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 likelihood ratio test to determine the number of common stochastic trends in a system of integrated time series. The asymptotic distribution of the test statistic is also derived as standard chi-square.

First Draft: Jan 13, 2006
This Version: October 4, 2007

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 NSF Grant No. SES-0453069 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 integrated latent variables.

However, CMP allow for only one integrated latent factor, and do 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 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 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 estimator 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 the usual likelihood ratio test can be applied in our model to test for the number of common stochastic trends, and that it has asymptotic chi-square 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 particular, 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. We conclude the paper in Section 5. 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. The likelihood ratio test for q 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, however, 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, so 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

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

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

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

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).

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

[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.

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.

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 $(y_t - \sum_{k=1}^{m} \Pi_k \Delta y_{t-k})$. 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 estimator 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 estimator 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 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.

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

$$s_n(\hat\theta_n) = s_n(\theta_0) + H_n(\theta_n)(\hat\theta_n - \theta_0), \tag{6}$$

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 (6) that we may write

$$\nu_n' T^{-1}(\hat\theta_n - \theta_0) = -[\nu_n^{-1}T'H_n(\theta_n)T\nu_n^{-1\prime}]^{-1}[\nu_n^{-1}T's_n(\theta_0)] \tag{7}$$

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: $-\nu_n^{-1}T'H_n(\theta_0)T\nu_n^{-1\prime} \to_d M > 0$ 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

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

where $\Theta_n = \{\theta | \|\mu_n' T^{-1}(\theta - \theta_0)\| \le 1\}$ 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 (7) and continuous mapping theorem that

$$\nu_n' T^{-1}(\hat{\theta}_n - \theta_0) = -[\nu_n^{-1} T' H_n(\theta_0) T \nu_n^{-1'}]^{-1}[\nu_n^{-1} T' s_n(\theta_0)] + o_p(1) \to_d M^{-1} N \qquad (8)$$

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

$$\nu_n^{-1} T' \left(H_n(\theta_n) - H_n(\theta_0)\right) T \nu_n^{-1\prime} \to_p 0 \qquad (9)$$

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,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n (\varepsilon_t^0 \varepsilon_t^{0'} - \Sigma_0) \to_d \mathbb{N}\left(0, (I + K)(\Sigma_0 \otimes \Sigma_0)\right) \qquad (10)$$

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

$$\sum_{t=1}^n (\varepsilon_t^0 \varepsilon_t^{0\prime} - \Sigma_0) \;\text{ and }\; \sum_{t=1}^n \frac{\partial \varepsilon_t^{0\prime}}{\partial\theta} \Sigma_0^{-1} \varepsilon_t^0 \;\text{ are asymptotically independent.} \qquad (11)$$

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 (10), (11) and

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial \varepsilon_t^{0\prime}}{\partial\theta} \Sigma_0^{-1} \varepsilon_t^0 \to_d \mathbb{N}\left(0, \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^n \frac{\partial \varepsilon_t^{0\prime}}{\partial\theta} \Sigma_0^{-1} \frac{\partial \varepsilon_t^0}{\partial\theta'}\right), \qquad (12)$$

which can be readily obtained by employing the standard martingale CLT. Of course, asymptotics in (12) 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

$$\theta = ((vecA)', v(\Lambda)')', \tag{13}$$

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

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

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

$$P = I_q - \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}A_0' \tag{14}$$

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 (14), we follow CMP and introduce the necessary rotation. Let B_0 be an $p \times (p-q)$ matrix satisfying the conditions

$$B_0'\Lambda_0^{-1}A_0 = 0 \quad \text{and} \quad B_0'\Lambda_0^{-1}B_0 = I_{p-q}. \tag{15}$$

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

$$P = I_q - \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}A_0' = \Lambda_0^{-1}B_0B_0', \tag{16}$$

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

$$T = (T_N, T_S), \tag{17}$$

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, (15) and (16) that

$$T_N' \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} = \left(B_0' \otimes I_q\right) \left(\frac{\partial \varepsilon_t^{0\prime}}{\partial vecA}\right) = -B_0' \otimes x_t + c_t^N(u,v) \tag{18}$$

and

$$T_S' \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} = \begin{pmatrix} \left[(A_0'\Lambda_0^{-1}A_0)^{-1/2}A_0' \otimes I_q\right] \dfrac{\partial \varepsilon_t^{0\prime}}{\partial vecA} \\ \dfrac{\partial \varepsilon_t^{0\prime}}{\partial v(\Lambda)} \end{pmatrix} = c_t^S(u,v) \tag{19}$$

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

$$T^{-1} = \begin{pmatrix} B_0'\Lambda_0^{-1} \otimes I_q & 0 \\ (A_0'\Lambda_0^{-1}A_0)^{-1/2}A_0'\Lambda_0^{-1} \otimes I_q & 0 \\ 0 & I_{p(p+1)/2} \end{pmatrix} \tag{20}$$

from our definition of T given above in (17).

