_q\right) vec\hat{A}_n \to_d Q \tag{25}$$

$$\sqrt{n}\left((A_0'\Lambda_0^{-1}A_0)^{-1/2}A_0'\Lambda_0^{-1} \otimes I_q\right)(vec\hat{A}_n - vecA_0) \to_d R, \tag{26}$$

similarly as in CMP. In particular, it follows immediately from (25) and (26) that

$$\sqrt{n}(vec\hat{A}_n - vecA_0) \to_d \left(A_0(A_0'\Lambda_0^{-1}A_0)^{-1/2} \otimes I_q\right) R,$$

which has a degenerate normal distribution, if $q < p$.

From Theorem 3.3 and the subsequent remarks, we know that the ML estimators $\hat{A}_n$ and $\hat{\Lambda}_n$ converge at the standard rate $\sqrt{n}$, and have normal limit distributions. However, in the case where $q < p$ the limit distribution of $\hat{A}_n$ is degenerate. In the direction of $B_0'\Lambda_0^{-1}$, it has a rate of convergence n and a mixed normal limit distribution. The normal and mixed normal asymptotic distributions of ML estimators validate the conventional inference for hypothesis testing in such state space models where multiple integrated latent variables are included.

As discussed in CMP, the asymptotic results for the ML estimator for our model also hold, at least qualitatively, for more general models, such as the type of the models including lagged terms in measurement equations. Even for the case where time series consists not only stochastic integrated trends, but deterministic linear time trend, after some proper rotation of the time series, see, e.g., Park (1992), our asymptotic theories are applicable for the rotated time series. The rotation simply separates out the component dominated by a deterministic linear time trend and the component represented as a purely stochastic integrated process.

4. Determination of Number of Common Trends

In the asymptotic analysis of the ML estimator for our model defined in (1), we assume that the number of common stochastic trends in (y_t) is known to be q. This of course is equivalent to assuming that the number of cointegrating relationships in the p-dimensional time series (y_t) is known to be $p-q$. From our analysis in the previous section, we may indeed readily deduce that

$$B_0'\Lambda_0^{-1}y_t = B_0'\Lambda_0^{-1}u_t \quad \text{and} \quad var(B_0'\Lambda_0^{-1}u_t) = I_{p-q}.$$

It is therefore clearly seen that $\Lambda_0^{-1}B_0$ is the matrix of $p-q$ cointegrating vectors, which yield cointegrating errors with identity covariance matrix. However, the number of common stochastic trends or the cointegrating relationships is typically unknown in empirical studies. In this section, we will develop a test based on a trace statistic for testing the number of common stochastic trends, and explain how we may use the test to determine the dimensionality of the latent integrated processes in our model.

Needless to say, testing for the number of common stochastic trends is equivalent to testing for the number of cointegrating relationships. Therefore, at least conceptually, we may use the existing test such as Johansen (1998, 1991) to determine $p-q$ or q, i.e., the number of cointegrating vectors or the number of common stochastic trends. However, using the methods based on a finite order VAR or ECM as Johansen's approach has two important shortcomings in our context. First, as we will show subsequently, our model cannot be represented as any finite order vector autoregression or error correction model. Any finite order VAR or ECM is therefore inconsistent with our model. Second, our model is potentially more useful for a large system of time series which share a few common stochastic trends. For such systems, VAR or ECM formulations often become too flexible, allowing too many parameters. In particular, it is impossible to use long VAR's or ECM's, trying to fit an infinite order VAR or ECM.

Proposition 4.1 We have

$$\triangle y_t = -B_0(\Lambda_0^{-1}B_0)'y_{t-1} - \sum_{k=1}^{t-1} C_k \triangle y_{t-k} + \varepsilon_t^0, \tag{27}$$

where $C_k = A_0(I_q - \Omega_0^{-1})^k (A_0 \Lambda_0^{-1} A_0)^{-1} A_0' \Lambda_0^{-1}$.

Proposition 4.1 makes clear the difference between our model and the conventional ECM. From (27), we may immediately see that (y_t) is generated as VAR(∞), which in particular implies that the our model is not representable as a finite-order VAR. Moreover, we have rank deficiencies in the short-run coefficients (C_k), as well as in error correction term $B_0(\Lambda_0^{-1}B_0)'$. Note that (C_k) are of rank q and $(\Lambda_0^{-1}B_0)'C_k = 0$ for all $k = 1, 2, \ldots$. In the conventional ECM, there is no such rank restriction imposed on the short-run coefficients. As a consequence, Johansen's approach, based on finite order ECM's, is not applicable in our model. This is also true for the general measurement equation (3). Indeed, it is easy to see that Proposition 4.1 continues to hold in this case only with (y_t) replaced by $(y_t - \sum_{k=1}^{m} \Pi_k \triangle y_{t-k})$. Clearly, in order to test the number of common stochastic trends in our framework, a new testing method is needed.

Now we consider the null hypothesis

$$\mathbb{H}_0 : \operatorname{rank} A_0 = q,$$

which will be tested against

$$\mathbb{H}_1 : \operatorname{rank} A_0 > q.$$

To determine the number of common trends, we test $\mathbb{H}_0$ sequentially starting from $q = 1$. If $\mathbb{H}_0$ is not rejected for $q = 1$, then we conclude that there exists a single common trend. If, on the other hand, $\mathbb{H}_0$ is rejected in favor of $q > 1$, then we test for the null hypothesis with $q = 2$. We may continue this procedure until $\mathbb{H}_0$ is not rejected. The number of common trends is then determined as the value of q, for which $\mathbb{H}_0$ is not rejected for the first time.

Our procedure here is in contrast with that of Johansen, which tests for the number of cointegrating relationships in a reversed order. In his approach, the null hypothesis of no cointegration (i.e., $p = q$) is first tested against one cointegrating relationship (i.e., $q = p - 1$), which will then be tested again two cointegrating relationships (i.e., $q = p - 2$) if the null hypothesis of no cointegration is rejected in favor of the alternative hypothesis. To determine the number of cointegrating relationships, we must continue the test until the null hypothesis is not rejected. This is the same as our procedure. In our framework, which seems more useful to analyze relatively large dimensional systems sharing a few common stochastic trends, we believe that our approach is more desirable. As q gets large, the estimation of our model becomes computationally quite burdensome.

The test will be based on the statistic defined as

$$\tau_n = (1/\hat{\omega}_n)\sqrt{n} \left[\operatorname{tr} \hat{B}_n' \hat{\Lambda}_n^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} y_t y_t' \right) \hat{\Lambda}_n^{-1} \hat{B}_n - (p - q) \right] \tag{28}$$

with

$$\hat{\omega}_n^2 = 2(p-q) - (vec\, I_{p-q})(\hat{B}_n' \hat{\Lambda}_n^{-1} \otimes \hat{B}_n' \hat{\Lambda}_n^{-1}) a\hat{v}ar(\hat{\Lambda}_n)(\hat{\Lambda}_n^{-1} \hat{B}_n \otimes \hat{\Lambda}_n^{-1} \hat{B}_n)(vec\, I_{p-q}),$$

where $a\hat{v}ar(\hat{\Lambda}_n)$ is a consistent estimate of the asymptotic variance of $\hat{\Lambda}_n$.

Under the null hypothesis $\mathbb{H}_0$, we have

$$\hat{B}_n'\hat{\Lambda}_n^{-1}\left(\frac{1}{n}\sum_{t=1}^n y_t y_t'\right)\hat{\Lambda}_n^{-1}\hat{B}_n \approx \hat{B}_n'\hat{\Lambda}_n^{-1}\left(\frac{1}{n}\sum_{t=1}^n u_t u_t'\right)\hat{\Lambda}_n^{-1}\hat{B}_n$$
$$\approx B_0\Lambda_0^{-1}\Lambda_0\Lambda_0^{-1}B_0 = I_{p-q},$$

due in particular to the fact that $B_0'\Lambda_0^{-1}A_0 = 0$. On the other hand, we have under the alternative hypothesis $\mathbb{H}_1$

$$\hat{B}_n'\hat{\Lambda}_n^{-1}\left(\frac{1}{n}\sum_{t=1}^n y_t y_t'\right)\hat{\Lambda}_n^{-1}\hat{B}_n \approx n\hat{B}_n'\hat{\Lambda}_n^{-1}A_0\left(\frac{1}{n^2}\sum_{t=1}^n x_t x_t'\right)A_0'\hat{\Lambda}_n^{-1}\hat{B}_n \to_p \infty,$$

since $\hat{B}_n'\hat{\Lambda}_n^{-1}A_0$ does not vanish as n gets large. Therefore, consistent is the test which rejects the null hypothesis $\mathbb{H}_0$ in favor of the alternative hypothesis $\mathbb{H}_1$ when the value of the statistic τ_n is large.

The following theorem establishes the limit null distribution of τ_n.

Theorem 4.2 Under the null hypothesis, we have

$$\tau_n \to_d \mathbb{N}(0,1)$$

as $n \to \infty$.

As shown in Theorem 4.2, the limit distribution of τ_n is standard normal. It does not include any nuisance parameters. Therefore, the implementation of the test is truly simple.

For the test statistic τ_n, we may use any consistent estimate of the asymptotic variance of $\hat{\Lambda}_n$. The most natural choice would be to use the negative hessian matrix. To be more precise, we define

$$\frac{\hat{T}_S'H_n(\hat{\theta}_n)\hat{T}_S}{n} = \bar{H}_S = \begin{pmatrix} \bar{H}_{11} & \bar{H}_{12} \\ \bar{H}_{21} & \bar{H}_{22} \end{pmatrix},$$

where $\hat{T}_S$ is the matrix defined similarly as T_S with A_0 replaced by $\hat{A}_n$ and the partition of $\bar{H}_S$ is made conformably with $\hat{T}_S$. Then a consistent estimate of the asymptotic variance of $v(\hat{\Lambda}_n)$ is given by

$$-\bar{H}_{22\cdot 1}^{-1} = -(\bar{H}_{22} - \bar{H}_{21}\bar{H}_{11}^{-1}\bar{H}_{12})^{-1}.$$

The corresponding asymptotic variance of $\hat{\Lambda}_n$ is given by

$$-D\bar{H}_{22\cdot 1}^{-1}D',$$

where D is the duplication matrix.

A similar approach is possible when the measurement equation is given more generally as in (3). In this case, we may simply modify the test statistic τ_n by replacing (y_t) with

$$y_t - \sum_{k=1}^m \hat{\Pi}_k \triangle y_{t-k},$$

where $(\hat{\Pi}_k)$ is the ML estimate of (Π_k) for $k = 1, \ldots, m$. This is shown in the proof of Theorem 4.2. Moreover, it is clear that $(\hat{\Pi}_k)$ is asymptotically independent of $(\hat{\Lambda}_n)$. Therefore, we may just ignore the blocks of the hessian matrix corresponding to $(\hat{\Pi}_k)$ and proceed as above, when we compute the consistent estimate of the asymptotic variance of $vec\,\hat{\Lambda}_n$.

