Automated Estimation of Vector Error Correction Models

Zhipeng Liao[*]   Peter C. B. Phillips[†]

Draft February, 2012

Abstract

Model selection and associated issues of post-model selection inference present well known challenges in empirical econometric research. These modeling issues are manifest in all applied work but they are particularly acute in multivariate time series settings such as cointegrated systems where multiple interconnected decisions can materially affect the form of the model and its interpretation. In cointegrated system modeling, empirical estimation typically proceeds in a stepwise manner that involves the determination of cointegrating rank and autoregressive lag order in a reduced rank vector autoregression followed by estimation and inference. This paper proposes an automated approach to cointegrated system modeling that uses adaptive shrinkage techniques to estimate vector error correction models with unknown cointegrating rank structure and unknown transient lag dynamic order. These methods enable simultaneous order estimation of the cointegrating rank and autoregressive order in conjunction with oracle-like efficient estimation of the cointegrating matrix and transient dynamics. As such they offer considerable advantages to the practitioner as an automated approach to the estimation of cointegrated systems. The paper develops the new methods, derives their limit theory, reports simulations and presents an empirical illustration.

Keywords: Adaptive shrinkage; Automation; Cointegrating rank, Lasso regression; Oracle efficiency; Transient dynamics; Vector error correction.

JEL classification: C22

[*] *University of California Los Angeles.* Support from an Anderson Fellowship at Yale University is gratefully acknowledged. Email: zhipeng-liao@hotmail.com

[†] *Yale University, University of Auckland, University of Southampton* and *Singapore Management University.* Support from the NSF under Grant No SES 09-56687 is gratefully acknowledged. Email: peter.phillips@yale.edu

1 Introduction

Cointegrated system modeling is now one of the main workhorses in empirical time series research. Much of this empirical research makes use of vector error correction (VECM) formulations. While there is often some prior information concerning the number of cointegrating vectors, most practical work involves (at least confirmatory) pre-testing to determine the cointegrating rank of the system as well as the lag order in the autoregressive component that embodies the transient dynamics. These order selection decisions can be made by sequential likelihood ratio tests (e.g. Johansen, 1988, for rank determination) or the application of suitable information criteria (Phillips, 1996). The latter approach offers several advantages such as joint determination of the cointegrating rank and autoregressive order, consistent estimation of both order parameters (Chao and Phillips, 1999), robustness to heterogeneity in the errors, and the convenience and generality of semi-parametric estimation in cases where the focus is simply the cointegrating rank (Cheng and Phillips, 2010). While appealing for practitioners, all of these methods are nonetheless subject to pre-test bias and post model selection inferential problems (Leeb and Pötscher, 2005).

The present paper explores a different approach. The goal is to liberate the empirical researcher from sequential testing procedures in inference about cointegrated systems and in policy work that relies on impulse responses. The ideas originate in recent work on sparse system estimation using shrinkage techniques such as lasso and bridge regression. These procedures utilize penalized least squares criteria in regression that can succeed, at least asymptotically, in selecting the correct regressors in a linear regression framework while consistently estimating the non-zero regression coefficients. While apparently effective asymptotically these procedures do not avoid post model selection inference issues in finite samples because the estimators implicitly carry effects from the implementation of shrinkage which can result in bias, multimodal distributions and difficulty discriminating local alternatives that can lead to unbounded risk (Leeb and Pötscher, 2008). On the other hand, the methods do radically simplify empirical research with large dimensional systems where order parameters must be chosen and sparsity is expected.

One of the contributions of this paper is to show how to develop adaptive versions of these shrinkage methods that apply in vector error correction modeling which by their nature involve reduced rank coefficient matrices and order parameters for lag polynomials and trend specifications. The implementation of these methods is not immediate. This is partly because of the nonlinearities involved in potential reduced rank structures and partly because of the interdependence of decision making concerning the form of the transient dynamics and the cointegrating rank structure. The paper designs a mechanism of estimation and selection that works through the eigenvalues of the levels coefficient matrix and the coefficient matrices of the transient dynamic components. The methods apply in quite general vector systems with unknown cointegrating rank structure and unknown lag dynamics. They permit simultaneous order estimation of the cointegrating rank and autoregressive order in conjunction with oracle-like efficient estimation of the cointegrating matrix and transient dynamics. As such they offer considerable advantages to the practitioner: in effect, it becomes unnecessary to implement pre-testing procedures because the empirical results reveal the order parameters as a consequence of the fitting procedure. In this sense, the methods provide an automated approach to the estimation of cointegrated systems. In the scalar case, the methods reduce to estimation in the presence or absence of a unit root and thereby implement an implicit unit root test procedure, as suggested in earlier work by Caner and Knight (2009).

The paper is organized as follows. Section 2 lays out the model and assumptions and shows how to implement adaptive shrinkage methods in VECM systems. Section 3 considers a simplified first order version of the VECM without lagged differences which reveals the approach to cointegrating rank selection and develops key elements in the limit theory. Here we show that the cointegrating rank r_o is identified by the number of zero eigenvalues of Π_o and the latter is consistently recovered by suitably designed shrinkage estimation. Section 4 extends this system and its asymptotics to the general case of cointegrated systems with weakly dependent errors. Here it is demonstrated that the cointegration rank r_o can be consistently selected despite the fact that Π_o may not be consistently estimable. Section 5 deals with the practically important case of a general VECM system driven by independent identically distributed (i.i.d.) shocks, where shrinkage estimation simultaneously performs consistent lag selection, cointegrating rank selection, and optimal estimation of the system coefficients. Section 6 considers adaptive selection of the tuning parameter. Section 7 reports some simulation findings and Section 8 provides an empirical illustration. Section 9 concludes and outlines some useful extensions of the methods and limit theory to other models. Proofs and some supplementary technical results are given in the Appendix.