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

$$\frac{1}{n} T_N' \sum_{t=1}^{n} \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1}\varepsilon_t^0 \to_d \int_0^1 V(r)dU(r), \tag{21}$$

and

$$\frac{1}{\sqrt{n}} T_S' \sum_{t=1}^{n} \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \Sigma_0^{-1} \varepsilon_t^0 \to_d W, \tag{22}$$

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 (21) is mixed normal. On the other hand, the independence of W and U renders the two limit distributions in (21) and (22) 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 θ_n of θ defined in (13). They are given by (8) with the rotation matrix T in (17) and the sequence of normalization matrix

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

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 stochastic common 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

$$\begin{pmatrix} R \\ S \end{pmatrix} = -[var(W) + var(Z)]^{-1}(W - Z), \tag{23}$$

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 3.2 that

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

and

$$n\left(B_0'\Lambda_0^{-1} \otimes I_q\right) vec\hat{A}_n \to_d Q \tag{24}$$

$$\sqrt{n}\left((A_0'\Lambda_0^{-1}A_0)^{-1/2}A_0'\Lambda_0^{-1} \otimes I_q\right)(vec\hat{A}_n - vecA_0) \to_d R, \tag{25}$$

similarly as in CMP. In particular, it follows immediately from (24) and (25) 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 3.2 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 q of stochastic common trends in (y_t) is known. 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 stochastic common trends or the cointegrating relationships is typically unknown in empirical studies. In this section, we will develop the likelihood ratio test for the number of stochastic common 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 stochastic common trends is equivalent to testing for the number of cointegrating relationships. Therefore, at least conceptually, we may use the existing tests, such as Park (1990), Phillips and Ouliaris (1990) and Shin () in regression framework or Johansen (1988, 1991) and Ahn and Reinsel () in vector autoregressions and error correction models, to find out the cointegrating rank to determine the rank of stochastic common trends. Nevertheless, none of the existing methods seems appropriate in our framework. The regression based methods require normalization, which assumes that the coefficients of the regressands are nonzero. And we do not have any natural normalization for our model.

Vector autoregression or error correction model based methods do not require such a priori normalization. However, they have two shortcomings. 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 stochastic common 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

$$\triangle y_t = -B_0(\Lambda_0^{-1}B_0)'y_{t-1} - \sum_{k=1}^{t-1} C_k \Delta y_{t-k} + \varepsilon_t^0, \tag{26}$$

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 (26), 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.

Now we consider the null hypothesis

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

which will be tested against

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

Let $\hat{\theta}_n$ be the $(pq + p(p+1)/2)$-dimensional ML estimator computed under the assumption that rank A_0 is q-dimensional, and let $\tilde{\theta}_n$ be the $(p^2 + p(p+1)/2)$-dimensional ML estimator obtained under the presumption that rank $A_0 = p$. If we denote by $\ell_n(\hat{\theta}_n)$ and $\ell_n(\tilde{\theta}_n)$ the log-likelihood values respectively for the ML estimators $\hat{\theta}_n$ and $\tilde{\theta}_n$, the likelihood ratio statistic τ_n for the null and alternative hypotheses introduced above is defined as

$$\tau_n = -2\left(\ell_n(\hat{\theta}_n) - \ell_n(\tilde{\theta}_n)\right).$$

The following theorem establishes the limit null distribution of τ_n.

Theorem 4.2 We have

$$\tau_n \to_d \chi^2_{p(p-q)}$$

as $n \to \infty$.

In a large dimensional system appearing to share a few common stochastic trends, it seems more practically useful to consider

$$\mathbb{H}_1 : q < \operatorname{rank} A_0 \leq r$$

for some $r < p$. In this case, we may similarly define the likelihood ratio statistic by defining $\tilde{\theta}_n$ to be $(pr + p(p+1)/2)$-dimensional, and the statistic now has limit distribution that is chi-square with $p(r-q)$-degrees of freedom. It is also possible to set $r = q + 1$, and test sequentially for smaller values of q until the null is rejected in favor of the alternative. In this setting, we test for p-number of additional restrictions in each step, and therefore the test statistics will have chi-square with p-degrees of freedom.

So far we have constructed a proper test for determining the number of common stochastic trends k, and derived the asymptotic theories about the ML parameter estimates. Our results for model (1) can then be used to decompose time series (y_t) into the permanent and transitory (PT) components. As in CMP, we denote them as (y_t^P) and (y_t^T) respectively, such that

$$y_t^P = A_0 x^0_{t|t-1} \quad \text{and} \quad y_t^T = y_t - A_0 x^0_{t|t-1}. \tag{27}$$

It is easy to see that y_t^P is I(1) and predictable and y_t^T is I(0) and mds.