5. Simulations

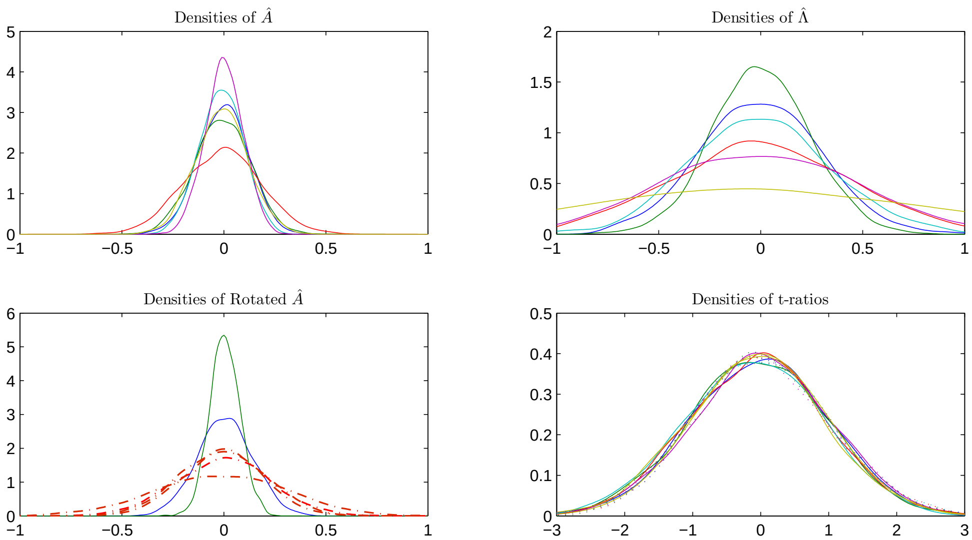

Figure 1: Densities of MLE and t-ratios, n=500

In this section, we perform a set of simulations to investigate the finite sample properties of the ML estimates. We look at a specific model of 3 observable time series with 2 common stochastic trends. In order to satisfy Assumption SSM4, the error terms (u_t) and (v_t) are generated as follows:

$$u_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} + \varepsilon_{3t} \\ \varepsilon_{2t} \end{pmatrix} \quad \text{and} \quad v_t = \begin{pmatrix} \varepsilon_{4t} \\ \varepsilon_{5t} \end{pmatrix},$$

where $(\varepsilon_{1t}) - (\varepsilon_{5t})$ are independent and randomly drawn from $N(0,1)$. The independence between (u_t) and (v_t) therefore follows and the covariance matrix of (v_t) is an identity matrix. Λ_0, the covariance matrix of (u_t), can be easily derived as well. We present Λ_0 and the arbitrarily selected true value of A_0 as follows:

$$A_0 = \begin{pmatrix} 1 & 2 \\ 2 & 0 \\ 3 & 2 \end{pmatrix} \quad \text{and} \quad \Lambda_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 4 \end{pmatrix}.$$

The initial value of the state variable x_0 is set to be 0. In this way, Assumption SSM5 is also satisfied and Theorem 3.3 is readily applicable to this model. According to the theorem, the ML estimators $\hat{A}$ and $\hat{\Lambda}$ converge to Gaussian distributions at the standard rate $\sqrt{n}$. However, the asymptotic distribution of $\hat{A}$ is degenerated normal. $\hat{A}$ converges to a mixed normal distribution at a faster rate n in the direction of $B_0'\Lambda_0^{-1}$, the cointegrating space of observable time series. However, in the directions orthogonal to the cointegrating space, it converges at standard $\sqrt{n}$ rate and has an asymptotically normal distribution.

In the simulation, the samples of size 500 are drawn 2000 times to estimate the ML estimators. The t-statistics based on these estimators are also derived. In estimating Λ_0, we estimate the cholesky triangle of Λ_0 instead. In that way, we only estimate the nonredundant parameters and also ensure the estimated covariance matrix to be positive definite. To choose the matrix B_0 in the rotation matrix, we first regress a randomly picked 3 dimensional vector on $\Lambda_0^{-1}A_0$, and then normalize the residual, such that the normalized residual e satisfies the condition $e'\Lambda_0^{-1}e = 1$. We take e as our B_0. It is clear that B_0 satisfies the constraints specified in (16). The simulation results are summarized in Figure 1. The distributions of the ML estimators are centered with their true values.

The finite sample behavior of the ML estimators are as expected. The distributions of $\hat{A}$ and $\hat{\Lambda}$ are symmetric and well centered as predicted by their asymptotic theories. The left bottom of 1 presents the distributions of rotated $\hat{A}$. The solid curves represent the distributions of $\hat{A}$ on the direction of $B_0'\Lambda_0^{-1}$ and the rest of the curves represent the distributions of $\hat{A}$ on the directions orthogonal to $B_0'\Lambda_0^{-1}$. As expected, the solid curves are steeper since the rotated $\hat{A}$ converges at a faster rate on the directions defined by the cointegrating space. Distributions of t-ratios are presented at the right bottom of 1. In total 13 curves are involved, 12 are the t-ratios of the ML estimators of the unknown parameters and one represents standard normal distribution. The thirteen curves seem to be very much overlapped, which is consistent with the asymptotic theory.

6. An Empirical Illustration

The topic of determining the relationship among the yields on default-free securities that differ only in their terms to maturity has long been a topic of concern for economists. Most researches are conducted in the framework of structural models. Structural models focus on explaining and testing the term premium. Depending on the assumption of the driving diffusion processes, structural models can be divided into one-factor and multi-factor models. One-factor models, like one-factor time-homogeneous models of Vasicek, Cox-Ingersoll-Ross, Dothan, and the Exponential Vasicek model, model the instantaneous spot interest rate via one driving diffusion process. Multi-factor models, such as the two-factor model in Hall and White (1994), assume that interest rates are affected by multiple correlated diffusion processes. According to Jamshidian and Zhu (1997), multi-factor models can explain much more variations in historical yield curves than one-factor models do. However, how many factors should be included to increase the explanatory power of models without causing the over-fitting problem is not clear. This problem becomes even less clear when we consider a group of interest rates altogether in one model. We expect some trends are common to different interest rates since under default-free assumption, the purchase of a long-term asset is equivalent to that of a sequence of short-term assets. Therefore, the number of trends must be much smaller than the number of the interest rates under consideration. In this section, we focus on testing and extracting the common unobservable stochastic trends of 9 default-free interest rates. Our analysis is based on the state space model described in Section 2.

The interest rates used here are Secondary Market Rate for Treasury Bills with maturities 3 and 6 months and Treasury Constant Maturity Rate of Treasury Bonds with maturities 1, 2, 3, 5, 7, 10, 30 years. The data is obtained from Federal Reserve Bank at St. Louis, and the selection of data is based on availability. The data is monthly and ranging from February 1st, 1977 to February 1st, 2002.

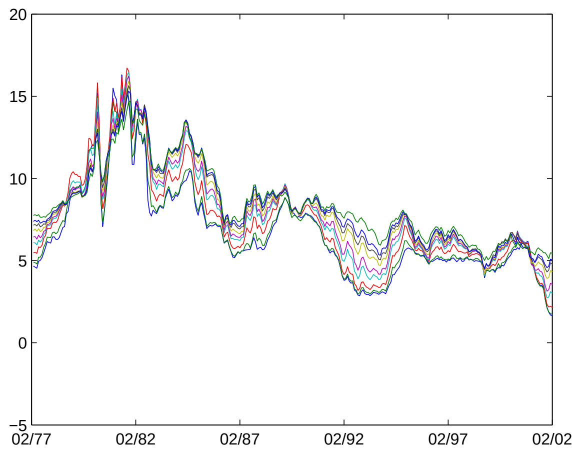

Figure 2: Default-Free Interest Rates with Different Maturities

A salient feature of interest rates is strong persistence. The persistence can be easily seen from Figure 2. Although strong persistence does not necessarily imply a presence of unit root, it is true that in many empirical studies, unit root tests fail to reject the null hypothesis that interest rates have a unit root. For example Nelson and Plosser (1982) investigate a set of macroeconomic variables by using Dickey-Fuller type tests that are developed in Dickey and Fuller (1979, 1981), and conclude that many macroeconomic time series, including bond yield, are better characterized as having a random walk component than as stationary with drift or trend stationary. In this paper, we revisit this problem by conducting the augmented Dickey Fuller test on the 9 interest rate series. The number of lags in the autoregression function is selected by Akaike Information Criteria and Bayesian Information Criteria. The autoregression has a constant term but not a time trend since no economic theory suggests that nominal interest rates should exhibit a deterministic time trend and if an interest were to be described by a stationary process, surely it would have a positive mean. The augmented Dickey Fuller tests on the nine interest rates fail to reject unit root hypotheses at 95% significance level. This gives us a reason to believe that interest rates are driven by some nonstationary stochastic processes. The details of the testing results are available upon request.

We use the model defined in (1) to extract the common trends of the 9 interest rates. Based on the asymptotic analysis in the previous sections, the model can be estimated by the ordinary Kalman filter and the ML estimates of model parameters are consistent and asymptotically Gaussian. Moreover, if the number of trends is known, the extracted trends only differ from the true trends by a stationary component. We apply the sequential test introduced in Section 4 in testing the number of common trends.

$$
\begin{aligned}
H_0 &: k = q \\
H_1 &: k > q
\end{aligned}
$$

where $q < 9$ is the number of trends under the null.

The test statistic is given in (28). Before calculating the test statistic under the null hypothesis, we first need to estimate the model with the number of trends assumed in the null. We shall always start from $q = 1$ and stop when we fail to reject the null. The likelihood function is formalized in the standard way as for general stationary state space models. The initial covariance matrix of the state variable is set to be $\Omega - I_q$, where Ω is the steady state value derived in Lemma 2.1:

$$\Omega = \frac{1}{2}(I_q + (I_q + 4(A'\Lambda^{-1}A)^{-1})^{1/2}).$$

The selection of the initial value of the state variable is not clear. Although it should not matter asymptotically as discussed in Lemma 2.2, it might still play a role in finite sample estimation. Kim and Nelson (1999) suggest dropping some of the initial observations when evaluating the likelihood function. This approach causes information loss as we can see. Another approach is to treat $x_{0|0}$ as another model parameter and estimate. This approach is applaudable when the number of trends is not big, i.e. the dimension of $x_{0|0}$ is low. In this application, we use the second approach. In estimating the covariance matrix of the error term in the measurement equation, in order to ensure that the estimated covariance matrix is positive definite and to estimate only non-redundant parameters we estimate its Cholesky decomposition instead.

We report the model estimators for $q = 1$ in Table 1. In order to get $\hat{B}_n$ in the trace statistic, we first regress a randomly generated 9×8 matrix on $\hat{\Lambda}^{-1}\hat{A}$, and then normalize the obtained error term so that for the normalized error e, $e'\hat{\Lambda}^{-1}e = 1$. It is clear that the normalized error satisfies the conditions in (16) and therefore can be used as $\hat{B}_n$. The trace statistic and its variance are now ready to be calculated. In our case,

$$\tau_n = -1.7019 \times 10^{-4}.$$

According to the test statistic, we fail to reject the null hypothesis of $q = 1$, i.e., the interest rates under consideration are driven by a single stochastic process.

To verify the test result, we also estimate the model under the assumption of $q = 2$, and extract the corresponding common trends. The maximum likelihood estimates for $q = 2$

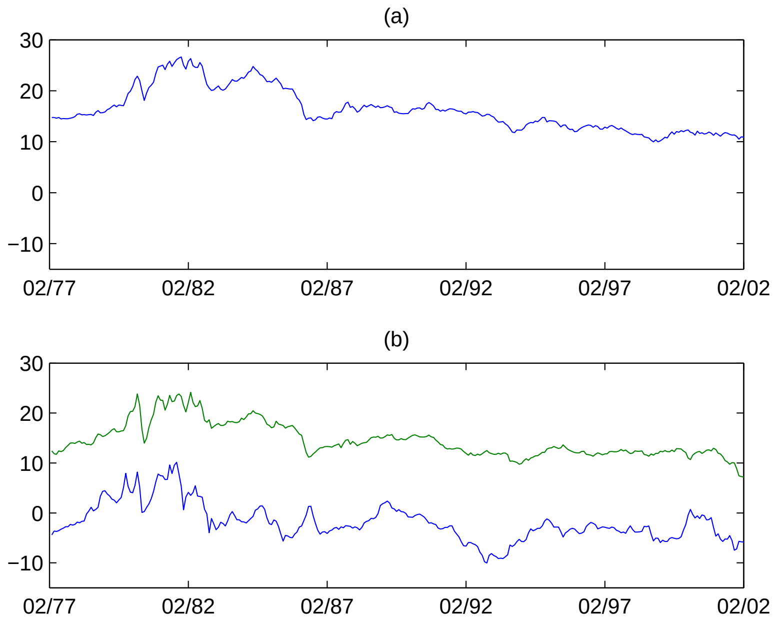

Figure 3: Extracted Common Trends

are provided in Table 2, and we present the extracted trends from the two models in Figure 3. The common trend extracted from the one-trend is displayed in 3 (a) and the trends extracted from the two-trend model is displayed in 3 (b). From Figure 3 (b), we can see that the two trends extracted with the two-trend model are almost identical to each other in terms of variations over time. Moreover, they are similar to the trend extracted by the one-trend model. This finding also implies that one trend is probably enough in describing the long-term behavior of the default-free interest rates.