2 Vector Error Correction and Adaptive Shrinkage

Throughout the paper we consider following parametric VECM representation of a cointegrated system

$$\Delta Y_t = \Pi_o Y_{t-1} + \sum_{j=1}^{p} B_{o,j} \Delta Y_{t-j} + u_t, \tag{2.1}$$

where $\Delta Y_t = Y_t - Y_{t-1}$, Y_t is an m-dimensional vector-valued time series, $\Pi_o = \alpha_o \beta_o'$ has rank $0 \leq r_o \leq m$, $B_{o,j}$ $(j = 1, ..., p)$ are $m \times m$ (transient) coefficient matrices and u_t is an m-vector error term with mean zero and nonsingular covariance matrix Σ_u. The rank r_o of Π_o is an order parameter measuring the cointegrating rank or the number of (long run) cointegrating relations in the system. The lag order p is a second order parameter, characterizing the transient dynamics in the system.

Conventional methods of estimation of (2.1) include reduced rank regression or maximum likelihood, based on the assumption of Gaussian u_t and a Gaussian likelihood. This approach relies on known r_o and known p, so implementation requires preliminary order parameter estimation. The system can also be estimated by unrestricted fully modified vector autoregression (Phillips, 1995), which leads to consistent estimation of the unit roots in (2.1), the cointegrating vectors and the transient dynamics. This method does not require knowledge of r_0 but does require knowledge of p. In addition, a semiparametric approach can be adopted in which r_o is estimated semiparametrically by order selection as in Cheng and Phillips (2010) followed by fully modified least squares regression to estimate the cointegrating matrix. This method achieves asymptotically efficient estimation of the long run relations (under Gaussianity) but does not estimate the transient relations.

The present paper explores the estimation of the parameters of (2.1) by Lasso-type regression, i.e. least squares (LS) regression with penalization. The resulting estimator is a shrinkage estimator. Specifically, the LS shrinkage estimator of (Π_o, B_o) where $B_o = (B_{o,1}, ..., B_{o,p})$ is defined as

$$\begin{aligned} (\widehat{\Pi}_n, \widehat{B}_n) \quad = \quad & \underset{\Pi \in R^{m\times m}, B_j \in R^{m\times m}}{\arg\min} \sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum\nolimits_{j=1}^{p} B_j \Delta Y_{t-j} \right\|^2 \\ & + n \sum\nolimits_{k=1}^{p} \widehat{P}_{\lambda_n}(B_j) + n \sum\nolimits_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\phi_k(\Pi)\right], \end{aligned} \tag{2.2}$$

where $\|\cdot\|$ is the Euclidean norm, $\widehat{P}_{\lambda_n}(\cdot)$ is penalty function with tuning parameter λ_n, and $\phi_k(\Pi)$ denotes the k-th largest modulus of the eigenvalues $\{\phi_k(\Pi)\}_{k=1}^{m}$ of the matrix

Π [1]. Let $S_{n,\phi} = \{k : \phi_k(\widehat{\Pi}_n) \neq 0\}$ be the index set of nonzero eigenvalues of $\widehat{\Pi}_n$. Since the eigenvalues of $\widehat{\Pi}_n$ are ordered as indicated, if there are $d_{\mathcal{S}_{n,\phi}}$ elements in $S_{n,\phi}$, then $S_{n,\phi} = \{m - d_{\mathcal{S}_{n,\phi}} + 1, ..., m\}$. Similarly, we index the zero components of $\widehat{\Pi}_n$ using the set $S^c_{n,\phi} = \{k : \phi_k(\widehat{\Pi}_n) = 0\}$ and set $S^c_{n,\phi} = \{1, ..., d_{\mathcal{S}_{n,\phi}}\}$. Given λ_n, this procedure delivers a one step estimator of the model (2.1) with an implied estimate of the cointegrating rank (based on the number of non-zero eigenvalues of $\widehat{\Pi}_n$ and an implied estimate of the transient dynamic order p and transient dynamic structure (that is, the non zero elements of B_o) based on the fitted value $\widehat{B}_n$.

In the literature of variable selection, there are many popular choices of the penalty function. One example for $\widehat{P}_{\lambda_n}(\cdot)$ is the adaptive Lasso penalty (Zou, 2006), defined as

$$\widehat{P}_{\lambda_n}(\phi) = \lambda_n \widehat{w}_\phi |\phi|, \tag{2.3}$$

where $\widehat{w}_\phi = |\widehat{\phi}_n|^{-\omega}$ with $\omega > 0$ and some consistent estimator $\widehat{\phi}_n$ of ϕ. Other examples are the bridge penalty (Knight, 2000), defined as

$$\widehat{P}_{\lambda_n}(\phi) = \lambda_n |\phi|^\gamma \tag{2.4}$$

where $\gamma \in (0, 1)$, and the SCAD penalty (Fan and Li, 2001), defined as

$$\widehat{P}_{\lambda_n}(\phi) = \begin{cases} \lambda_n|\phi| & |\phi| \leq \lambda_n \\ \frac{\lambda_n a|\phi|}{(a-1)} - \frac{\phi^2 + \lambda_n^2}{2(a-1)} & \lambda_n < |\phi| \leq a\lambda_n \\ \frac{(a+1)\lambda_n^2}{2} & a\lambda_n < |\phi| \end{cases}, \tag{2.5}$$

where a is some positive real number strictly larger than 2. In the context of variable selection, the LS shrinkage estimator with the above penalty functions has "oracle" properties in the sense that any zero coefficients are estimated as zeros with probability approaching 1(w.p.a.1). Correspondingly, the non-zero coefficients are estimated as if the true model were known and implemented in estimation (Fan and Li, 2001).