5. Conclusion

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: Mathematical Proofs

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

$$\Omega_{t+1|t} - I_q = \Omega_{t|t-1} - \Omega_{t|t-1}A'(A\Omega_{t|t-1}A' + \Lambda)^{-1}A\Omega_{t|t-1}. \tag{28}$$

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

$$X - I_q = X - XA'(AXA' + \Lambda)^{-1}AX. \tag{29}$$

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 (29), we have

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

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

$$\begin{aligned}(AXA' + \Lambda)^{-1} &= \Lambda^{-1} - \Lambda^{-1}AX(X + XA'\Lambda^{-1}AX)^{-1}XA'\Lambda^{-1} \\ &= \Lambda^{-1} - \Lambda^{-1}A(I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}, \end{aligned} \tag{31}$$

and therefore,

$$
\begin{aligned}
&XA'(AXA'+\Lambda)^{-1}AX \\
&\quad = XA'\Lambda^{-1}AX - XA'\Lambda^{-1}A(I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX \\
&\quad = XA'\Lambda^{-1}AX - (I_q + XA'\Lambda^{-1}A)(I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX \\
&\qquad + (I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX \\
&\quad = (I_q + XA'\Lambda^{-1}A)^{-1}XA'\Lambda^{-1}AX. \qquad (32)
\end{aligned}
$$

Consequently, we have

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

i.e.,

$$X(A'\Lambda^{-1}A)X = I_q + XA'\Lambda^{-1}A, \qquad (33)$$

due to (30) and (32).

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 (33). 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

$$
\begin{aligned}
x_{t+1|t} &= x_{t|t-1} + \Omega A'\Sigma^{-1}(y_t - y_{t|t-1}) \\
&= x_{t|t-1} + \Omega A'\Sigma^{-1}(y_t - Ax_{t|t-1}) \\
&= (I_q - \Omega A'\Sigma^{-1}A)x_{t|t-1} + \Omega A'\Sigma^{-1}y_t \\
&= (I_q - \Omega A'\Sigma^{-1}A)x_{t|t-1} + \Omega A'\Sigma^{-1}y_t \qquad (34)
\end{aligned}
$$

with the steady state values Ω and Σ. However, it follows from (28) that

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

i.e.,

$$\Omega A'\Sigma^{-1}A = \Omega^{-1}. \qquad (35)$$

We may also deduce from (31) that

$$\Sigma^{-1} = (A\Omega A' + \Lambda)^{-1} = \Lambda^{-1} - \Lambda^{-1}A(I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}, \tag{36}$$

which yields

$$\begin{aligned}\Omega A'\Sigma^{-1}A &= \Omega A'\Lambda^{-1}A - \Omega A'\Lambda^{-1}A(I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}A \\ &= \Omega A'\Lambda^{-1}A\left[I_q - (I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}A\right]. \end{aligned}\tag{37}$$

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

$$I_q - (I_q + \Omega A'\Lambda^{-1}A)^{-1}\Omega A'\Lambda^{-1}A = (\Omega A'\Lambda^{-1}A)^{-1}\Omega^{-1}. \tag{38}$$

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

$$\Omega A'\Sigma^{-1} = \Omega\left[\Lambda^{-1}A(\Omega A'\Lambda^{-1}A)^{-1}\Omega^{-1}\right]' = \Omega^{-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}, \tag{39}$$

due to (36) and (38).

Now we have from (34), (35) and (39) 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,

$$x_{t|t-1} = \sum_{k=1}^{t-1}(I_q - \Omega^{-1})^{k-1}\Omega^{-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_{t-k} + (I_q - \Omega^{-1})^{t-1}x_{1|0}. \tag{40}$$

Moreover,

$$\begin{aligned}&\sum_{k=1}^{t-1}(I_q - \Omega^{-1})^{k-1}\Omega^{-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_{t-k} \\ &\quad= \sum_{k=1}^{t-1}(I_q - \Omega^{-1})^{k-1}\left[I_q - (I_q - \Omega^{-1})\right](A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_{t-k} \\ &\quad= (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t - \sum_{k=0}^{t-2}(I_q - \Omega^{-1})^{k}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\triangle y_{t-k} \\ &\qquad - (I_q - \Omega^{-1})^{t-1}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_1 \\ &\quad= (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t - \sum_{k=0}^{t-1}(I_q - \Omega^{-1})^{k}(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\triangle y_{t-k}.\end{aligned}\tag{41}$$

The stated result now follows directly from (40) and (41). 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 equation (**??**), the readers are referred to the proof of Proposition 2.4 in Chang et al. (2007) 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 equation (**??**). It follows from Lemma 2.2 that

$$\begin{aligned} x^0_{t|t-1} =& (A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}(A_0x_t + u_t) \\ & - \sum_{k=0}^{t-1}(I_q - \Omega_0^{-1})^k(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}(A_0v_{t-k} + (u_{t-k} - u_{t-k-1})) \\ =& x_t + (A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_t - \sum_{k=0}^{t-1}(I_q - \Omega_0^{-1})^k(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}(u_{t-k} - u_{t-k-1}) \\ & - \sum_{k=0}^{t-1}(I_q - \Omega_0^{-1})^k v_{t-k}. \end{aligned} \tag{42}$$

However, we may easily deduce that

$$\begin{aligned} & \sum_{k=0}^{t-1}(I_q - \Omega_0^{-1})^k(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}(u_{t-k} - u_{t-k-1}) \\ & = (A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_t - \Omega_0^{-1}\sum_{k=1}^{t-1}(I_q - \Omega_0^{-1})^k(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}u_{t-k}. \end{aligned} \tag{43}$$

The stated result now follows immediately from (42) and (43). □

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

$$x_{t|t-1} = (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}y_t + w_t, \tag{44}$$

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

$$K_{ab}vecA = vecA' \tag{45}$$

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

$$\frac{\partial \varepsilon_t}{\partial (vecA)'} = -A\frac{\partial x_{t|t-1}}{\partial (vecA)'} - I_p \otimes x'_{t|t-1}. \tag{46}$$

Moreover, it follows that

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

The partial derivatives of ε_t with respect to $vecA$ and $vec\Lambda$ may therefore be easily obtained from (46) and (47), once we find the partial derivatives of $x_{t|t-1}$ with respect to $vecA$ and $vec\Lambda$ in (44).