We can also consider this problem by studying the estimated errors in the measurement equation. As discussed earlier, the model (1) provides a natural way for decomposing a nonstationary time series into permanent and transitory components, which can be written as:

$$
\begin{aligned}
i_t^P &= \hat{A}x_{t|t-1} \\
i_t^T &= i_t - \hat{A}x_{t|t-1},
\end{aligned}
$$

where i_t^P and i_t^T represent permanent and transitory components of interest rate i_t, respectively. The transitory component is also called the prediction error. If x_t catches all the trends indeed, the transitory component should be stationary, since the extracted trends $x_{t|t-1}$ and real trends x_t only differ by a stationary component according to Proposition 2.3. On the other hand, if the number of underlying trends were bigger than what assumed in the model, at least some series in the transitory component would show nonstationarity. This implication gives us another way to verify the test result by testing the stationarity

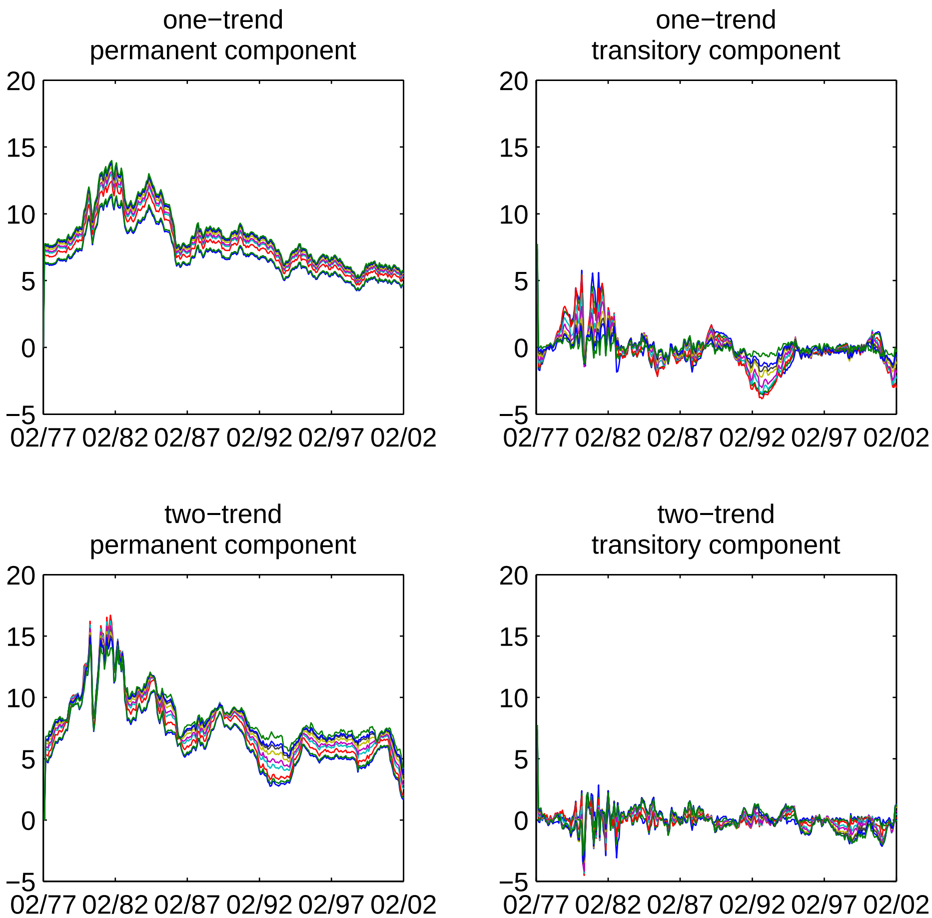

Figure 4: Permanent and Transitory Components from Models with $q = 1,\ 2$

of the transitory components. As a comparison, we present the permanent and transitory components from the one-trend model and the two-trend model in Figure 4. The transitory components from both models look quite stationary although they still show some persistence. The persistence in the transitory components is expectable since the true trends x_t differ from the extracted trends $x_{t|t-1}$ by a summation of two AR(1) processesas indicated in Proposition 2.3. We conduct the ADF tests with no drift and no time trend on the two transitory components, and we are able to reject at 95% level the unit root hypothesis on both series. However, as we can anticipate, the two-trend model catches more variations of the interest rates and the corresponding transitory component shows much less persistence than that from the one-trend model.

Based on the testing result, one common trend is sufficient in describing the long-term behavior of the interest rates with different maturities. Figure 5 shows the relationship between the extracted common trend and each interest rate. For a better comparison,

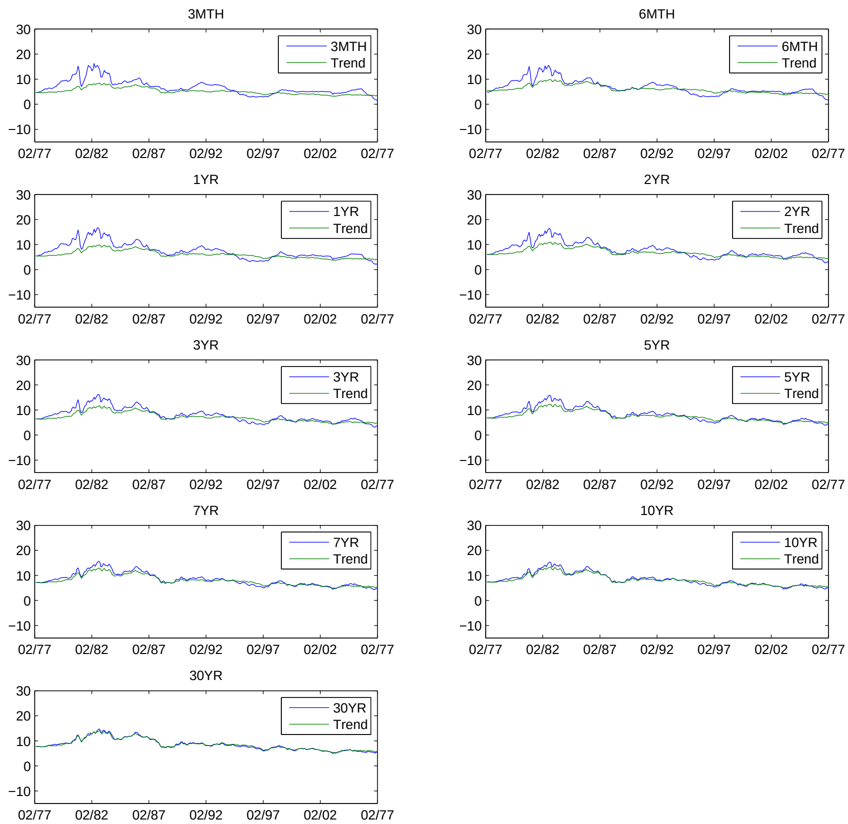

Figure 5: The Extracted Common Trend and Interest Rates

we rescale the trend by multiplying an array of constants so that both series start at the same value. The rescaling should not matter since the variation not the magnitude of the common trend is important. Actually, in model (1), we normalize the trend by assuming the covariance matrix of the innovation being an identity matrix. From Figure 5, we can see that the behavior of the common trend is very close to that of 30 year treasury bond. It tells us that the return on the long-term bonds are mostly determined by the trend which is generated by persistent shocks. Shocks with transitory effects have very limited effect on long-term assets. On the other hand, we can see that the short-term treasury bills vary more around the common trend, although they do not go far from it. It implies that short-term assets are affected by both the long-term trend and the transitory shocks. These findings are consistent with our intuition. Currently, the Federal Reserve Banks are trying to influence the rate of return on assets with long maturities by controlling short-term interest rates. The long-term interest rates are more important in the sense that most consumers and firms are borrowing and lending based on relatively long term rates. Figure 5 tells us that in order to effectively offset undesirable changes in long-term interest rates, government should respond with policies that have persistent effect instead of transitory effect. Policies, such as tax policy, that produce persistent shifts in interest rates are most likely to be more effective than monetary policies which are believed to have temporary effects.

7. Conclusion

In this paper, we consider a state-space model with multiple integrated latent factors. The model provides a new framework, within which we may effectively specify and analyze common stochastic trends in a cointegrated system as latent factors. The standard Kalman filter is used to estimate the model and to extract the common stochastic trends. We establish the consistency and asymptotic normality of the ML estimates of the model parameters, and therefore validate the conventional method of inference based on ML estimators for this class of models. In particular, the ML estimator for the factor loading coefficient has a mixed rate of convergence. It converges at n rate in the direction of cointegration, while the overall convergence rate is $\sqrt{n}$ as in the standard stationary model. Its asymptotic distribution is therefore degenerate if normalized with the conventional $\sqrt{n}$ rate.

In order to determine the number of common stochastic trends, or equivalently the number of cointegrating relationships, we derive a new test. The existing methods relying on the ECM such as Johansen's test are not applicable to our model, since it cannot be represented as a finite order VAR. Our test is based on a trace statistic. The trace statistic is shown to be normal distributed, and therefore, it is very simple to implement in practical applications. Moreover, the statistic diverges whenever the model has more number of common stochastic trends than is assumed. If applied sequentially, we may find the number of stochastic trends in the model. Our method is particularly appropriate to deal with a large dimensional system sharing a few common stochastic trends. The simulation reported in the paper shows that the new test performs reasonably well in finite samples. For the empirical illustration, we analyze the system of interest rates with different maturities, and obtain a strong evidence that they share a single common stochastic trend.

References

Chang, Y., J.I. Miller and J.Y. Park (2007). "Extracting a common stochastic trend: Theory with some applications," forthcoming in *Journal of Econometrics*.

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.

Johansen, S. (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica*, 59: 1551-80.

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.

Park, J.Y. (1992). "Canonical cointegrating regressions," *Econometrica*, 60: 119-143.

Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear regressions with integrated time series," *Econometrica*, 69: 117-161.

Phillips, P.C.B. (1991). "Optimal inference in cointegrated systems," *Econometrica*, 59: 283-306.

Stuart, A. and J.K. Ord (1991). *Kendall's Advanced Theory of Statistics,* Vol. 2, 5th ed, 605-870. New York, N.Y.: Oxford University Press.

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.

Appendix A: Mathematical Proofs

Proof of Lemma 2.1 According to the prediction and updating steps, we have

$$\Omega_{t+1|t} - I_q = \Omega_{t|t-1} - \Omega_{t|t-1}A'(A\Omega_{t|t-1}A' + \Lambda)^{-1}A\Omega_{t|t-1}. \tag{29}$$

In order to show that the steady state value of Ω uniquely exists, we consider the matrix equation given by

$$X - I_q = X - XA'(AXA' + \Lambda)^{-1}AX. \tag{30}$$

Here the unknown matrix X is a $q \times q$ positive definite matrix. We need to check if there exists one and only one positive definite matrix X satisfying the matrix equation.

From function (30), we have

$$XA'(AXA' + \Lambda)^{-1}AX = I_q. \tag{31}$$

Moreover, using the standard rules for matrix algebra, we may easily deduce that

$$\begin{aligned}(AXA' + \Lambda)^{-1} &= \Lambda^{-1} - \Lambda^{-1}AX(X + XA'\Lambda^{-1}AX)^{-1}XA'\Lambda^{-1} \\ &= \Lambda^{-1} - \Lambda^{-1}A(I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1},\end{aligned} \tag{32}$$

and therefore,

$$\begin{aligned}&XA'(AXA' + \Lambda)^{-1}AX \\ &\quad = XA'\Lambda^{-1}AX - XA'\Lambda^{-1}A(I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX \\ &\quad = XA'\Lambda^{-1}AX - (I_q + XA'\Lambda^{-1}A)(I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX \\ &\qquad + (I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX \\ &\quad = (I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX.\end{aligned} \tag{33}$$

Consequently, we have

$$(I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX = I_q,$$

i.e.,

$$X(A'\Lambda^{-1}A)X = I_q + XA'\Lambda^{-1}A, \tag{34}$$

due to (31) and (33).