Let $\phi(\Pi) = (\phi_1(\Pi), ..., \phi_m(\Pi))$ denote the vector containing the m ordered eigenvalues of the matrix Π in $R^{m\times m}$. When $\{u_t\}_{t\geq 1}$ is an i.i.d. process or a martingale difference sequence, the LS estimators $(\widehat{\Pi}_{ols}, \widehat{B}_{ols})$ of (Π_o, B_o) are well known to be consistent. Thus

[1] Throughout this paper, for any $m \times m$ matrix Π, we order the eigenvalues of Π in ascending order by their moduli, i.e. $|\phi_1(\Pi)| \leq |\phi_2(\Pi)| \leq ... \leq |\phi_m(\Pi)|$. When there is a pair of complex conjugate eigenvalues, we order the one with a negative imaginary part before the other.

it seems intuitively clear that some form of adaptive penalization can be devised to consistently distinguish the zero and nonzero components in B_o and $\phi(\Pi_o)$. We show that the shrinkage LS estimator defined in (2.2) enjoys these oracle-like properties, in the sense that the zero components in B_o and $\phi(\Pi_o)$ are estimated as zeros w.p.a.1. Thus, Π_o and the non-zero elements in B_o are estimated as if the form of the true model were known and inferences can be conducted as if we knew the true cointegration rank r_o.

If the transient behavior of (2.1) is misspecified and (for some given lag order p) the error process $\{u_t\}_{t\geq 1}$ is weakly dependent and $r_o > 0$, then consistent estimators of (Π_o, B_o) are typically unavailable without further assumptions. However, we show here that the $m - r_o$ zero eigenvalues of Π_o can still be consistently estimated with an order n convergence rate, while the rest of the eigenvalues of Π_o are estimated with asymptotic bias at a $\sqrt{n}$ convergence rate. The different convergence rates of the eigenvalues are important, because when the non-zero eigenvalues of Π_o are occasionally (asymptotically) estimated as zeros, the different convergence rates are useful in consistently distinguishing the zero eigenvalues from the biasedly estimated non-zero eigenvalues of Π_o. Specifically, we show that if the estimator of some non-zero eigenvalue of Π_o has probability limit zero, then this estimator will converge in probability to zero at the rate $\sqrt{n}$, while estimates of the zero eigenvalues of Π_o all have convergence rate n. Hence the adaptive penalties associated with estimates of zero eigenvalues of Π_o will diverge to infinity at a rate faster than those of estimates of the nonzero eigenvalues of Π_o, even though the latter also converge to zero in probability. As we have prior knowledge about these different divergence rates in a potentially cointegrated system, we can impose explicit conditions on the convergence rate of the tuning parameter to ensure that only the estimates of zero eigenvalues of Π_o are adaptively shrunk to zero in finite samples.

For empirical implementation of our approach, we provide data-driven procedures for selecting the tuning parameter of the penalty function in finite samples. Our method is executed as follows.

(i) Solve the minimization problem in (6.2) to get the data determined tuning parameter $\widehat{\lambda}_n^*$.

(ii) Using the tuning parameter $\widehat{\lambda}_n^*$, obtain the LS shrinkage estimator $(\widehat{\Pi}_n, \widehat{B}_n)$ of (Π_o, B_o).

(iii) The cointegration rank selected by the shrinkage method is implied by the rank of the shrinkage estimator $\widehat{\Pi}_n$ and the lagged differences selected by the shrinkage method are implied by the nonzero matrices in $\widehat{B}_n$.

(iv) The LS shrinkage estimator $(\widehat{\Pi}_n, \widehat{B}_{\mathcal{S}_{n,B}})$ contains shrinkage bias introduced by the penalty on the nonzero eigenvalues of $\widehat{\Pi}_n$ and nonzero matrices in $\widehat{B}$. To remove this bias, run a second-step reduced rank regression based on the cointegration rank and the model selected by the LS shrinkage estimation given $\widehat{\lambda}_n^*$.

Notation is standard. For vector-valued, zero mean, covariance stationary stochastic processes $\{a_t\}_{t\geq 1}$ and $\{b_t\}_{t\geq 1}$, $\Gamma_{ab}(h) = E[a_t b'_{t+h}]$ and $\Lambda_{ab} = \sum_{h=0}^{\infty} \Gamma_{ab}(h)$ denote the lag h autocovariance matrix and one-sided long-run covariance matrix. The symbolism $a_n \asymp b_n$ means that $(1 + o_p(1))b_n = a_n$ or vice versa; the expression $a_n = o_p(b_n)$ signifies that $\Pr(|a_n/b_n| \geq \epsilon) \to 0$ for all $\epsilon > 0$ as n go to infinity; and $a_n = O_p(b_n)$ when $\Pr(|a_n/b_n| \geq M) \to 0$ as n and M go to infinity. As usual, "$\to_p$" and "$\to_d$" imply convergence in probability and convergence in distribution, respectively.

3 The First Order VECM

This section considers the following simplified first order version of (2.1),

$$\Delta Y_t = \Pi_o Y_{t-1} + u_t = \alpha_o \beta_o' Y_{t-1} + u_t. \tag{3.1}$$

The model contains no deterministic trend and no lagged differences. Our focus in this simplified system is to outline the approach to cointegrating rank selection and develop key elements in the limit theory, showing consistency in rank selection and reduced rank coefficient matrix estimation. The theory is then generalized in subsequent sections.

We impose the following assumptions on the innovation u_t.

Assumption 3.1 (WN) $\{u_t\}_{t\geq 1}$ *is an m-dimensional i.i.d. process with mean zero, and nonsingular covariance matrix* Ω_u.

Assumption 3.1 ensures that the parameter matrix Π_o is consistently estimable in this simplified system. It could be relaxed to allow for martingale difference innovations. Under 3.1, partial sums of u_t satisfy the functional law

$$n^{-\frac{1}{2}} \sum_{t=1}^{[n\cdot]} u_t \to_d B_u(\cdot), \tag{3.2}$$

where $B_u(\cdot)$ is vector Brownian motion with variance matrix Ω_u.