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 (44) 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

$$\begin{aligned} \frac{\partial x_{t|t-1}}{\partial (vecA)'} =& - \left[(A'\Lambda^{-1}A)^{-1} \otimes y'_t\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1}\right] K_{pq} \\ &- (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y'_t\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1} \\ &+ \left[(A'\Lambda^{-1}A)^{-1} \otimes y'_t\Lambda^{-1}\right] K_{pq} + w_t \\ =& - y'_t\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes (A'\Lambda^{-1}A)^{-1} \\ &- (A'\Lambda^{-1}A)^{-1}A'\Lambda^{-1} \otimes y'_t\Lambda^{-1}A(A'\Lambda^{-1}A)^{-1} \\ &+ y'_t\Lambda^{-1} \otimes (A'\Lambda^{-1}A)^{-1} + w_t \end{aligned} \tag{48}$$

Now we may easily deduce from (48) that

$$\begin{aligned}\frac{\partial x^{0\prime}_{t|t-1}}{\partial vecA} =& -(\Lambda_0^{-1}A_0x_t + w_t)\otimes(A_0'\Lambda_0^{-1}A_0)^{-1} - \Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}\otimes(x_t+w_t)\\ &+(\Lambda_0^{-1}A_0x_t + w_t)\otimes(A_0'\Lambda_0^{-1}A_0)^{-1}\\ =& -\Lambda_0^{-1}A_0(A_0'\Lambda_0^{-1}A_0)^{-1}\otimes x_t + w_t,\end{aligned} \tag{49}$$

and subsequently from (46) that

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

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

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

due to (47). The proof is therefore complete. □

Proof of Lemma 3.2 It follows immediately from (18) that

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

Moreover, due to (19), $T_s'(\partial\varepsilon_t^{0\prime}/\partial\theta)$ 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 3.2 The proof will be done in three steps, each of which will establish ML1, ML2 and ML3. As in Chang et al. (2007), 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 Chang et al. (2005).

Second Step Now we establish ML2. It is shown in Chang et al. (2007) 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

$$\frac{1}{n}T_S'H_n(\theta_0)T_S \to_p -[var(W)+var(Z)]. \tag{50}$$

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

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

As shown in Chang et al. (2007),

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

for large n. Therefore, it suffices to show that

$$C_n = \begin{pmatrix} C_n(A,A) & C_n(A,\Lambda)\\ C_n(\Lambda,A) & C_n(\Lambda,\Lambda)\end{pmatrix} = O_p(n^{-1/2}) \tag{51}$$

to deduce (50). Note that we have from (48)

$$I_q\otimes x_{t|t-1} + (A'\otimes I_q)\frac{\partial x_{t|t-1}'}{\partial vecA} = \overline{w}_{t-1}, \tag{52}$$

which will be used below in the proof of (51).

First, we prove

$$C_n(A,A) = O_p(n^{-1/2}). \tag{53}$$

It follows from (46) 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

$$vec\frac{\partial x'_{t|t-1}}{\partial vecA}A' = (I_{pq}\otimes A)vec\frac{\partial x'_{t|t-1}}{\partial vecA}$$
$$= \left(\frac{\partial x'_{t|t-1}}{\partial vecA}\otimes I_p\right)vecA' = \left(\frac{\partial x'_{t|t-1}}{\partial vecA}\otimes I_p\right)K_{pq}vecA,$$

we have

$$\frac{\partial}{\partial(vecA)'}vec\frac{\partial\varepsilon'_t}{\partial vecA} = -(I_p\otimes K_{pq})\left[(vecI_p)\otimes\frac{\partial x_{t|t-1}}{\partial(vecA)'}\right]$$
$$-\left(\frac{\partial x'_{t|t-1}}{\partial vecA}\otimes I_p\right)K_{pq} - (I_{pq}\otimes A)\frac{\partial}{\partial(vecA)'}vec\frac{\partial x'_{t|t-1}}{\partial vecA}. \qquad (54)$$

In what follows, we will use (54) to show

$$(A'_0\otimes I_q)\left(I_{pq}\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}\right)\left(\frac{\partial}{\partial(vecA)'}vec\frac{\partial\varepsilon_t^{0\prime}}{\partial vecA}\right)(A_0\otimes I_q) = \overline{w}_{t-1}\varepsilon_t^0, \qquad (55)$$

from which (53) follows immediately.