Now it easy to check that

$$\begin{aligned} X_1 &= \frac{1}{2}(A'\Lambda^{-1}A)^{-1}\left((A'\Lambda^{-1}A) + \left[(A'\Lambda^{-1}A)^2 + 4(A'\Lambda^{-1}A)\right]^{1/2}\right) \\ &= \frac{1}{2}\left(I_q + \left[I_q + 4(A'\Lambda^{-1}A)^{-1}\right]^{1/2}\right) \\ X_2 &= \frac{1}{2}\left(I_q - \left[I_q + 4(A'\Lambda^{-1}A)^{-1}\right]^{1/2}\right) \end{aligned}$$

are the two solutions for X in matrix equation (34). Because X_2 is negative definite, it does not satisfy the properties of X. Therefore, X_1 which is positive definite is the only solution for our problem, i.e, the steady state value of Ω uniquely exists. The steady state value for Σ follows immediately with $\Sigma = A\Omega A' + \Lambda$. □

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

$$\begin{aligned} x_{t+1|t} &= x_{t|t-1} + \Omega A'\Sigma^{-1}(y_t - y_{t|t-1}) \\ &= x_{t|t-1} + \Omega A'\Sigma^{-1}(y_t - Ax_{t|t-1}) \\ &= (I_q - \Omega A'\Sigma^{-1}A)x_{t|t-1} + \Omega A'\Sigma^{-1}y_t \\ &= (I_q - \Omega A'\Sigma^{-1}A)x_{t|t-1} + \Omega A'\Sigma^{-1}y_t \end{aligned} \tag{35}$$

with the steady state values Ω and Σ. However, it follows from (29) that

$$\Omega A'\Sigma^{-1}A\Omega = I_q,$$

i.e.,

$$\Omega A'\Sigma^{-1}A = \Omega^{-1}. \tag{36}$$

We may also deduce from (32) that

$$\Sigma^{-1} = (A\Omega A' + \Lambda)^{-1} = \Lambda^{-1} - \Lambda^{-1}A(I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}, \tag{37}$$

which yields

$$\begin{aligned} \Omega A'\Sigma^{-1}A &= \Omega A'\Lambda^{-1}A - \Omega A'\Lambda^{-1}A(I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}A \\ &= \Omega A'\Lambda^{-1}A\left[I_q - (I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}A\right]. \end{aligned} \tag{38}$$

Therefore, it follows from (36) and (38) that

$$I_q - (I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}A = (\Omega A'\Lambda^{-1}A)^{-1}\Omega^{-1}. \tag{39}$$

Furthermore, we have

$$\begin{aligned}
\Sigma^{-1}A &= \Lambda^{-1}A - \Lambda^{-1}A(I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}A \\
&= \Lambda^{-1}A\left[I_q - (I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}A\right] \\
&= \Lambda^{-1}A\left(\Omega A'\Lambda^{-1}A\right)^{-1}\Omega^{-1}
\end{aligned}$$

and

$$\Omega A'\Sigma^{-1} = \Omega\left[\Lambda^{-1}A(\Omega A'\Lambda^{-1}A)^{-1}\Omega^{-1}\right]' = \Omega^{-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}, \tag{40}$$

due to (37) and (39).

Now we have from (35), (36) and (40) that

$$x_{t+1|t} = (I_q - \Omega^{-1})x_{t|t-1} + \Omega^{-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t,$$

and consequently,

$$x_{t|t-1} = \sum_{k=1}^{t-1}(I_q - \Omega^{-1})^{k-1}\Omega^{-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_{t-k} + (I_q - \Omega^{-1})^{t-1}x_{1|0}. \tag{41}$$

Moreover,

$$\begin{aligned}
&\sum_{k=1}^{t-1}(I_q - \Omega^{-1})^{k-1}\Omega^{-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_{t-k} \\
&\quad = \sum_{k=1}^{t-1}(I_q - \Omega^{-1})^{k-1}\left[I_q - (I_q - \Omega^{-1})\right](A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_{t-k} \\
&\quad = (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t - \sum_{k=0}^{t-2}(I_q - \Omega^{-1})^{k}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\triangle y_{t-k} \\
&\qquad - (I_q - \Omega^{-1})^{t-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_1 \\
&\quad = (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t - \sum_{k=0}^{t-1}(I_q - \Omega^{-1})^{k}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\triangle y_{t-k}.
\end{aligned} \tag{42}$$

The stated result now follows directly from (41) and (42). Note that $x_{1|0} = x_{0|0} = x_0$ and $y_0 = 0$. The proof is therefore complete. □

Proof of Proposition 2.3 For the proof of Proposition 2.3, the readers are referred to the proof of Proposition 2.4 in CMP for the details. In order to fit our model, we only need to replace ω_0 with Ω_0 and $1/\omega_0$ with Ω_0^{-1}. Now let us look at the proof of Proposition 2.3. It follows from Lemma 2.2 that

$$\begin{aligned} x^0_{t|t-1} =&(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}(A_0x_t+u_t)\\ &-\sum_{k=0}^{t-1}(I_q-\Omega_0^{-1})^k(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}(A_0v_{t-k}+(u_{t-k}-u_{t-k-1}))\\ =&x_t+(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_t-\sum_{k=0}^{t-1}(I_q-\Omega_0^{-1})^k(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}(u_{t-k}-u_{t-k-1})\\ &-\sum_{k=0}^{t-1}(I_q-\Omega_0^{-1})^k v_{t-k}. \end{aligned} \tag{43}$$

However, we may easily deduce that

$$\begin{aligned} &\sum_{k=0}^{t-1}(I_q-\Omega_0^{-1})^k(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}(u_{t-k}-u_{t-k-1})\\ &\quad=(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_t-\Omega_0^{-1}\sum_{k=1}^{t-1}(I_q-\Omega_0^{-1})^k(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_{t-k}. \end{aligned} \tag{44}$$

The stated result now follows immediately from (43) and (44). □

Proof of Lemma 3.1 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. It follows from Lemma 2.2 that

$$x_{t|t-1}=(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t+w_t, \tag{45}$$

under our convention here. We define the commutation matrix K_{ab} by

$$K_{ab}vecA=vecA' \tag{46}$$

for $a\times b$ matrix A. Note that we define vec to be the operator stacking rows, not the columns, of a matrix. Therefore, if we let $\overline{vec}$ be the operator stacking columns of a matrix, and let $\overline{K}_{ab}$ be the commutation matrix such that $\overline{K}_{ab}\,\overline{vec}A=\overline{vec}A'$, then we have $K_{ab}=\overline{K}_{ba}$. The readers are referred to Magnus and Neudecker (1988) for more on the commutation matrix.

Since

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

and

$$vec\,Ax_{t|t-1}=(I_p\otimes x_{t|t-1}')vecA,$$

we may easily deduce that

$$\frac{\partial\varepsilon_t}{\partial(vecA)'}=-A\frac{\partial x_{t|t-1}}{\partial(vecA)'}-I_p\otimes x_{t|t-1}'. \tag{47}$$

Moreover, it follows that

$$\frac{\partial \varepsilon_t}{\partial (vec\Lambda)'} = -A\frac{\partial x_{t|t-1}}{\partial (vec\Lambda)'}. \tag{48}$$

The partial derivatives of ε_t with respect to $vecA$ and $vec\Lambda$ may therefore be easily obtained from (47) and (48), once we find the partial derivatives of $x_{t|t-1}$ with respect to $vecA$ and $vec\Lambda$ in (45).

Firstly, in order to get the partial derivative of $x_{t|t-1}$ with respect to A, we assume Λ to be fixed. Then it follows from (45) that

$$\begin{aligned} dx_{t|t-1} =& -(A'\Lambda^{-1}A)^{-1}(dA'\Lambda^{-1}A + A'\Lambda^{-1}dA)(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t \\ &+ (A'\Lambda^{-1}A)^{-1}dA'\Lambda^{-1}y_t + w_t \\ =& -(A'\Lambda^{-1}A)^{-1}dA'(\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t) \\ &- ((A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1})dA((A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t) \\ &+ (A'\Lambda^{-1}A)^{-1}dA'(\Lambda^{-1}y_t) + w_t, \end{aligned}$$

and that

$$\begin{aligned} dx_{t|t-1} =& -\left[(A'\Lambda^{-1}A)^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right] dvecA' \\ &- \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}\right] dvecA \\ &+ \left[(A'\Lambda^{-1}A)^{-1} \otimes y_t'\Lambda^{-1}\right] dvecA' + w_t \\ =& -\left[(A'\Lambda^{-1}A)^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right] K_{pq}dvecA \\ &- \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}\right] dvecA \\ &+ \left[(A'\Lambda^{-1}A)^{-1} \otimes y_t'\Lambda^{-1}\right] K_{pq}dvecA + w_t. \end{aligned}$$

Consequently, we have

$$\begin{aligned} \frac{\partial x_{t|t-1}}{\partial (vecA)'} =& -\left[(A'\Lambda^{-1}A)^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right] K_{pq} \\ &- (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1} \\ &+ \left[(A'\Lambda^{-1}A)^{-1} \otimes y_t'\Lambda^{-1}\right] K_{pq} + w_t \\ =& -y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes (A'\Lambda^{-1}A)^{-1} \\ &- (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1} \\ &+ y_t'\Lambda^{-1} \otimes (A'\Lambda^{-1}A)^{-1} + w_t \end{aligned} \tag{49}$$

Now we may easily deduce from (49) that

$$\begin{aligned} \frac{\partial x^{0\prime}_{t|t-1}}{\partial vecA} =& -(\Lambda_0^{-1}A_0x_t + w_t) \otimes (A_0'\Lambda_0^{-1}A_0)^{-1} - \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1} \otimes (x_t + w_t) \\ &+ (\Lambda_0^{-1}A_0x_t + w_t) \otimes (A_0'\Lambda_0^{-1}A_0)^{-1} \\ =& -\Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1} \otimes x_t + w_t, \end{aligned} \tag{50}$$

and subsequently from (47) that

$$\begin{aligned}\frac{\partial \varepsilon_t^{0\prime}}{\partial vecA} &= -\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}A_0' - I_p \otimes x_{t|t-1}^0 \\ &= \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}A_0' \otimes x_t - I_p \otimes x_{t|t-1}^0 \\ &= -\left[I_p - \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\right] \otimes x_t + w_t,\end{aligned}$$

as was to be shown.

Secondly, we consider the partial derivative of $x_{t|t-1}$ with respect to $vec\Lambda$. Assuming A is fixed, we have

$$\begin{aligned}dx_{t|t-1} =& -(A'\Lambda^{-1}A)^{-1}A'(-\Lambda^{-1}d\Lambda\Lambda^{-1})A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t \\ &+ (A'\Lambda^{-1}A)^{-1}A'(-\Lambda^{-1}d\Lambda\Lambda^{-1})y_t + w_t \\ =& \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right]d\Lambda\left[\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t\right] \\ &- \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right]d\Lambda(\Lambda^{-1}y_t) + w_t\end{aligned}$$

and

$$\begin{aligned}dx_{t|t-1} =& \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right]dvec\Lambda \\ &- \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1}\right]dvec\Lambda + w_t.\end{aligned}$$

Consequently, we have

$$\begin{aligned}\frac{\partial x_{t|t-1}}{\partial (vec\Lambda)'} =& (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \\ &- (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1} + w_t \\ =& (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y_t'\Lambda^{-1}\left[A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} - I_p\right] + w_t,\end{aligned}$$

from which we have

$$\frac{\partial x_{t|t-1}^{0\prime}}{\partial vec\Lambda} = w_t$$

due to (48). The proof is therefore complete. □

Proof of Lemma 3.2 It follows immediately from (19) that

$$V_n(r) = -B_0' \otimes \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} v_t + o_p(1).$$

Moreover, due to (20), $T_s'(\partial \varepsilon_t^{0\prime}/\partial\theta)$ is a stationary linear process and $\mathcal{F}_{t-1}$-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 Theorem 3.3 The proof will be done in three steps, each of which will establish ML1, ML2 and ML3. As in CMP, we use the following notational convention in the proof:

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

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

First Step ML1 holds with N given in the theorem, as shown in the proof of Theorem 3.5 of CMP. In particular, we have

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

and

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

as $n\to\infty$. □

Second Step Now we establish ML2. It is shown in CMP that

$$\frac{1}{n^2}T_N' H_n(\theta_0)T_N \to_d -\int_0^1 V(r)\Sigma_0^{-1}V(r)'dr$$

as $n\to\infty$, and that

$$\frac{1}{n^{3/2}}T_N' H_n(\theta_0)T_S = O_p(n^{-1/2})$$

for large n, which are in particular due to

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

for large n.