Assumption 3.2 (RR) *(i) The determinantal equation $|I-(I+\Pi_o)\lambda|=0$ has roots on or outside the unit circle; (ii) the matrix Π_o has rank r_o, with $0\leq r_o\leq m$; (iii) if $r_o>0$, then the matrix $R=I_{r_o}+\beta_o'\alpha_o$ has eigenvalues within the unit circle.*

As $\Pi_o=\alpha_o\beta_o'$ has rank r_o, we can choose α_o and β_o to be $m\times r_o$ matrices with full rank. When $r_o=0$, we simply take $\Pi_o=0$. Let $\alpha_{o,\perp}$ and $\beta_{o,\perp}$ be the matrix orthogonal complements of α_o and β_o and, without loss of generality, assume that $\alpha_{o,\perp}'\alpha_{o,\perp}=I_{m-r_o}$ and $\beta_{o,\perp}'\beta_{o,\perp}=I_{m-r_o}$. Assumption 3.2.(iii) implies $\beta_o'\alpha_o$ is nonsingular, which further implies that $\alpha_{o,\perp}'\beta_{o,\perp}$ is nonsingular.

Suppose $\Pi_o\neq 0$ and define $Q=[\beta_o,\alpha_{o\perp}]'$. In view of the well known relation (e.g., Johansen, 1995)

$$\alpha_o(\beta_o'\alpha_o)^{-1}\beta_o'+\beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\alpha_{o,\perp}'=I_m,$$

it follows that $Q^{-1}=\left[\alpha_o(\beta_o'\alpha_o)^{-1},\beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\right]$ and then

$$Q\Pi_oQ^{-1}=\begin{pmatrix}\beta_o'\alpha_o & 0\\ 0 & 0\end{pmatrix}. \tag{3.3}$$

Cointegrating rank is the number r_o of non-zero eigenvalues of Π_o and $\beta_{o,\perp}$ is a matrix of normalized eigenvectors corresponding to those eigenvalues. When $\Pi_o=0$, then the result holds trivially with $r_o=0$ and $\beta_{o,\perp}=I_m$.

Let $S_\phi=\{k:\phi_k(\Pi_o)\neq 0\}$ be the index set of nonzero eigenvalues of Π_o and similarly $S_\phi^c=\{k:\phi_k(\Pi_o)=0\}$ denote the index set of zero eigenvalues of Π_o. By the ordering of the eigenvalues of Π_o, we know that $S_\phi=\{m-d_{\mathcal{S}_\phi}+1,...,m\}$ and $S_\phi^c=\{1,...,m-d_{\mathcal{S}_\phi}\}$, where $d_{\mathcal{S}_\phi}$ is the cardinality of the set S_ϕ. Based on the above discussion, the following equality holds

$$d_{\mathcal{S}_\phi}=r_o. \tag{3.4}$$

Consistent selection of the rank of Π_o is equivalent to the consistent recovery of the zero components in $\phi(\Pi_o)$. We show that the latter is achieved by the sparsity of our shrinkage estimator of the eigenvalues of Π_o.

Using the matrix Q, we can transform (3.1) as

$$\Delta Z_t=\Xi_oZ_{t-1}+w_t, \tag{3.5}$$

where

$$Z_t = \begin{pmatrix} \beta_o' Y_t \\ \alpha_{o,\perp}' Y_t \end{pmatrix} \equiv \begin{pmatrix} Z_{1,t} \\ Z_{2,t} \end{pmatrix}, \; w_t = \begin{pmatrix} \beta_o' u_t \\ \alpha_{o,\perp}' u_t \end{pmatrix} \equiv \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix}$$

and $\Xi_o = Q\Pi_o Q^{-1}$. Assumption 3.2 leads to the following Wold representation for $Z_{1,t}$

$$Z_{1,t} = \beta_o' Y_t = \sum_{i=0}^{\infty} R^i \beta_o' u_{t-i} = R(L)\beta_o' u_t, \tag{3.6}$$

and the partial sum Granger representation,

$$Y_t = C \sum_{s=1}^{t} u_s + \alpha_o(\beta_o'\alpha_o)^{-1} R(L)\beta_o' u_t + CY_0, \tag{3.7}$$

where $C = \beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\alpha_{o,\perp}'$. Under Assumption 3.2 and (3.2), we have the functional law

$$n^{-\frac{1}{2}} \sum_{t=1}^{[n\cdot]} w_t \to_d B_w(\cdot) = QB_u(\cdot) = \begin{bmatrix} \beta_o' B_u(\cdot) \\ \alpha_{o,\perp}' B_u(\cdot) \end{bmatrix} \equiv \begin{bmatrix} B_{w_1}(\cdot) \\ B_{w_2}(\cdot) \end{bmatrix}$$

for $w_t = Qu_t$, so that

$$n^{-\frac{1}{2}} \sum_{t=1}^{[n\cdot]} Z_{1,t} = n^{-\frac{1}{2}} \sum_{t=1}^{[n\cdot]} \beta_o' Y_t \to_d -(\beta_o'\alpha_o)^{-1} B_{w_1}(\cdot), \tag{3.8}$$

since $R(1) = \sum_{i=0}^{\infty} R^i = (I - R)^{-1} = -(\beta_o'\alpha_o)^{-1}$. Also

$$n^{-1} \sum_{t=1}^{n} Z_{1,t-1} Z_{1,t-1}' = n^{-1} \sum_{t=1}^{n} \beta_o' Y_{t-1} Y_{t-1}' \beta_o \to_p \Sigma_{z_1 z_1},$$

where $\Sigma_{z_1 z_1} \equiv Var\left[\beta_o' Y_t\right] = \sum_{i=0}^{\infty} R^i \beta_o' \Omega_u \beta_o R^{i\prime}$.