For the first term in (54), we have

$$(A'_0\otimes I_q)\left(I_{pq}\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_p\otimes K_{pq})\left[(vecI_p)\otimes\frac{\partial x^0_{t|t-1}}{\partial(vecA)'}\right](A_0\otimes I_q)$$
$$= \left(A'_0\otimes I_q\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_p\otimes K_{pq})\left[(vecI_p)\otimes\frac{\partial x^0_{t|t-1}}{\partial(vecA)'}\right](A_0\otimes I_q)$$
$$= A'_0\otimes\left(I_q\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}\right)K_{pq}\left[(vecI_p)\otimes\frac{\partial x^0_{t|t-1}}{\partial(vecA)'}\right](A_0\otimes I_q)$$
$$= \left(A'_0\otimes\varepsilon_t^{0\prime}\Sigma_0^{-1}\otimes I_q\right)\left[(vecI_p)\otimes\frac{\partial x^0_{t|t-1}}{\partial(vecA)'}(A_0\otimes I_q)\right]$$
$$= A'_0\Sigma_0^{-1}\varepsilon_t^0\otimes\left[\frac{\partial x^0_{t|t-1}}{\partial(vecA)'}(A_0\otimes I_q)\right]$$
$$= A'_0\Sigma_0^{-1}\varepsilon_t^0\otimes I_q\otimes x'_t + \overline{w}_{t-1}\varepsilon_t^0. \qquad (56)$$

For the second term in (54), we may deduce that

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

similarly as for the first term in (54).

The third term in (54) are written as

$$
\begin{aligned}
&(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_{pq} \otimes A_0)\left(\frac{\partial}{\partial(vecA)'} vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right)(A_0 \otimes I_q) \\
&\quad = \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_{pq} \otimes A_0)\left(\frac{\partial}{\partial(vecA)'} vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right)(A_0 \otimes I_q) \\
&\quad = \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\right)\left(\frac{\partial}{\partial(vecA)'} vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right)(A_0 \otimes I_q),
\end{aligned} \tag{58}
$$

and analyzed using the identity introduced in (52). It follows from (52) that

$$
\begin{aligned}
&(I_q \otimes K_{qq})\left[(vecI_q) \otimes \frac{\partial x_{t|t-1}^0}{\partial(vecA)'}\right] + (A' \otimes I_q \otimes I_q)\left(\frac{\partial}{\partial(vecA)'} vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\right) \\
&\qquad + \left(I_q \otimes I_q \otimes \frac{\partial x_{t|t-1}^0}{\partial(vecA)'}\right)(I_q \otimes K_{pq} \otimes I_q)[K_{pq} \otimes (vecI_q)] = \overline{w}_{t-1},
\end{aligned} \tag{59}
$$

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

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

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 (59) 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 (59) becomes

$$(I_q \otimes I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0)(I_q \otimes K_{qq}) \left[(vecI_q) \otimes \frac{\partial x^0_{t|t-1}}{\partial (vecA)'} \right] (A_0 \otimes I_q)$$
$$= \left(I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \otimes I_q \right) \left[(vecI_q) \otimes \left(\frac{\partial x^0_{t|t-1}}{\partial (vecA)'} (A_0 \otimes I_q) \right) \right]$$
$$= A_0' \Sigma_0^{-1} \varepsilon_t^0 \otimes I_q \otimes x_t' + \overline{w}_{t-1} \varepsilon_t^0. \tag{60}$$

On the other hand, the third term in (59) reduces to

$$(I_q \otimes I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0) \left(I_q \otimes I_q \otimes \frac{\partial x^0_{t|t-1}}{\partial (vecA)'} \right) (I_q \otimes K_{pq} \otimes I_q)[K_{pq} \otimes (vecI_q)](A_0 \otimes I_q)$$
$$= \left[I_q \otimes I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \left(A_0' \Lambda_0^{-1} A_0 \right)^{-1} A_0' \Lambda_0^{-1} \otimes x_t' \right] (I_q \otimes K_{pq} \otimes I_q)[K_{pq}(A_0 \otimes I_q) \otimes (vecI_q)]$$
$$+ \overline{w}_{t-1} \varepsilon_t^0$$
$$= \left[I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \left(A_0' \Lambda_0^{-1} A_0 \right)^{-1} A_0' \Lambda_0^{-1} \otimes I_q \otimes x_t' \right] (I_q \otimes K_{pq} \otimes I_q)[K_{pq}(A_0 \otimes I_q) \otimes (vecI_q)]$$
$$+ \overline{w}_{t-1} \varepsilon_t^0$$
$$= \left[\varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \left(A_0' \Lambda_0^{-1} A_0 \right)^{-1} A_0' \Lambda_0^{-1} \otimes I_q \right] (A_0 \otimes I_q) \otimes x_t + \overline{w}_{t-1} \varepsilon_t$$
$$= \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \otimes I_q \otimes x_t + \overline{w}_{t-1} \varepsilon_t^0. \tag{61}$$

Therefore, it follows from (59), (60) and (61) that

$$(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0) \left(\frac{\partial}{\partial (vecA)'} vec \frac{\partial x^{0\prime}_{t|t-1}}{\partial vecA} \right) (A_0 \otimes I_q)$$
$$= A_0' \Sigma_0^{-1} \varepsilon_t^0 \otimes I_q \otimes x_t' + \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \otimes I_q \otimes x_t + \overline{w}_{t-1} \varepsilon_t^0, \tag{62}$$

which establishes the required result for the third term of (54), as shown in (58). Consequently, we may deduce (55) from (56), (57) and (62).