In order to establish ML2, we only need to show

$$\frac{1}{n}T_S'H_n(\theta_0)T_S\to_p-[var(W)+var(Z)]. \tag{51}$$

Notice that

$$\frac{1}{n}T_S'H_n(\theta_0)T_S=A_n+B_n+C_n+(D_n+D_n')+o_p(1),$$

where

$$A_n=-\frac{1}{2}T_S'\left[\frac{\partial(vec\,\Sigma_0)'}{\partial\theta}(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\frac{\partial(vec\,\Sigma_0)}{\partial\theta'}\right]T_S+o_p(1)$$
$$B_n=-\frac{1}{n}\sum_{t=1}^{n}T_S'\left(\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\Sigma_0^{-1}\frac{\partial\varepsilon_t^0}{\partial\theta'}\right)T_S$$
$$C_n=-\frac{1}{n}\sum_{t=1}^{n}T_S'\left[(I\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1})\left(\frac{\partial^2}{\partial\theta\partial\theta'}\otimes\varepsilon_t^0\right)\right]T_S$$
$$D_n=T_S'\left[\frac{\partial(vec\,\Sigma_0)'}{\partial\theta}(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial\varepsilon_t^0}{\partial\theta'}\otimes\varepsilon_t^0\right)\right]T_S.$$

As shown in CMP,

$$A_n=-var(Z)+o_p(1)$$
$$B_n=-var(W)+o_p(1)$$
$$D_n=O_p(n^{-1/2})$$

for large n. Therefore, it suffices to show that

$$C_n=\begin{pmatrix} C_n(A,A) & C_n(A,\Lambda)\\ C_n(\Lambda,A) & C_n(\Lambda,\Lambda)\end{pmatrix}=O_p(n^{-1/2}) \tag{52}$$

to deduce (51). Note that we have from (49)

$$I_q\otimes x_{t|t-1}+(A'\otimes I_q)\frac{\partial x_{t|t-1}'}{\partial vecA}=\overline{w}_{t-1}, \tag{53}$$

which will be used below in the proof of (52).

First, we prove

$$C_n(A,A) = O_p(n^{-1/2}). \tag{54}$$

It follows from (47) that

$$vec\frac{\partial \varepsilon_t'}{\partial vecA} = -vec(I_p \otimes x_{t|t-1}) - vec\frac{\partial x_{t|t-1}'}{\partial vecA}A',$$

and since

$$vec(I_p \otimes x_{t|t-1}) = (I_p \otimes K_{pq})[(vecI_p) \otimes x_{t|t-1}]$$

and

$$\begin{aligned} vec\frac{\partial x_{t|t-1}'}{\partial vecA}A' &= (I_{pq} \otimes A)vec\frac{\partial x_{t|t-1}'}{\partial vecA} \\ &= \left(\frac{\partial x_{t|t-1}'}{\partial vecA} \otimes I_p\right) vecA' = \left(\frac{\partial x_{t|t-1}'}{\partial vecA} \otimes I_p\right) K_{pq}vecA, \end{aligned}$$

we have

$$\begin{aligned} \frac{\partial}{\partial (vecA)'}vec\frac{\partial \varepsilon_t'}{\partial vecA} &= -(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}}{\partial (vecA)'}\right] \\ &\quad - \left(\frac{\partial x_{t|t-1}'}{\partial vecA} \otimes I_p\right) K_{pq} - (I_{pq} \otimes A)\frac{\partial}{\partial (vecA)'}vec\frac{\partial x_{t|t-1}'}{\partial vecA}. \end{aligned} \tag{55}$$

In what follows, we will use (55) to show

$$(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)\left(\frac{\partial}{\partial (vecA)'}vec\frac{\partial \varepsilon_t^{0\prime}}{\partial vecA}\right)(A_0 \otimes I_q) = \overline{w}_{t-1}\varepsilon_t^0, \tag{56}$$

from which (54) follows immediately.

For the first term in (55), we have

$$\begin{aligned} &(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial (vecA)'}\right](A_0 \otimes I_q) \\ &= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial (vecA)'}\right](A_0 \otimes I_q) \\ &= A_0' \otimes \left(I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)K_{pq}\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial (vecA)'}\right](A_0 \otimes I_q) \\ &= \left(A_0' \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1} \otimes I_q\right)\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial (vecA)'}(A_0 \otimes I_q)\right] \\ &= A_0'\Sigma_0^{-1}\varepsilon_t^0 \otimes \left[\frac{\partial x_{t|t-1}^0}{\partial (vecA)'}(A_0 \otimes I_q)\right] \\ &= A_0'\Sigma_0^{-1}\varepsilon_t^0 \otimes I_q \otimes x_t' + \overline{w}_{t-1}\varepsilon_t^0. \end{aligned} \tag{57}$$

For the second term in (55), we may deduce that

$$
\begin{aligned}
&(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)\left(\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA} \otimes I_q\right) K_{pq}(A_0 \otimes I_q) \\
&= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)\left(\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA} \otimes I_q\right)(I_q \otimes A_0)K_{qq} \\
&= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)\left(\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA} \otimes A_0\right)K_{qq} \\
&= \left[(A_0' \otimes I_q)\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\right]K_{qq} \\
&= \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0 \otimes \left[(A_0' \otimes I_q)\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right] \\
&= \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0 \otimes I_q \otimes x_t + \overline{w}_{t-1}\varepsilon_t^0, \qquad (58)
\end{aligned}
$$

similarly as for the first term in (55).

The third term in (55) are written as

$$
\begin{aligned}
&(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_{pq} \otimes A_0)\left(\frac{\partial}{\partial(vecA)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right)(A_0 \otimes I_q) \\
&= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_{pq} \otimes A_0)\left(\frac{\partial}{\partial(vecA)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right)(A_0 \otimes I_q) \\
&= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\right)\left(\frac{\partial}{\partial(vecA)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right)(A_0 \otimes I_q), \qquad (59)
\end{aligned}
$$

and analyzed using the identity introduced in (53). It follows from (53) that

$$
\begin{aligned}
&(I_q \otimes K_{qq})\left[(vecI_q) \otimes \frac{\partial x_{t|t-1}^0}{\partial(vecA)'}\right] + (A' \otimes I_q \otimes I_q)\left(\frac{\partial}{\partial(vecA)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right) \\
&\quad + \left(I_q \otimes I_q \otimes \frac{\partial x_{t|t-1}^0}{\partial(vecA)'}\right)(I_q \otimes K_{pq} \otimes I_q)[K_{pq} \otimes (vecI_q)] = \overline{w}_{t-1}, \qquad (60)
\end{aligned}
$$

since

$$
vec(I_q \otimes x_{t|t-1}) = (I_q \otimes K_{qq})\left[(vecI_q) \otimes x_{t|t-1}\right]
$$

and

$$
\begin{aligned}
vec(A' \otimes I_q)\frac{\partial x_{t|t-1}'}{\partial vecA} &= (A' \otimes I_q \otimes I_q)vec\frac{\partial x_{t|t-1}'}{\partial vecA} \\
&= \left(I_q \otimes I_q \otimes \frac{\partial x_{t|t-1}}{\partial(vecA)'}\right)vec(A' \otimes I_q)
\end{aligned}
$$

with

$$
\begin{aligned}
vec(A'\otimes I_q) &= (I_q\otimes K_{pq}\otimes I_q)[(vecA')\otimes(vecI_q)]\\
&= (I_q\otimes K_{pq}\otimes I_q)[(K_{pq}vecA)\otimes(vecI_q)].
\end{aligned}
$$

See, e.g., Magnus and Neudecker (1988) for the rules in matrix algebra used here.

Now we pre- and post-multiply all three terms in (60) by

$$
I_q\otimes I_q\otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0 \quad \text{and} \quad A_0\otimes I_q.
$$

The first term in (60) becomes

$$
\begin{aligned}
&(I_q\otimes I_q\otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0)(I_q\otimes K_{qq})\left[(vecI_q)\otimes\frac{\partial x^0_{t|t-1}}{\partial(vecA)'}\right](A_0\otimes I_q)\\
&\quad=\left(I_q\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\otimes I_q\right)\left[(vecI_q)\otimes\left(\frac{\partial x^0_{t|t-1}}{\partial(vecA)'}(A_0\otimes I_q)\right)\right]\\
&\quad=A_0'\Sigma_0^{-1}\varepsilon_t^0\otimes I_q\otimes x_t'+\overline{w}_{t-1}\varepsilon_t^0. \qquad (61)
\end{aligned}
$$

On the other hand, the third term in (60) reduces to

$$
\begin{aligned}
&(I_q\otimes I_q\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}A_0)\left(I_q\otimes I_q\otimes\frac{\partial x^0_{t|t-1}}{\partial(vecA)'}\right)(I_q\otimes K_{pq}\otimes I_q)[K_{pq}\otimes(vecI_q)](A_0\otimes I_q)\\
&=\left[I_q\otimes I_q\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\left(A_0'\Lambda_0^{-1}A_0\right)^{-1}A_0'\Lambda_0^{-1}\otimes x_t'\right](I_q\otimes K_{pq}\otimes I_q)[K_{pq}(A_0\otimes I_q)\otimes(vecI_q)]\\
&\quad+\overline{w}_{t-1}\varepsilon_t^0\\
&=\left[I_q\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\left(A_0'\Lambda_0^{-1}A_0\right)^{-1}A_0'\Lambda_0^{-1}\otimes I_q\otimes x_t'\right](I_q\otimes K_{pq}\otimes I_q)[K_{pq}(A_0\otimes I_q)\otimes(vecI_q)]\\
&\quad+\overline{w}_{t-1}\varepsilon_t^0\\
&=\left[\varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\left(A_0'\Lambda_0^{-1}A_0\right)^{-1}A_0'\Lambda_0^{-1}\otimes I_q\right](A_0\otimes I_q)\otimes x_t+\overline{w}_{t-1}\varepsilon_t\\
&=\varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\otimes I_q\otimes x_t+\overline{w}_{t-1}\varepsilon_t^0. \qquad (62)
\end{aligned}
$$

Therefore, it follows from (60), (61) and (62) that

$$
\begin{aligned}
&(A_0'\otimes I_q\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}A_0)\left(\frac{\partial}{\partial(vecA)'}vec\frac{\partial x^{0\prime}_{t|t-1}}{\partial vecA}\right)(A_0\otimes I_q)\\
&\quad=A_0'\Sigma_0^{-1}\varepsilon_t^0\otimes I_q\otimes x_t'+\varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\otimes I_q\otimes x_t+\overline{w}_{t-1}\varepsilon_t^0, \qquad (63)
\end{aligned}
$$

which establishes the required result for the third term of (55), as shown in (59). Consequently, we may deduce (56) from (57), (58) and (63).

Second, we prove that

$$
C_n(A,\Lambda)=O_p(n^{-1/2}). \qquad (64)
$$

As we have shown earlier, we have

$$vec\frac{\partial \varepsilon_t'}{\partial vecA} = -(I_p \otimes K_{pq})[(vecI_p) \otimes x_{t|t-1}] - (I_{pq} \otimes A)vec\frac{\partial x_{t|t-1}'}{\partial vecA},$$

and it follows that

$$\begin{aligned}&\frac{\partial}{\partial (vec\Lambda)'}vec\frac{\partial \varepsilon_t'}{\partial vecA}\\ &= -(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}}{\partial (\Lambda)'}\right] - (I_{pq} \otimes A)\frac{\partial}{\partial (vec\Lambda)'}vec\frac{\partial x_{t|t-1}'}{\partial vecA}.\end{aligned} \tag{65}$$

In what follows, it will be shown that

$$(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)\left[\frac{\partial}{\partial (vec\Lambda)'}vec\frac{\partial \varepsilon_t^{0\prime}}{\partial vecA}\right]\lambda = \overline{w}_{t-1}\varepsilon_t^0 \tag{66}$$

for any p^2-dimensional vector λ. Clearly, (64) can be deduced immediately from (66).

For the first term in (65), we have

$$\begin{aligned}&(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial (vec\Lambda)'}\right]\lambda\\ &= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial (vec\Lambda)'}\right]\lambda\\ &= \left(A_0' \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1} \otimes I_q\right)(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial (vec\Lambda)'}\right]\lambda\\ &= A_0'\Sigma_0^{-1}\varepsilon_t^0 \otimes \left[\frac{\partial x_{t|t-1}^0}{\partial (vec\Lambda)'}\lambda\right] = \overline{w}_{t-1}\varepsilon_t^0.\end{aligned} \tag{67}$$

The proof for (66) will be finished, if we show that the second term in (65) also yields

$$\begin{aligned}&(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_{pq} \otimes A)\frac{\partial}{\partial (vec\Lambda)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\lambda\\ &= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_{pq} \otimes A)\frac{\partial}{\partial (vec\Lambda)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\lambda\\ &= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\right)\frac{\partial}{\partial (vec\Lambda)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\lambda = \overline{w}_{t-1}\varepsilon_t^0,\end{aligned} \tag{68}$$

similarly as the first term in (65).