The shrinkage LS estimator $\widehat{\Pi}_n$ of Π_o is defined as

$$\widehat{\Pi}_n = \arg \min_{\Pi \in R^{m\times m}} \sum_{t=1}^{n} \|\Delta Y_t - \Pi Y_{t-1}\|_E^2 + n \sum_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi)\right]. \tag{3.9}$$

The unrestricted LS estimator $\widehat{\Pi}_{ols}$ of Π_o is

$$\widehat{\Pi}_{ols} = \arg\min_{\Pi \in R^{m\times m}} \sum_{t=1}^{n} \|\Delta Y_t - \Pi Y_{t-1}\|_E^2 = \left(\sum_{t=1}^{T} \Delta Y_t Y_{t-1}'\right)\left(\sum_{t=1}^{T} Y_{t-1} Y_{t-1}'\right)^{-1}. \quad (3.10)$$

Let $\phi(\widehat{\Pi}_{ols}) = [\phi_1(\widehat{\Pi}_{ols}), ..., \phi_k(\widehat{\Pi}_{ols})]$ be the vector of ordered eigenvalues of $\widehat{\Pi}_{ols}$. The asymptotic properties of $\widehat{\Pi}_{ols}$ and its eigenvalues are described in the following result.

Lemma 3.1 *Under Assumptions 3.1 and 3.2, we have:*

(a) $\widehat{\Pi}_{ols}$ *satisfies*

$$Q\left(\widehat{\Pi}_{ols} - \Pi_o\right) Q^{-1} D_n^{-1} = O_p(1) \quad (3.11)$$

where $D_n = diag(n^{-\frac{1}{2}} I_{r_o}, n^{-1} I_{m-r_o})$;

(b) the eigenvalues of $\widehat{\Pi}_{ols}$ *satisfy*

$$\left[\phi_1(\widehat{\Pi}_{ols}), ..., \phi_{m-r_o}(\widehat{\Pi}_{ols})\right] \rightarrow_p (0, ..., 0)_{1\times(m-r_o)} \quad (3.12)$$

and

$$\left[\phi_{m-r_o+1}(\widehat{\Pi}_{ols}), ..., \phi_m(\widehat{\Pi}_{ols})\right] \rightarrow_p \left[\phi_{m-r_o+1}(\Pi_o), ..., \phi_m(\Pi_o)\right]; \quad (3.13)$$

(c) the first $m - r_o$ *ordered eigenvalues of* $\widehat{\Pi}_{ols}$ *satisfy*

$$n\left[\phi_1(\widehat{\Pi}_{ols}), ..., \phi_{m-r_o}(\widehat{\Pi}_{ols})\right] \rightarrow_d \left(\widetilde{\phi}_{o,1}, ..., \widetilde{\phi}_{o,m-r_o}\right), \quad (3.14)$$

where the $\widetilde{\phi}_{o,j}$ $(j = 1, ..., m - r_o)$ *are solutions of the following determinantal equation*

$$\left|\mu I_{m-r_0} - \left(\int dB_{w_2} B_{w_2}'\right)\left(\int B_{w_2} B_{w_2}'\right)^{-1}\right| = 0, \quad (3.15)$$

where $B_{w_2} = \alpha_{o,\perp}' B_u$ *is Brownian motion with covariance matrix* $\alpha_{o,\perp}' \Omega_u \alpha_{o,\perp}$.

The results of Lemma 3.1 are useful in constructing the adaptive Lasso penalty function because the OLS estimate $\widehat{\Pi}_{ols}$ and the related eigenvalue estimates $\phi_k(\widehat{\Pi}_{ols})$ of $\phi_k(\Pi_o)$ $(k = 1, ..., m)$ can be used as first step estimates in the penalty function. The convergence rates of $\widehat{\Pi}_{ols}$ and $\phi_k(\widehat{\Pi}_{ols})$ are important for delivering consistent model selection and cointegrated rank selection. In the univariate case where $m = 1$ and $r_o = 0$, we immediately have (see Lemma 1.(iii) in the Appendix)

$$n\phi_{\mathcal{S}_\phi^c}(\widehat{\Pi}_{ols}) \to_d \frac{\int B_u(a)dB_u(a)}{\int B_u^2(a)da}, \tag{3.16}$$

coinciding with the asymptotic distribution of the LS estimator in a regression with a unit-root.

The following assumptions provide some regularity conditions on the penalty function $\widehat{P}_{\lambda_n}(\cdot)$, which are sufficient for establishing consistency and the convergence rate of $\widehat{\Pi}_n$, as well as the sparsity of $\phi(\widehat{\Pi}_n)$.

Assumption 3.3 (PF-I) *(i) The penalty function $\widehat{P}_{\lambda_n}(\cdot)$ is non-negative with $\widehat{P}_{\lambda_n}(0) = 0$ and satisfies*

$$\widehat{P}_{\lambda_n}(\phi) = o_p(1) \tag{3.17}$$

for all $\phi \in (0, \infty)$; (ii) $\widehat{P}_{\lambda_n}(\phi)$ is twice continuously differentiable at $\phi \in (0, \infty)$ with $\widehat{P}''_{\lambda_n}(\phi) = o_p(1)$ for all $\phi \in (0, \infty)$; (iii) $\widehat{P}_{\lambda_n}(\cdot)$ satisfies

$$n^{\frac{1}{2}} \left| \widehat{P}'_{\lambda_n}(\phi) \right| = o_p(1), \tag{3.18}$$

for any $\phi \in (0, \infty)$; (iv) $\widehat{P}_{\lambda_n}(\cdot)$ is an increasing function in $(0, +\infty)$ and for each $k \in \mathcal{S}_\phi^c$, there exists some sequence $v_{n,k} = o(1)$ such that

$$\liminf_{n\to\infty} \left[n\widehat{P}_{\lambda_n}(n^{-1}v_{n,k}) \right] \to \infty \text{ } a.e. \tag{3.19}$$

Assumption 3.3.(i) essentially states that the shrinkage effect on estimates of the non-zero eigenvalues goes to zero asymptotically. This requirement is important for establishing the consistency of the shrinkage estimator $\widehat{\Pi}_n$. Assumption 3.3.(ii) is a useful regularity condition in deriving the convergence rate of $\widehat{\Pi}_n$ and Assumption 3.3.(iii) assures a $\sqrt{n}$ convergence rate. Assumption 3.3.(iv) ensures that the penalty $\widehat{P}_{\lambda_n}(\cdot)$ attaches to estimates of the zero eigenvalues diverges to infinity, which ensures the zero eigenvalues of Π_o are estimated with zero w.p.a.1 and consistent cointegration rank selection. Similar general assumptions are imposed on the penalty function $\widehat{P}_{\lambda_n}(\cdot)$ in Liao (2010), where shrinkage techniques are employed to perform consistent moment selection in a GMM framework.