Second, we prove that

$$C_n(A, \Lambda) = O_p(n^{-1/2}). \tag{63}$$

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

$$\begin{aligned}
&\frac{\partial}{\partial(vec\Lambda)'}vec\frac{\partial\varepsilon_t'}{\partial vecA} \\
&= -(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}}{\partial(\Lambda)'}\right] - (I_{pq} \otimes A)\frac{\partial}{\partial(vec\Lambda)'}vec\frac{\partial x_{t|t-1}'}{\partial vecA}. \qquad (64)
\end{aligned}$$

In what follows, it will be shown that

$$(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)\left[\frac{\partial}{\partial(vec\Lambda)'}vec\frac{\partial\varepsilon_t^{0\prime}}{\partial vecA}\right]\lambda = \overline{w}_{t-1}\varepsilon_t^0 \qquad (65)$$

for any p^2-dimensional vector λ. Clearly, (63) can be deduced immediately from (65).

For the first term in (64), we have

$$\begin{aligned}
&(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial(vec\Lambda)'}\right]\lambda \\
&= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial(vec\Lambda)'}\right]\lambda \\
&= \left(A_0' \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1} \otimes I_q\right)(I_p \otimes K_{pq})\left[(vecI_p) \otimes \frac{\partial x_{t|t-1}^0}{\partial(vec\Lambda)'}\right]\lambda \\
&= A_0'\Sigma_0^{-1}\varepsilon_t^0 \otimes \left[\frac{\partial x_{t|t-1}^0}{\partial(vec\Lambda)'}\lambda\right] = \overline{w}_{t-1}\varepsilon_t^0. \qquad (66)
\end{aligned}$$

The proof for (65) will be finished, if we show that the second term in (64) also yields

$$\begin{aligned}
&(A_0' \otimes I_q)\left(I_{pq} \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_{pq} \otimes A)\frac{\partial}{\partial(vec\Lambda)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\lambda \\
&= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}\right)(I_{pq} \otimes A)\frac{\partial}{\partial(vec\Lambda)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\lambda \\
&= \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime}\Sigma_0^{-1}A_0\right)\frac{\partial}{\partial(vec\Lambda)'}vec\frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA}\lambda = \overline{w}_{t-1}\varepsilon_t^0, \qquad (67)
\end{aligned}$$

similarly as the first term in (64).

To establish (67), we use the identity in (52). 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

$$(I_q \otimes K_{qq}) \left[(vecI_q) \otimes \frac{\partial x_{t|t-1}}{\partial (vec\Lambda)'} \right] + (A' \otimes I_q \otimes I_q) \frac{\partial}{\partial (vec\Lambda)'} vec \frac{\partial x'_{t|t-1}}{\partial vecA} = \overline{w}_{t-1}. \qquad (68)$$

Now we may evaluate (68) 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

$$A_0' \Sigma_0^{-1} \varepsilon_t^0 \otimes \frac{\partial x_{t|t-1}^0}{\partial (vec\Lambda)'} \lambda + \left(A_0' \otimes I_q \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1} A_0 \right) \frac{\partial}{\partial (vec\Lambda)'} vec \frac{\partial x_{t|t-1}^{0\prime}}{\partial vecA} \lambda = \overline{w}_{t-1} \varepsilon_t^0. \qquad (69)$$

Note that

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

The proof for (63) is complete, since (67) can be deduced readily from from (69).

The proof for $C_n(\Lambda, \Lambda)$ is straightforward, as in Chang el al. (2007). Therefore, we have established (51), and the proof for the second step is complete. □

Third Step To establish ML3, as Chang et al. (2007), 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

$$vecA = vecA_0 + O_p(n^{-1/2+\delta}) \qquad (70)$$

$$vec\Lambda = vec\Lambda_0 + O_p(n^{-1/2+\delta}). \qquad (71)$$

We will show that

$$\frac{1}{n^{2(1-\delta)}} T_N' \left[\sum_{t=1}^n \left(\frac{\partial \varepsilon_t'}{\partial \theta} - \frac{\partial \varepsilon_t^{0\prime}}{\partial \theta} \right) \Sigma_0^{-1} \frac{\partial \varepsilon_t^0}{\partial \theta'} \right] T_N \quad \to_p \quad 0 \tag{72}$$

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

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

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

$$\frac{1}{n^{1-\delta}} \sum_{t=1}^n T_S' \left[(I \otimes \varepsilon_t^{0\prime} \Sigma_0^{-1}) \left(\frac{\partial^2}{\partial \theta \partial \theta'} \otimes (\varepsilon_t - \varepsilon_t^0) \right) \right] T_s \quad \to_p \quad 0 \tag{76}$$

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

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

and

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

$$\frac{1}{n^{1-\delta}} \sum_{t=1}^n T_S' \left[(I \otimes (\varepsilon_t' - \varepsilon_t^{0\prime}) \Sigma_0^{-1}) \left(\frac{\partial^2}{\partial \theta \partial \theta'} \otimes (\varepsilon_t - \varepsilon_t^0) \right) \right] T_S \quad \to_p \quad 0 \tag{80}$$