To establish (68), we use the identity in (53). We may write it as

$$(I_q \otimes K_{qq})[(vecI_q) \otimes x_{t|t-1}] + (A' \otimes I_q \otimes I_q)vec\frac{\partial x_{t|t-1}'}{\partial vecA} = \overline{w}_{t-1},$$

from which it follows that

$$(I_q \otimes K_{qq}) \left[(vecI_q) \otimes \frac{\partial x_{t|t-1}}{\partial (vec\Lambda)'} \right] + (A' \otimes I_q \otimes I_q) \frac{\partial}{\partial (vec\Lambda)'} vec \frac{\partial x'_{t|t-1}}{\partial vecA} = \overline{w}_{t-1}. \tag{69}$$

Now we may evaluate (69) at the true values of parameters A and Λ, and pre- and post-multiply both sides by

$$I_q \otimes I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \quad \text{and} \quad \lambda$$

respectively, to get

$$A_0' \Sigma_0^{-1} \varepsilon_t^0 \otimes \frac{\partial x^0_{t|t-1}}{\partial (vec\Lambda)'} \lambda + \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \right) \frac{\partial}{\partial (vec\Lambda)'} vec \frac{\partial x^{0\prime}_{t|t-1}}{\partial vecA} \lambda = \overline{w}_{t-1} \varepsilon_t^0. \tag{70}$$

Note that

$$\begin{aligned} &(I_q \otimes I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0)(I_q \otimes K_{qq}) \left[(vecI_q) \otimes \frac{\partial x_{t|t-1}}{\partial (vec\Lambda)'} \right] \lambda \\ &\quad = \left(I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \otimes I_q \right) \left[(vecI_q) \otimes \frac{\partial x_{t|t-1}}{\partial (vec\Lambda)'} \right] = A_0' \Sigma_0^{-1} \varepsilon_t^0 \otimes \left[\frac{\partial x_{t|t-1}}{\partial (vec\Lambda)'} \lambda \right]. \end{aligned}$$

The proof for (64) is complete, since (68) can be deduced readily from from (70).

The proof for $C_n(\Lambda, \Lambda)$ is straightforward, as in Chang el al. (2007). Therefore, we have established (52), and the proof for the second step is complete. □

Third Step To establish ML3, as in CMP, we let

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

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

$$\begin{aligned} \left(B_0' \Lambda_0^{-1} \otimes I_k \right) (vecA - vecA_0) &= O(n^{-1+\delta}) \\ \left((A_0' \Lambda_0^{-1} A_0)^{-1/2} A_0' \Lambda_0^{-1} \otimes I_k \right) (vecA - vecA_0) &= O(n^{-1/2+\delta}) \\ vec\Lambda - vec\Lambda_0 &= O(n^{-1/2+\delta}), \end{aligned}$$

we have

$$vecA = vecA_0 + O_p(n^{-1/2+\delta}) \tag{71}$$

$$vec\Lambda = vec\Lambda_0 + O_p(n^{-1/2+\delta}). \tag{72}$$

We will show that

$$\frac{1}{n^{2(1-\delta)}}T_N'\left[\sum_{t=1}^{n}\left(\frac{\partial\varepsilon_t'}{\partial\theta}-\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\right)\Sigma_0^{-1}\frac{\partial\varepsilon_t^0}{\partial\theta'}\right]T_N \quad \to_p \quad 0 \tag{73}$$

$$\frac{1}{n^{2(1-\delta)}}T_N'\left[\sum_{t=1}^{n}\left(\frac{\partial\varepsilon_t'}{\partial\theta}-\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\right)\Sigma_0^{-1}\left(\frac{\partial\varepsilon_t}{\partial\theta'}-\frac{\partial\varepsilon_t^0}{\partial\theta'}\right)\right]T_N \quad \to_p \quad 0 \tag{74}$$

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

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

and

$$\frac{1}{n^{1-\delta}}T_S'\left[\sum_{t=1}^{n}\left(\frac{\partial\varepsilon_t'}{\partial\theta}-\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\right)\Sigma_0^{-1}\frac{\partial\varepsilon_t^0}{\partial\theta'}\right]T_S \quad \to_p \quad 0 \tag{77}$$

$$T_S'\left[\frac{\partial(vec\,\Sigma_0)'}{\partial\theta}(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\frac{1}{n^{1-\delta}}\sum_{t=1}^{n}\left(\left(\frac{\partial\varepsilon_t}{\partial\theta'}-\frac{\partial\varepsilon_t^0}{\partial\theta'}\right)\otimes\varepsilon_t^0\right)\right]T_S \quad \to_p \quad 0 \tag{78}$$

$$\frac{1}{n^{1-\delta}}T_S'\left[\sum_{t=1}^{n}\left(\frac{\partial\varepsilon_t'}{\partial\theta}-\frac{\partial\varepsilon_t^{0\prime}}{\partial\theta}\right)\Sigma_0^{-1}\left(\frac{\partial\varepsilon_t}{\partial\theta'}-\frac{\partial\varepsilon_t^0}{\partial\theta'}\right)\right]T_S \quad \to_p \quad 0 \tag{79}$$

$$T_S'\left[\frac{\partial(vec\,\Sigma_0)'}{\partial\theta}(\Sigma_0^{-1}\otimes\Sigma_0^{-1})\frac{1}{n^{1-\delta}}\sum_{t=1}^{n}\left(\left(\frac{\partial\varepsilon_t}{\partial\theta'}-\frac{\partial\varepsilon_t^0}{\partial\theta'}\right)\otimes(\varepsilon_t-\varepsilon_t^0)\right)\right]T_S \quad \to_p \quad 0 \tag{80}$$

$$\frac{1}{n^{1-\delta}}\sum_{t=1}^{n}T_S'\left[(I\otimes(\varepsilon_t'-\varepsilon_t^{0\prime})\Sigma_0^{-1})\left(\frac{\partial^2}{\partial\theta\partial\theta'}\otimes\varepsilon_t^0\right)\right]T_s \quad \to_p \quad 0 \tag{81}$$

$$\frac{1}{n^{1-\delta}}\sum_{t=1}^{n}T_S'\left[(I\otimes(\varepsilon_t'-\varepsilon_t^{0\prime})\Sigma_0^{-1})\left(\frac{\partial^2}{\partial\theta\partial\theta'}\otimes(\varepsilon_t-\varepsilon_t^0)\right)\right]T_S \quad \to_p \quad 0 \tag{82}$$

for all A and Λ satisfying (71) and (72). In what follows, we use the generic notation $\Delta(n^\kappa d_t)$ to denote the terms which include n^κ (or a lower order) times (d_t), (d_t) can be stationary or nonstationary. Clearly, we have

$$\varepsilon_t-\varepsilon_t^0,\frac{\partial\varepsilon_t}{\partial\theta'}-\frac{\partial\varepsilon_t^0}{\partial\theta'},\frac{\partial^2}{\partial\theta\partial\theta'}\otimes(\varepsilon_t-\varepsilon_t^0)=\Delta(n^{-1/2+\delta}x_t)+w_t, \tag{83}$$

since both $A=A_0+O(n^{-1/2+\delta})$ and $\Lambda=\Lambda_0+O(n^{-1/2}+\delta)$. The results in (73)-(76) follow immediately from (83). In (76), note that

$$\sum_{t=1}^{n}\left(\frac{\partial\varepsilon_t^0}{\partial\theta'}\otimes(\varepsilon_t-\varepsilon_t^0)\right)T_S=\sum_{t=1}^{n}\left(\frac{\partial\varepsilon_t^0}{\partial\theta'}T_S\otimes(\varepsilon_t-\varepsilon_t^0)\right)\quad\text{and}\quad\frac{\partial\varepsilon_t^0}{\partial\theta'}T_S=w_t.$$

The proof for (77)-(82) are more involved. In doing that, we need to show that

$$T_s' \left(\frac{\partial \varepsilon_t'}{\partial \theta} - \frac{\partial \varepsilon_t^{0'}}{\partial \theta} \right) = \Delta(n^{-1/2+\delta} w_t) + \Delta(n^{-1+2\delta} d_t) \tag{84}$$

which is equivalent to show the following two equations:

$$(A_0' \otimes I_k) \left(\frac{\partial \varepsilon_t'}{\partial vecA} - \frac{\partial \varepsilon_t^{0'}}{\partial vecA} \right) = \Delta(n^{-1/2+\delta} w_t) + \Delta(n^{-1+2\delta} d_t) \tag{85}$$

$$\frac{\partial \varepsilon_t'}{\partial vec\Lambda} - \frac{\partial \varepsilon_t^{0'}}{\partial vec\Lambda} = \Delta(n^{-1/2+\delta} w_t) + \Delta(n^{-1+2\delta} d_t) \tag{86}$$

Since

$$\begin{aligned} vec\left(\frac{\partial \varepsilon_t'}{\partial vec\Lambda} - \frac{\partial \varepsilon_t^{0'}}{\partial vec\Lambda} \right) &= \frac{\partial}{\partial (vec\Lambda)'} vec \frac{\partial \varepsilon_t^{0'}}{\partial vec\Lambda} (vec\Lambda - vec\Lambda_0) \\ &\quad + \frac{\partial}{\partial (vecA)'} vec \frac{\partial \varepsilon_t^{0'}}{\partial vec\Lambda} (vecA - vecA_0) + \Delta(n^{-1+2\delta} w_t), \end{aligned}$$

(86) follows immediately from Lemma 3.1. In order to show (85), it is useful to notice that the left hand side of equation (85) is a matrix with elements

$$A_{i0}' \left(\frac{\partial \varepsilon_{tj}}{\partial A_k} - \frac{\partial \varepsilon_{tj}^0}{\partial A_k} \right),$$

where A_i and A_k represent the ith and kth columns of A, respectively, and A_{i0} is the true value of A_i. ε_{tj} is the jth element of ε_t. Here i, $k = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, p$. We will follow the same notations in the rest of the proof. It is rather clear that to show equation (85) is equivalent to show:

$$\begin{aligned} A_{i0}' \left(\frac{\partial \varepsilon_{tj}}{\partial A_k} - \frac{\partial \varepsilon_{tj}^0}{\partial A_k} \right) &= A_{i0}' \left(\frac{\partial \varepsilon_{tj}^0}{\partial A_k \partial A_1'} (A_1 - A_{10}) + \cdots + \frac{\partial \varepsilon_{tj}^0}{\partial A_k \partial A_q'} (A_q - A_{q0}) \right) \\ &\quad + A_{i0}' \frac{\partial \varepsilon_{tj}}{\partial A_k \partial vec\Lambda'} (vec\Lambda - vec\Lambda_0) + \Delta(n^{-1+2\delta} d_t) \\ &= \Delta(n^{-1/2+\delta} w_t) + \Delta(n^{-1+2\delta} d_t). \end{aligned} \tag{87}$$

Since $(A_0' \otimes I_q) \frac{\partial \varepsilon_t^{0}{}'}{\partial vecA} = w_t$, we have

$$A_{i0}' \frac{\partial \varepsilon_{tj}^0}{\partial A_k} = w_t, \tag{88}$$

which implies

$$
\begin{aligned}
A_{i0}' \frac{\partial \varepsilon_{tj}^0}{\partial A_k \partial A_l'} &= w_t, \quad \text{for} \quad l \neq i && (89)\\
A_{i0}' \frac{\partial \varepsilon_{tj}^0}{\partial A_k \partial A_i'} + \frac{\partial \varepsilon_{tj}^0}{\partial A_k'} &= w_t && (90)\\
A_{i0}' \frac{\partial \varepsilon_{tj}^0}{\partial A_k \partial vec\Lambda'} &= w_t. && (91)
\end{aligned}
$$

Because of (88) and (90), we have

$$
A_{i0}' \frac{\partial \varepsilon_{tj}^0}{\partial A_k \partial A_i'}(A_i - A_{i0}) = \left(w_t - \frac{\partial \varepsilon_{tj}^0}{\partial A_k'} \right)(A_i - A_{i0}) \quad = \quad \Delta(n^{1/2+\delta} w_t). \tag{92}
$$