Theorem 3.1 (Consistency) *Suppose Assumptions WN, RR and PF-I.(i) are satisfied. Then the shrinkage LS estimator $\widehat{\Pi}_n$ is consistent, i.e. $\widehat{\Pi}_n - \Pi_o = o_p(1)$.*

When consistent shrinkage estimators are considered, Theorem 3.1 extends Theorem 1 of Caner and Knight (2009), where shrinkage techniques are used to perform a unit root test. As the eigenvalues $\phi(\Pi)$ of the matrix Π are continuous functions of Π, we deduce from the consistency of $\widehat{\Pi}_n$ and continuous mapping that $\phi_k(\widehat{\Pi}_n) \to_p \phi_k(\Pi_o)$ for all $k = 1, ..., m$. Theorem 3.1 implies that the nonzero eigenvalues of Π_o are estimated as non-zeros, which means that the rank of Π_o will not be under-selected. However, consistency of the estimates of the non-zero eigenvalues is not necessary for consistent cointegration rank selection. In that case what is essential is that the probability limits of the estimates of those (non-zero) eigenvalues are not zeros or at least that their convergence rates are slower than those of estimates of the zero eigenvalues. This point will be pursued in the following section where it is demonstrated that consistent estimation of the cointegrating rank continues to hold for weakly dependent innovations $\{u_t\}_{t\geq 1}$ even though full consistency of $\widehat{\Pi}_n$ does not generally apply in that case.

Our next result gives the convergence rate of the shrinkage estimator $\widehat{\Pi}_n$.

Theorem 3.2 (Rate of Convergence) *Denote* $\delta_n = \max_{k\in\mathcal{S}_\phi} \left|\widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right|$. *Under Assumption WN, RR and Assumption PF.(i)-(ii), the shrinkage LS estimator* $\widehat{\Pi}_n$ *satisfies the following:*

(a) if $r_o = 0$*, then* $\widehat{\Pi}_n - \Pi_o = O_p(n^{-1} + n^{-1}\delta_n)$*;*

(b) if $0 < r_o \leq m$*, then* $\left(\widehat{\Pi}_n - \Pi_o\right) Q^{-1} D_n^{-1} = O_p(1 + n^{\frac{1}{2}}\delta_n)$.

The term δ_n represents the shrinkage bias that the penalty function introduces to the LS shrinkage estimator. If the convergence rate of λ_n is fast enough such that Assumption PF.(iii) is satisfied, then Theorem 3.2 implies that $\widehat{\Pi}_n - \Pi_o = O_p(n^{-1})$ when $r_o = 0$ and $\left(\widehat{\Pi}_n - \Pi_o\right) Q^{-1} D_n^{-1} = O_p(1)$ otherwise. Hence, under Assumption WN, RR and Assumption PF.(i)-(iii), the LS shrinkage estimator $\widehat{\Pi}_n$ has the same stochastic properties of the LS estimator $\widehat{\Pi}_{ols}$. However, we next show that if the tuning parameter λ_n converges to zero not too fast and such that Assumptions PF.(iv) is also satisfied, then the correct rank restriction $r = r_o$ is automatically imposed on the LS shrinkage estimator $\widehat{\Pi}_n$ w.p.a.1.

Recall that $S_{n,\phi}$ is the index set of the largest (or last) r_o ordered eigenvalues of $\widehat{\Pi}_n$ and $S^c_{n,\phi}$ is the index set of the smallest (or first) $m - r_o$ ordered eigenvalues of $\widehat{\Pi}_n$. By virtue of the consistency of $\phi(\widehat{\Pi}_n) = \left[\phi_1(\widehat{\Pi}_n), ..., \phi_m(\widehat{\Pi}_n)\right]$, we can deduce that $S^c_{n,\phi} \subseteq S^c_\phi$. To show $S^c_{n,\phi} = S^c_\phi$, it is suffices to show that $S^c_\phi \subseteq S^c_{n,\phi}$.

Theorem 3.3 (Sparsity-I) *Under Assumption WN, RR and Assumption PF-I,*

$$\Pr\left(\phi_k(\widehat{\Pi}_n) = 0\right) \to 1 \text{ as } n \to \infty \tag{3.20}$$

for all $k \in \mathcal{S}_{\phi}^{c}$.

Proof. By the definition of $\widehat{\Pi}_n$,

$$\sum_{t=1}^{n}\left(\left\|\Delta Y_t - \widehat{\Pi}_n Y_{t-1}\right\|_E^2 - \|\Delta Y_t - \Pi_o)Y_{t-1}\|_E^2\right) \leq n\sum_{k=1}^{m}\left\{\widehat{P}_{\lambda_n}\left[\bar{\phi}_k\left(\Pi_o\right)\right)\right] - \widehat{P}_{\lambda_n}\left[\bar{\phi}_k\left(\widehat{\Pi}_n\right)\right]\right\}. \tag{3.21}$$

The left-hand side of the inequality in (3.21) can be rewritten as

$$\begin{aligned} L_n \equiv & \left[vec\left(\widehat{\Pi}_n - \Pi_o\right)\right]'\left(\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' \otimes I_m\right)\left[vec\left(\widehat{\Pi}_n - \Pi_o\right)\right] \\ & - 2\left[vec\left(\widehat{\Pi}_n - \Pi_o\right)\right]' vec\left(\sum_{t=1}^{n} Y_{t-1}u_t'\right) \end{aligned} \tag{3.22}$$

By Lemma 10.1 and Theorem 3.2, we can deduce that $L_n = O_p(1)$.