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

for all A and Λ satisfying (70) and (71). Here we only prove that the nonstationary components in (72)-(81) satisfy the required conditions. It is rather obvious that the required conditions hold for the stationary components. In what follows, we use the generic notation $\Delta(n^\kappa x_t)$ to denote the terms which include n^κ (or of a lower order) times (x_t). Clearly, we have

$$\varepsilon_t - \varepsilon_t^0, \frac{\partial \varepsilon_t}{\partial \theta'} - \frac{\partial \varepsilon_t^0}{\partial \theta'}, \frac{\partial^2}{\partial \theta \partial \theta'} \otimes (\varepsilon_t - \varepsilon_t^0) = \Delta(n^{-1/2+\delta} x_t) + w_t, \tag{82}$$

since both $A = A_0 + O(n^{-1/2+\delta})$ and $\Lambda = \Lambda_0 + O(n^{-1/2} + \delta)$. The results in (72)-(78) follow immediately from (82). The proof for (79)-(81) are more involved. For (80) and (81), we need to show

$$x_{t|t-1}^0 = x_t + w_t \tag{83}$$

$$x_{t|t-1} - x_{t|t-1}^0 = -n^{-1/2+\delta} x_t \Delta(n^{-1+2\delta} x_t) + w_t. \tag{84}$$

The result in (83) follows directly from (**??**). To establish (84), note that

$$\begin{aligned} x_{t|t-1} - x^0_{t|t-1} &= \frac{\partial x^0_{t|t-1}}{\partial (vecA)'}(vecA - vecA_0) + \frac{\partial x^0_{t|t-1}}{\partial (vec\Lambda)'}(vec\Lambda - vec\Lambda_0) + \Delta(n^{-1+2\delta}x_t) + w_t \\ &= -n^{-1/2+\delta}x_t\Delta(n^{-1+2\delta}x_t) + w_t, \end{aligned}$$

due to (**??**) and (49). Now it follows immediately from (83) and (84) that

$$\varepsilon_t - \varepsilon^0_t = -(A - A_0)x^0_{t|t-1} - A(x_{t|t-1} - x^0_{t|t-1}) = \Delta(n^{-1+2\delta}x_t) + w_t, \tag{85}$$

from which, together with (82), we may easily deduce (80) and (81).

Finally we prove (79). To do so, we first show that

$$(A_0' \otimes I_k)\frac{\partial x^0_{t|t-1}}{\partial (vecA)'} = -x_t + w_t \tag{86}$$

$$(A_0' \otimes I_k)\left(\frac{\partial x^0_{t|t-1}}{\partial (vecA)'} - \frac{\partial x^0_{t|t-1}}{\partial (vecA)'}\right) = 2n^{-1/2+\delta}x_t + \Delta(n^{-1+2\delta}) + w_t. \tag{87}$$

The result in (86) follows immediately from (49). To derive (87),
we note that

$$\begin{aligned} \frac{\partial x^0_{t|t-1}}{\partial (vecA)'} - \frac{\partial x^0_{t|t-1}}{\partial (vecA)'} &= \frac{\partial^2 x^0_{t|t-1}}{\partial vecA \partial (vecA)'}(vecA - vecA_0) + \frac{\partial^2 x^0_{t|t-1}}{\partial vecA \partial (vec\Lambda)'}(vec\Lambda - vec\Lambda_0) \\ &\quad + \Delta(n^{-1+2\delta}x_t) + w_t, \end{aligned}$$

and that *********** subsequent proof

Proof of Proposition 4.1 The stated result follows immediately from Lemma 2.2 and equation (**??**). Note that we have from Lemma 2.2

$$A_0 x^0_{t|t-1} = A_0(A_0'\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}y_t - \sum_{k=0}^{t-1} A_0(I_q - \Omega_0^{-1})^k(A_0\Lambda_0^{-1}A_0)^{-1}A_0'\Lambda_0^{-1}\triangle y_{t-k}$$

under the convention $x_0 = 0$, and

$$A_0 x^0_{t|t-1} = y_t - \varepsilon^0_t$$

due to the definition of (ε^0_t). □

Proof of Theorem 4.2 Throughout the proof, we let Λ be known and $\theta = vecA$. The proof of the general case with Λ unknown is essentially identical, only with added complexity in notation and exposition. In what follows, we denote by $\tilde{A}_n$, with its vectorized version $vec\tilde{A}_n$, the MLE of A for the case of $p = q$, i.e., there are p-common stochastic trends and no cointegration. Under this convention, we write the likelihood function as

$$\ell_n(A) = \ell(\tilde{A}_n) - \frac{1}{2}(vecA - vec\tilde{A}_n)'H_n(A, \tilde{A})(vecA - vec\tilde{A}_n), \tag{88}$$

where $H_n(A, \tilde{A})$ is the Hessian matrix evaluated at some value $\bar{A}$, say, of A which lie in the line connecting $\tilde{A}_n$ and A.