Equation (87) follows immediately from (89), (91) and (92). Based on (83) and (84), equations (77)-(80) follow. In (80), it is useful to note that

$$
\left[\left(\frac{\partial \varepsilon_t}{\partial \theta'} - \frac{\partial \varepsilon_t^0}{\partial \theta'} \right) \otimes \varepsilon_t^0 \right] T_S = \left(\frac{\partial \varepsilon_t}{\partial \theta'} - \frac{\partial \varepsilon_t^0}{\partial \theta'} \right) T_S \otimes \varepsilon_t^0
$$

Moreover, equation (84) implies that

$$
\left(\frac{\partial}{\partial \theta \partial \theta'} \otimes \varepsilon_t^0 \right) T_s = w_t,
$$

which finishes the proofs of equations (81) and (82). $\square$

Proof of Proposition 4.1 The stated result follows immediately from Lemma 2.2 and equation (17). To see this, note that we have from Lemma 2.2

$$
\begin{aligned}
A_0 x_{t|t-1}^0 &= A_0 (A_0' \Lambda_0^{-1} A_0)^{-1} A_0' \Lambda_0^{-1} y_t - \sum_{k=0}^{t-1} A_0 (I_q - \Omega_0^{-1})^k (A_0 \Lambda_0^{-1} A_0)^{-1} A_0' \Lambda_0^{-1} \triangle y_{t-k} \\
&= A_0 (A_0' \Lambda_0^{-1} A_0)^{-1} A_0' \Lambda_0^{-1} y_{t-1} - \sum_{k=1}^{t-1} C_k \triangle y_{t-k}
\end{aligned}
$$

under the convention $x_0 = 0$. Moreover, it follows that

$$
\begin{aligned}
\triangle y_t &= (A_0 x_{t|t-1}^0 - y_{t-1}) + \varepsilon_t^0 \\
&= -(I_p - A_0 (A_0' \Lambda_0^{-1} A_0)^{-1} A_0' \Lambda_0^{-1}) y_{t-1} - \sum_{k=1}^{t-1} C_k \triangle y_{t-k} + \varepsilon_t^0 \\
&= -B_0 B_0' \Lambda_0^{-1} y_{t-1} - \sum_{k=1}^{t-1} C_k \triangle y_{t-k} + \varepsilon_t^0,
\end{aligned}
$$

due to the definition of (ε_t^0) and B_0. $\square$

Proof of Theorem 4.2 We let

$$\tilde{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^0 \varepsilon_t^{0\prime}$$

and

$$G_n = \hat{B}_n' \hat{\Lambda}_n^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} y_t y_t' \right) \hat{\Lambda}_n^{-1} \hat{B}_n$$

throughout the proof. The proof will be done in three steps.

First Step First, we establish that

$$G_n = I_{p-q} + B_0' \Lambda_0^{-1} \left[(\tilde{\Sigma}_n - \Sigma_0) - (\hat{\Lambda}_n - \Lambda_0) \right] \Lambda_0^{-1} B_0 + o_p(n^{-1/2}). \tag{93}$$

To show this, note first that we have

$$\begin{aligned} O_p(n^{-1}) &= (\hat{A}_n - A_0)'(\hat{\Lambda}_n^{-1}\hat{B}_n - \Lambda_0^{-1}B_0) \\ &= -A_0'\hat{\Lambda}_n^{-1}\hat{B}_n - \hat{A}_n'\Lambda_0^{-1}B_0 \\ &= -A_0'\hat{\Lambda}_n^{-1}\hat{B}_n + O_p(n^{-1}), \end{aligned} \tag{94}$$

since

$$\hat{A}_n \Lambda_n^{-1} \hat{B}_n = A_0' \Lambda_0^{-1} B_0 = 0$$

by the definition of B_0 and $\hat{B}_n$, and

$$\hat{A}_n = A_0 + O_p(n^{-1/2}),\ \hat{B}_n = B_0 + O_p(n^{-1/2}),\ \hat{\Lambda}_n = \Lambda_0 + O_p(n^{-1/2})$$
$$B_0' \Lambda_0^{-1} \hat{A}_n = O_p(n^{-1}),$$

as we have shown earlier.

Therefore, it follows from (94) that

$$A_0' \hat{\Lambda}_n^{-1} \hat{B}_n = O_p(n^{-1}), \tag{95}$$

and consequently,

$$\begin{aligned} G_n &= \hat{B}_n' \hat{\Lambda}_n^{-1} \tilde{\Sigma}_n \hat{\Lambda}_n^{-1} \hat{B}_n + O_p(n^{-1}) \\ &= I_{p-q} + \hat{B}_n' \hat{\Lambda}_n^{-1} \left(\tilde{\Sigma}_n - \hat{\Lambda}_n \right) \hat{\Lambda}_n^{-1} \hat{B}_n + O_p(n^{-1}) \\ &= I_{p-q} + B_0' \Lambda_0^{-1} \left(\tilde{\Sigma}_n - \hat{\Lambda}_n \right) \hat{\Lambda}_0^{-1} B_0 + o_p(n^{-1/2}), \end{aligned} \tag{96}$$

due in particular to (95). Finally, note that

$$B_0' \Lambda_0^{-1} \Sigma_0 \Lambda_0^{-1} B_0 = B_0' \Lambda_0^{-1} \Lambda_0 \Lambda_0^{-1} B_0 = I_{p-q}$$

from which we may readily deduce (93) from (96).

Our result in (93) holds if G_n is defined as

$$G_n = \hat{B}_n' \hat{\Lambda}_n^{-1} \left[\frac{1}{n} \sum_{t=1}^{n} \left(y_t - \sum_{k=1}^{m} \hat{\Pi}_k \triangle y_{t-k} \right) \left(y_t - \sum_{k=1}^{m} \hat{\Pi}_k \triangle y_{t-k} \right)' \right] \hat{\Lambda}_n^{-1} \hat{B}_n$$

for more general model in (3), where $(\hat{\Pi}_k)$ are the ML estimates of (Π_k). In this case, we have

$$y_t - \sum_{k=1}^{m} \hat{\Pi}_k \triangle y_{t-k} = A_0 x_{t|t-1} + \varepsilon_t^0 - \sum_{k=1}^{m} (\hat{\Pi}_k - \Pi_k) \triangle y_{t-k}.$$

Therefore, it is easy to see that (93) continues to hold, since

$$\hat{\Pi}_k = \Pi_k + O_p(n^{-1/2}), \ \sum_{t=1}^{n} \triangle y_{t-k} \varepsilon_t^0 = O_p(n^{1/2}), \ \sum_{t=1}^{n} x_{t|t-1} \triangle y_{t-k} = O_p(n)$$

for all $k = 1, \ldots, m$.

Second Step Secondly, we show that

$$\sqrt{n} \left(\tilde{\Sigma}_n - \Sigma_0 \right) - \sqrt{n} \left(\hat{\Lambda}_n - \Lambda_0 \right) \to_d \mathbb{N}(0, \Psi) \tag{97}$$

with some $\Psi > 0$. For this, we let

$$\frac{1}{n} T_S' H_n(\theta_0) T_S \to_p H_S = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix},$$

where the partition of matrix H_S is made conformably with T_S. Then we have

$$\begin{aligned} \sqrt{n} \left(vec\, \hat{\Lambda}_n - vec\, \Lambda_0 \right) &= D \sqrt{n} [v(\hat{\Lambda}_n) - v(\Lambda_0)] \\ &= -D H_{22 \cdot 1}^{-1} (-H_{21} H_{11}^{-1}, I) \left(\frac{T_S' s_n(\theta_0)}{\sqrt{n}} \right) + o_p(1), \end{aligned} \tag{98}$$

where D is the duplication matrix, $H_{22\cdot 1} = H_{22} - H_{21} H_{11}^{-1} H_{12}$, and

$$\begin{aligned} \frac{T_S' s_n(\theta_0)}{\sqrt{n}} =& \frac{1}{2} T_S' \frac{\partial (vec \Sigma_0)'}{\partial \theta} (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \sqrt{n} \left(vec\, \tilde{\Sigma} - vec\, \Sigma_0 \right) \\ & - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} T_S' \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1} \varepsilon_t^0 \end{aligned} \tag{99}$$

as we derived earlier.

Now it is clear that $\sqrt{n}(\tilde{\Sigma}_n - \Sigma_0)$ and $\sqrt{n}(\hat{\Lambda}_n - \Lambda_0)$ are jointly normal asymptotically. To find the asymptotic variance Ψ, we first note that

$$\Psi = avar(\tilde{\Sigma}_n) + avar(\hat{\Lambda}_n) - acov(\tilde{\Sigma}_n, \hat{\Lambda}_n) - acov(\hat{\Lambda}_n, \tilde{\Sigma}_n). \tag{100}$$

We have

$$avar(\tilde{\Sigma}_n) = (I_{p^2} + K_{pp})(\Sigma_0 \otimes \Sigma_0) \tag{101}$$

$$avar(\hat{\Lambda}_n) = -DH_{22\cdot 1}^{-1}D'. \tag{102}$$

Moreover, it follows from (99) that

$$\begin{aligned} &acov\left(\frac{T_S' s_n(\theta_0)}{\sqrt{n}}, \sqrt{n}(vec\,\tilde{\Sigma}_n - vec\,\Sigma_0)\right) \\ &= \frac{1}{2}T_S'\frac{\partial(vec\Sigma_0)'}{\partial\theta}(\Sigma_0^{-1} \otimes \Sigma_0^{-1})\,avar\left(\sqrt{n}(vec\,\tilde{\Sigma}_n - vec\,\Sigma_0)\right) \\ &= \frac{1}{2}T_S'\frac{\partial(vec\Sigma_0)'}{\partial\theta}(\Sigma_0^{-1} \otimes \Sigma_0^{-1})(I_{p^2} + K_{pp})(\Sigma_0 \otimes \Sigma_0) \\ &= \frac{1}{2}T_S'\frac{\partial v(\Sigma_0)'}{\partial\theta}D'(\Sigma_0^{-1} \otimes \Sigma_0^{-1})(I_{p^2} + K_{pp})(\Sigma_0 \otimes \Sigma_0) \\ &= T_S'\frac{\partial v(\Sigma_0)'}{\partial\theta}D' \\ &= T_S'\frac{\partial(vec\Sigma_0)'}{\partial\theta}, \end{aligned}$$

and subsequently from (98) that

$$acov(\hat{\Lambda}_n, \tilde{\Sigma}_n) = -DH_{22\cdot 1}^{-1}(-H_{21}H_{11}^{-1}, I)T_S'\frac{\partial(vec\Sigma_0)'}{\partial\theta}. \tag{103}$$

Note in particular that $KD = D$ for any commutation and duplication matrices of conformable dimensions.

Now we calculate $(\partial/\partial\theta')vec\,\Sigma$ to obtain the asymptotic covariance in (103) more explicitly. For this, we write

$$\Sigma = A\Omega A' + \Lambda$$

with

$$\Omega = \frac{1}{2}\left[I_q + \left(I_q + 4(A'\Lambda^{-1}A)^{-1}\right)^{1/2}\right].$$

First, with respect to A, we have

$$\begin{aligned} vec\,d\Sigma &= vec(dA\Omega A' + Ad\Omega A' + A\Omega dA') \\ &= (I_p \otimes A\Omega)vec\,dA + (A \otimes A)vec\,d\Omega + (A\Omega \otimes I_p)K_{pq}vec\,dA. \end{aligned}$$

Moreover, we may easily deduce that

$$\begin{aligned} vec\,d\Omega =& \frac{1}{2}\left[\left(I_q + 4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2} \otimes I_q + I_q \otimes \left(I_q + 4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\right] \\ & 4vec\,d(A'\Lambda^{-1}A)^{-1} \end{aligned}$$

and that

$$\begin{aligned} vec\,d(A'\Lambda^{-1}A)^{-1} &= -vec\left[(A'\Lambda^{-1}A)^{-1}d(A'\Lambda^{-1}A)(A'\Lambda^{-1}A)^{-1}\right] \\ &= -\left[(A'\Lambda^{-1}A)^{-1} \otimes (A'\Lambda^{-1}A)^{-1}\right]vec\,d(A'\Lambda^{-1}A) \\ &= -\left[(A'\Lambda^{-1}A)^{-1} \otimes (A'\Lambda^{-1}A)^{-1}\right]\left[(I_q \otimes A'\Lambda^{-1})K_{pq} + (A'\Lambda^{-1} \otimes I_q)\right]vec\,dA. \end{aligned}$$

Therefore, we have

$$\begin{aligned}vec\, d\Omega = & -2\left[\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\otimes I_q + I_q\otimes\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\right]\\ & \left[(A'\Lambda^{-1}A)^{-1}\otimes(A'\Lambda^{-1}A)^{-1}\right]\left[(I_q\otimes A'\Lambda^{-1})K_{pq}+(A'\Lambda^{-1}\otimes I_q)\right] vec\, dA\end{aligned}$$

and

$$\begin{aligned}vec\, d\Sigma = & (I_p\otimes A\Omega) vec\, dA + (A\Omega\otimes I_p)K_{pq} vec\, dA\\ & -2(A\otimes A)\left[\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\otimes I_q + I_q\otimes\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\right]\\ & \left[(A'\Lambda^{-1}A)^{-1}\otimes(A'\Lambda^{-1}A)^{-1}\right]\left[(I_q\otimes A'\Lambda^{-1})K_{pq}+(A'\Lambda^{-1}\otimes I_q)\right] vec\, dA.\end{aligned}$$

Consequently,

$$\begin{aligned}\frac{\partial vec\,\Sigma}{\partial(vec\, A)'} = & (I_{p^2}+K_{pp})(I_p\otimes A\Omega) - 2(I_{p^2}+K_{pp})(A\otimes A)\left[\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\otimes I_q\right.\\ & \left. + I_q\otimes\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\right]\left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\otimes(A'\Lambda^{-1}A)^{-1}\right] \qquad (104)\end{aligned}$$

as we may readily deduce.