On the event $\left\{\phi_{k_o}(\widehat{\Pi}_n) \neq 0\right\}$ for some $k_o \in S_{\phi}^{c}$, the right-hand side of the inequality in (3.21) can be bounded in the following way

$$\begin{aligned} R_n \equiv & \; n\sum_{k\in S_{\phi}}\left\{\widehat{P}_{\lambda_n}\left[\bar{\phi}_k\left(\Pi_o\right)\right] - \widehat{P}_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)]\right\} - n\sum_{k\in S_{\phi}^{c}}\widehat{P}_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)] \\ \leq & \; -n\widehat{P}_{\lambda_n}[\bar{\phi}_{k_o}(\widehat{\Pi}_n)] + n\sum_{k\in S_{\phi}}\left\{\widehat{P}_{\lambda_n}\left[\bar{\phi}_k\left(\Pi_o\right)\right] - \widehat{P}_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)]\right\}. \end{aligned} \tag{3.23}$$

Using Assumptions PF.(ii)-(iii), the Bauer-Fiker Theorem and Theorem 3.2, we have

$$\begin{aligned} & \left\|n\sum_{k\in S_{\phi}}\left\{\widehat{P}_{\lambda_n}\left[\bar{\phi}_k\left(\Pi_o\right)\right] - \widehat{P}_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)]\right\}\right\| \\ \leq & \; n^{\frac{1}{2}}\widehat{P}_{\lambda_n}'\left[\bar{\phi}_k\left(\Pi_o\right)\right]\left\|n^{\frac{1}{2}}(\widehat{\Pi}_n - \Pi_o)\right\|_E \\ & + [1 + o_p(1)]\,\widehat{P}_{\lambda_n}''\left[\bar{\phi}_k\left(\Pi_o\right)\right]\left\|n^{\frac{1}{2}}(\widehat{\Pi}_n - \Pi_o)\right\|_E^2 \\ = & \; O_p(1). \end{aligned} \tag{3.24}$$

Suppose that the event $\{\phi_{k_o}(\widehat{\Pi}_n) \neq 0\}$ has nonzero probability measure as $n \to \infty$. Conditional on this event, the restriction $\phi_{k_o}(\Pi) = 0$ is not imposed on the LS estimator $\widehat{\Pi}_n$ asymptotically. Hence we can use Assumptions PF.(iv) and Theorem 4.1 in Liao and Phillips (2010) to deduce that

$$n\widehat{P}_{\lambda_n}[\bar{\phi}_{k_o}(\widehat{\Pi}_n)] = n\widehat{P}_{\lambda_n}\left\{n^{-1}n\bar{\phi}_{k_o}(\widehat{\Pi}_n)\right\} \geq n\widehat{P}_{\lambda_n}\left[n^{-1}v_n\right] \to_p \infty \tag{3.25}$$

as $n \to \infty$. From the results in (3.23)-(3.25), we have $R_n \to -\infty$ as $n \to \infty$ and $L_n > R_n$ with nontrivial probability, which contradicts the definition of $\widehat{\Pi}_n$. Hence, it follows that $\Pr\left(\phi_k(\widehat{\Pi}_n) = 0\right) \to 1$ for all $k \in \mathcal{S}_\phi^c$, as $n \to \infty$. ■

Theorem 3.1 and the CMT imply that

$$\phi_k(\widehat{\Pi}_n) \to_p \phi_k(\Pi_o). \tag{3.26}$$

Combining Theorem 3.1 and Theorem 3.3, we deduce that

$$\Pr(S_{n,\phi} = S_\phi) \to 1, \tag{3.27}$$

which implies consistent cointegration rank selection, giving the following result.

Corollary 3.4 *Under Assumption WN, RR and PF-I, we have*

$$\Pr\left(r(\widehat{\Pi}_n) = r_o\right) \to 1 \tag{3.28}$$

as $n \to \infty$, where $r(\widehat{\Pi}_n)$ denotes the rank of $\widehat{\Pi}_n$.

It is clear that Assumption PF-I plays an important role in deriving consistent cointegration rank selection. We next show that the adaptive Lasso and bridge penalty functions satisfy Assumption PF-I; the SCAD penalty satisfies Assumptions PF-I.(i)-(iii) but fails to satisfy Assumptions PF-I.(iv).

Corollary 3.5 *The adaptive Lasso, bridge and SCAD penalty functions have the following properties:*

(a) If $\lambda_n = o(1)$, then the adaptive Lasso and bridge penalty functions satisfy Assumptions PF.(i)-(ii) and the SCAD penalty function satisfies Assumptions PF.(i)-(iii);

(b) if $n^{\frac{1}{2}}\lambda_n = o(1)$*, then the adaptive Lasso and bridge penalty functions satisfy Assumptions PF.(ii)-(iii);*

(c) if $\lambda_n n^{\omega} \to \infty$*, then the adaptive Lasso penalty satisfies Assumption PF.(v); if* $n^{1-\gamma}\lambda_n \to \infty$*, then the bridge penalty satisfies Assumption PF.(v);*

(d) the SCAD penalty function does not satisfy Assumption PF.(iii).