Under the null hypothesis, there are only $q < p$ stochastic common trends so that A is $p \times q$, instead of $p \times p$. In what follows, we denote by $\hat{A}$ and $vec\hat{A}_n$ the MLE of A under the null hypothesis. Note that $vec\hat{A}_n$ is pq-dimensional vector, whereas $vec\tilde{A}_n$ is p^2-dimensional. Moreover, we define

$$\hat{T}_N = \hat{B}_n \otimes I_p \quad \text{and} \quad \hat{T}_S = \hat{A}_n \left(\hat{A}_n' \hat{\Lambda}_n^{-1/2} \hat{A}_n \right)^{-1/2} \otimes I_p,$$

where $\hat{B}_n$ and $\hat{\Lambda}_n$ are the MLE's of B and Λ obtained together with $\hat{A}_n$ from the model with q-stochastic common trends. Finally, we let

$$\hat{T}_n = (\hat{T}_N, \hat{T}_S),$$

and

$$\nu_n = \text{diag}\,(nI_{p(p-q)}, \sqrt{n} I_{pq}),$$

a diagonal matrix. Note that

$$\hat{T}_n^{-1} = \begin{pmatrix} \hat{B}_n' \hat{\Lambda}_n^{-1} \\ (\hat{A}_n' \hat{\Lambda}_n^{-1} \hat{A}_n)^{-1/2} \hat{A}_n' \hat{\Lambda}_n^{-1} \end{pmatrix} \otimes I_p, \tag{89}$$

similarly as in (20).

As we have shown in the proof of Theorem 3.2,

$$\nu_n^{-1} \hat{T}_n' H_n(A, \tilde{A}) \hat{T}_n \nu_n^{-1} = \nu_n^{-1} \hat{T}_n' H_n(A_0) \hat{T}_n \nu_n^{-1} + o_p(1)$$

for large n. Therefore, we may write (88) as

$$\begin{aligned} &\ell_n(A) - \ell(\tilde{A}_n) \\ &= -\frac{1}{2} \left[(vecA - vec\tilde{A}_n)' \hat{T}_n^{-1\prime} \nu_n \right] \left[\nu_n^{-1} \hat{T}_n' H_n(A_0) \hat{T}_n \nu_n^{-1} \right] \left[\nu_n \hat{T}_n^{-1} (vecA - vec\tilde{A}_n) \right], \end{aligned} \tag{90}$$

up to the term of order $o_p(1)$, uniformly in a neighborhood of A_0. Furthermore, we define

$$\delta_n = \nu_n \hat{T}_n^{-1} vecA \quad \text{and} \quad \tilde{\delta}_n = \nu_n \hat{T}_n^{-1} vec\tilde{A}$$

and

$$M_n = \nu_n^{-1} \hat{T}_n' H_n(A_0) \hat{T}_n \nu_n^{-1},$$

and rewrite (90) as

$$\ell_n(A) - \ell(\tilde{A}_n) = \frac{1}{2} (\delta_n - \tilde{\delta}_n)' M_n (\delta_n - \tilde{\delta}_n), \tag{91}$$

which is maximized with respect to an $p \times q$-dimensional matrix A.

We know that the maximizer of A in (91) is given by $\hat{A}_n$, up to an error of order $o_p(1)$. Moreover, we have

$$\hat{B}_n' \hat{\Lambda}_n^{-1} \hat{A}_n = 0.$$

Therefore, if we define

$$\delta_n = (\delta'_{1n}, \delta'_{2n})',$$

then we may set the first $p(p-q)$-elements δ_{1n} of δ_n to be zero in our maximization problem (91) up to an error which is asymptotically negligible. Consequently, if we let

$$\tilde{\delta}_n = (\tilde{\delta}'_{1n}, \tilde{\delta}'_{2n})'$$

analogously with δ_n, and

$$M_n = \begin{pmatrix} M^n_{11} & M^n_{12} \\ M^n_{21} & M^n_{22} \end{pmatrix},$$

where the partition is made conformably with $\tilde{\delta}_n$ and δ_n, then it follows that

$$\tau_n = 2\left[\ell_n(\hat{A}_n) - \ell(\tilde{A}_n)\right] = \tilde{\delta}'_{1n}\left(M^n_{11} - M^n_{12}M^{n-1}_{22}M^n_{21}\right)\tilde{\delta}_{1n} + o_p(1), \tag{92}$$

from which we may easily derive the asymptotic distribution for the likelihood ratio test statistic τ_n.

As we have shown in the proof of Theorem 3.2, we have under the null hypothesis

$$\tilde{\delta}_{1n} \to_d \left(\int_0^1 V(r)\Sigma_0^{-1}V(r)'dr\right)^{-1}\int_0^1 V(r)\,dU(r)$$

and

$$M^n_{11} \to_d \int_0^1 V(r)\Sigma_0^{-1}V(r)'dr.$$

Moreover,

$$M^n_{12}, M^n_{21} \to_p 0$$

and M^n_{22} converges in probability to a positive definite matrix. Consequently, we have

$$\tau_n \to_d \int_0^1 dU(r)V(r)'\left(\int_0^1 V(r)\Sigma_0^{-1}V(r)'dr\right)^{-1}\int_0^1 V(r)\,dU(r) =_d \chi^2_{p(p-q)}$$

as $n \to \infty$, as was to be shown. The proof is therefore complete. $\square$