Next, with respect to Λ, we have

$$vec\, d\Sigma = (A\otimes A) vec\, d\Omega + vec\,\Lambda.$$

Moreover, similarly as above, we have

$$\begin{aligned}vec\, d\Omega = & \frac{1}{2}\left[\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\otimes I_q + I_q\otimes\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\right]\\ & 4 vec\, d(A'\Lambda^{-1}A)^{-1}\end{aligned}$$

and that

$$\begin{aligned}vec\, d(A'\Lambda^{-1}A)^{-1} & = -vec\left[(A'\Lambda^{-1}A)^{-1}d(A'\Lambda^{-1}A)(A'\Lambda^{-1}A)^{-1}\right]\\ & = -\left[(A'\Lambda^{-1}A)^{-1}\otimes(A'\Lambda^{-1}A)^{-1}\right] vec\, d(A'\Lambda^{-1}A)\\ & = \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\otimes(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right] vec\,\Lambda.\end{aligned}$$

Therefore, it follows that

$$\begin{aligned}vec\, d\Sigma = & 2(A\otimes A)\left[\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\otimes I_q + I_q\otimes\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\right]\\ & \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\otimes(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right] vec\,\Lambda + vec\,\Lambda.\end{aligned}$$

Consequently, we have

$$\begin{aligned}\frac{\partial vec\,\Sigma}{\partial(vec\,\Lambda)'} = & 2(A\otimes A)\left[\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\otimes I_q + I_q\otimes\left(I_q+4(A'\Lambda^{-1}A)^{-1}\right)^{-1/2}\right]\\ & \left[(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\otimes(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right] + I_{p^2}. \qquad (105)\end{aligned}$$

Due to (104) and (105), the asymptotic covariance in (103) can now be expressed solely in terms of model parameters.

Third Step It follows immediately from (93) and (97) that

$$\sqrt{n}(G_n - I_{p-q}) \to_d \mathbb{N}\left(0, (B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})\Psi(\Lambda_0^{-1}B_0 \otimes \Lambda_0^{-1}B_0)\right)$$

as $n \to \infty$. Therefore, we may easily deduce that

$$\sqrt{n}\,\mathrm{tr}\,(G_n - I_{p-q}) \to_d \mathbb{N}(0, \omega^2) \tag{106}$$

with

$$\omega^2 = (vec\, I_{p-q})'(B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})\Psi(\Lambda_0^{-1}B_0 \otimes \Lambda_0^{-1}B_0)(vec\, I_{p-q}),$$

since in particular

$$\mathrm{tr}\,(G_n - I_{p-q}) = (vec\, I_{p-q})'vec\,(G_n - I_{p-q}).$$

The proof is now complete if we show that

$$\omega^2 = 2(p-q) - (vec\, I_{p-q})'(B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})avar(\hat{\Lambda}_n)(\Lambda_0^{-1}B_0 \otimes \Lambda_0^{-1}B_0)(vec\, I_{p-q}), \tag{107}$$

due to (106). Note that $\hat{\omega}_n^2 \to_p \omega^2$.

First, we note that

$$\begin{aligned}
&(vecI_{p-q})'(B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})avar(\tilde{\Sigma}_n)(\Lambda_0^{-1}B_0 \otimes \Lambda_0^{-1}B_0)(vecI_{p-q}) \\
&\qquad = 2(vecI_{p-q})'(B_0'\Lambda_0^{-1}\Sigma_0\Lambda_0^{-1}B_0 \otimes B_0'\Lambda_0^{-1}\Sigma_0\Lambda_0^{-1}B_0)(vecI_{p-q}) \\
&\qquad = 2(vecI_{p-q})'(B_0'\Lambda_0^{-1}B_0 \otimes B_0'\Lambda_0^{-1}B_0)(vecI_{p-q}) \\
&\qquad = 2(p-q),
\end{aligned} \tag{108}$$

due to (101) and

$$\Sigma_0\Lambda_0^{-1}B_0 = B_0 \qquad \text{and} \qquad B_0'\Lambda_0^{-1}B_0 = I_{p-q}$$

by the definition of B_0.

Second, it follows from (104) and (105) that

$$\begin{aligned}
(B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})\frac{\partial vec\,\Sigma_0}{\partial(vec\,A)'} &= 0 \\
(B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})\frac{\partial vec\,\Sigma_0}{\partial v(\Lambda)'} &= (B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})D.
\end{aligned}$$

As a result, we may deduce from (102) and (103) that

$$\begin{aligned}
&(vecI_{p-q})'(B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})acov(\hat{\Lambda}_n, \tilde{\Sigma}_n)(\Lambda_0^{-1}B_0 \otimes \Lambda_0^{-1}B_0)(vecI_{p-q}) \\
&\qquad = (vecI_{p-q})'(B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})acov(\tilde{\Sigma}_n, \hat{\Lambda}_n)(\Lambda_0^{-1}B_0 \otimes \Lambda_0^{-1}B_0)(vecI_{p-q}) \\
&\qquad = (vecI_{p-q})'(B_0'\Lambda_0^{-1} \otimes B_0'\Lambda_0^{-1})avar(\hat{\Lambda}_n)(\Lambda_0^{-1}B_0 \otimes \Lambda_0^{-1}B_0)(vecI_{p-q}).
\end{aligned} \tag{109}$$

Consequently, (107) follows immediately from (100), (108) and (109). □

Appendix B: Tables

Table 1: Parameter Estimates from One-Trend Model

Parameter	Estimates	Parameter	Estimates	Parameter	Estimates
a_1	0.4243 (0.0053)	π_{22}	-0.2114 (0.0105)	π_{47}	-0.1287 (0.0471)
a_2	0.4310(0.0047)	π_{23}	-0.3800 (0.0222)	*π_{48}	-0.0751 (0.0506)
a_3	0.4677 (0.0044)	π_{24}	-0.4257 (0.0353)	*π_{49}	0.0160 (0.0610)
a_4	0.4872 (0.0022)	π_{25}	-0.4141 (0.0419)	π_{55}	-0.0603 (0.0050)
a_5	0.4945 (0.0007)	π_{26}	-0.3634 (0.0484)	π_{56}	-0.0523 (0.0176)
a_6	0.5065 (0.0008)	π_{27}	-0.3308 (0.0521)	*π_{57}	-0.0325 (0.0222)
a_7	0.5155 (0.0010)	π_{28}	-0.3010 (0.0547)	*π_{58}	-0.0327 (0.0296)
a_8	0.5192 (0.0018)	π_{29}	-0.2204 (0.0590)	*π_{59}	0.0032 (0.0400)
a_9	0.5251 (0.0033)	π_{33}	0.1580 (0.0076)	π_{66}	-0.0654 (0.0134)
π_{11}	-1.6620 (0.2065)	π_{34}	0.2292 (0.0191)	*π_{67}	-0.0386 (0.0237)
π_{21}	-1.5518 (0.2059)	π_{35}	0.2617 (0.0268)	*π_{68}	-0.0410 (0.0373)
π_{31}	-1.5583 (0.2211)	π_{36}	0.2679 (0.0359)	*π_{69}	0.0330 (0.0517)
π_{41}	-1.1825 (0.2085)	π_{37}	0.2396 (0.0412)	π_{77}	0.0470 (0.0114)
π_{51}	-0.9529 (0.1988)	π_{38}	0.2345 (0.0450)	*π_{78}	0.0372 (0.0247)
π_{61}	-0.6567 (0.1857)	π_{39}	0.1476 (0.0523)	*π_{79}	-0.0192 (0.0454)
π_{71}	-0.4834 (0.1787)	π_{44}	-0.1425 (0.0161)	π_{88}	-0.0632 (0.0103)
π_{81}	-0.3598 (0.1728)	π_{45}	-0.1695 (0.0266)	*π_{89}	-0.0334 (0.0556)
*π_{91}	-0.1055 (0.1591)	π_{46}	-0.1391 (0.0388)	*π_{99}	0.0001 (0.0723)

[1] Parameters with * are not significant at the level of $\alpha = 0.05$

Table 2: Parameter Estimates from Two-Trend Model

Parameter	Estimates	Parameter	Estimates	Parameter	Estimates
a_{11}	0.3543 (0.0174)	π_{41}	0.4762 (0.1610)	π_{44}	0.1416 (0.0356)
a_{21}	0.3387 (0.0179)	π_{51}	0.5683 (0.1816)	π_{54}	0.2523 (0.0501)
a_{31}	0.3575 (0.0195)	π_{61}	0.6479 (0.2011)	π_{64}	0.4294 (0.0551)
a_{41}	0.2820 (0.0172)	π_{71}	0.6717 (0.2079)	π_{74}	0.4829 (0.0583)
a_{51}	0.2350 (0.0132)	π_{81}	0.6971 (0.2127)	π_{84}	0.5668 (0.0421)
a_{61}	0.1714 (0.0052)	π_{91}	0.6591 (0.2038)	π_{94}	0.5777 (0.0281)
a_{71}	0.1278 (0.0094)	*π_{22}	-0.0042 (0.0313)	π_{55}	0.0562 (0.0102)
a_{81}	0.1011 (0.0123)	π_{32}	0.0917 (0.0548)	*π_{65}	0.0258 (0.0346)
a_{91}	0.0259 (0.0123)	π_{42}	0.3313 (0.0688)	*π_{75}	0.0221 (0.0403)
a_{12}	0.5087 (0.0046)	π_{52}	0.4233 (0.0746)	*π_{85}	0.0128 (0.0566)
a_{22}	0.5147 (0.0044)	π_{62}	0.4712 (0.0811)	*π_{95}	0.0060 (0.0692)
a_{32}	0.5570 (0.0045)	π_{72}	0.4932 (0.0839)	*π_{66}	-0.0274 (0.0282)
a_{42}	0.5715 (0.0027)	π_{82}	0.4770 (0.0885)	*π_{76}	-0.0495 (0.0436)
a_{52}	0.5751 (0.0016)	π_{92}	0.4016 (0.0918)	*π_{86}	-0.0016 (0.0672)
a_{62}	0.5823 (0.0007)	π_{33}	0.1420 (0.0189)	*π_{96}	0.0289 (0.0762)
a_{72}	0.5886 (0.0013)	π_{43}	0.2235 (0.0616)	π_{77}	0.0060 (0.0077)
a_{82}	0.5902 (0.0019)	π_{53}	0.3018 (0.0787)	π_{87}	0.0959 (0.0243)
a_{92}	0.5905 (0.0033)	π_{63}	0.4178 (0.0916)	π_{97}	0.0945 (0.0454)
π_{11}	-0.2337 (0.0843)	π_{73}	0.4463 (0.0977)	*π_{88}	-0.0044 (0.0480)
*π_{21}	0.0183 (0.0948)	π_{83}	0.5105 (0.0991)	*π_{98}	-0.0073 (0.0819)
π_{31}	0.2583 (0.1234)	π_{93}	0.5535 (0.0914)	π_{99}	0.0340 (0.0052)

[1] Parameters with * are not significant at the level of $\alpha = 0.05$