Proof. (a). First it is clear that the adaptive Lasso, bridge and SCAD penalty functions all satisfy $\widehat{P}_{\lambda_n}(0) = 0$. For the adaptive Lasso penalty,

$$\widehat{P}_{\lambda_n}[\bar{\phi}_k(\Pi_o)] = \lambda_n |\phi_k(\widehat{\Pi}_{ols})|^{-\omega} |\phi_k(\Pi_o)| = o_p(1)$$

for all $k \in \mathcal{S}_\phi$, by the consistency of $\phi_k(\widehat{\Pi}_{ols})$ and the Slutsky Theorem. The adaptive Lasso penalty is clearly twice differentiable with $\widehat{P}''_{\lambda_n}(\phi) = 0$ for all $\phi \in (0, \infty)$, and thus satisfies Assumptions PF.(i)-(ii). For the bridge penalty,

$$\lambda_n \widehat{P}_{\lambda_n}[\bar{\phi}_k(\Pi_o)] = \lambda_n |\phi_k(\Pi_o)|^{\gamma} = o(1),$$

for all k and

$$\widehat{P}''_{\lambda_n}(\phi) = (\gamma - 1)\lambda_n |\phi|^{\gamma-2} \to 0,$$

for any $\phi \in (0, \infty)$. Hence the bridge penalty satisfies Assumptions PF.(i)-(ii). For the SCAD penalty, when n is sufficiently large, we have

$$\widehat{P}_{\lambda_n}[\bar{\phi}_k(\Pi_o)] = \frac{(a+1)\lambda_n^2}{2} = o(1),$$

for all $k \in \mathcal{S}_\phi$. Furthermore when n is sufficiently large, the SCAD penalty function is clearly twice differentiable at $\phi_k(\Pi_o)$ ($k \in \mathcal{S}_\phi$) with

$$\sqrt{n}\widehat{P}'_{\lambda_n}[\bar{\phi}_k(\Pi_o)] = \sqrt{n}\frac{[\lambda_n a - |\phi_k(\Pi_o)|]_+}{(a-1)} I(\beta > \lambda_n) = 0, \tag{3.29}$$

where $[x]_+ = \max\{0, x\}$, and $\widehat{P}''_{\lambda_n}[\bar{\phi}_k(\Pi_o)] = 0$ for all $k \in \mathcal{S}_\phi$. Hence the SCAD penalty satisfies Assumptions PF.(i)-(iii).

(b). For the adaptive Lasso penalty,

$$n^{\frac{1}{2}}\widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right] = n^{\frac{1}{2}}\lambda_n |\phi_k(\widehat{\Pi}_{ols})|^{-\omega} = o_p(1),$$

where the last equality is by the consistency of $\widehat{\phi}_{k,n}$ and Slutsky Theorem. For the bridge penalty,

$$n^{\frac{1}{2}}\widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right] = n^{\frac{1}{2}}\lambda_n|\phi_k(\Pi_o)|^{\gamma-1} = o(1).$$

Hence both the adaptive Lasso penalty and bridge penalty functions satisfy Assumptions PF.(iii).

(c) For the adaptive Lasso penalty, by Lemma **3.1**, we can choose $v_{n,k} = n^{-\frac{\omega}{2}}\lambda_n^{-\frac{1}{2}} = o(1)$ such that

$$n\widehat{P}_{\lambda_n}\left(n^{-1}v_{n,k}\right) = |n\phi_k(\widehat{\Pi}_n)|^{-\omega}|n^{\omega}\lambda_n v_{n,k}| = |n\phi_k(\widehat{\Pi}_n)|^{-\omega}|n^{\frac{\omega}{2}}\lambda_n^{\frac{1}{2}}| \to_p \infty,$$

for all $k \in \mathcal{S}_{n,\phi}^c$. Thus the adaptive Lasso penalty satisfies Assumption PF.(iv). For the bridge penalty, we can choose $v_n = (n^{1-\gamma}\lambda_n)^{-\frac{1}{2\gamma}} = o(1)$ such that

$$n\widehat{P}_{\lambda_n}\left(n^{-1}v_{n,k}\right) = n^{1-\gamma}\lambda_n|v_{n,k}|^{\gamma} = \left(n^{1-\gamma}\lambda_n\right)^{\frac{1}{2}} \to \infty,$$

for all $k \in \mathcal{S}_{n,\phi}^c$. Hence the bridge penalty function satisfies Assumption PF.(iii).

(d) For the SCAD penalty, if $n\lambda_n \to \infty$ and $\lambda_n \to 0$, then

$$n\widehat{P}_{\lambda_n}\left(n^{-1}v_{n,k}\right) = \lambda_n|v_{n,k}| \to 0,$$

for any $v_{n,k} = o(1)$. If $n\lambda_n \to 0$, then

$$n\widehat{P}_{\lambda_n}\left(n^{-1}v_{n,k}\right) = \frac{a\lambda_n|v_{n,k}|}{(a-1)} - \frac{n^{-1}v_{n,k}^2 + n\lambda_n^2}{2(a-1)} \to 0,$$

or

$$n\widehat{P}_{\lambda_n}\left(n^{-1}v_{n,k}\right) = \frac{(a+1)n\lambda_n^2}{2} \to 0.$$

Hence there exists no sequence $v_{n,k}$ such that Assumption PF.(iii) is satisfied. ■

From the sparsity and consistency of $\phi_k(\widehat{\Pi}_n)$ $(k = 1, ..., m)$, we can deduce that the rank constraint $r(\Pi) = r_o$ is imposed on the LS shrinkage estimator $\widehat{\Pi}_n$ w.p.a.1. Hence, for large enough n the shrinkage estimator $\widehat{\Pi}_n$ can be decomposed as $\widehat{\alpha}_n\widehat{\beta}_n'$ w.p.a.1, where $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ are some $m \times r_o$ matrices. Without loss of generality, we assume the first r_o columns of Π_o are linearly independent. To ensure identification, we normalize β_o as $\beta_o = [I_{r_o}, O_{r_o}]'$

where O_{r_o} is some $r_o \times (m - r_o)$ matrix such that

$$\Pi_o = \alpha_o \beta_o' = [\alpha_o, \alpha_o O_{r_o}]. \tag{3.30}$$

Hence α_o is the first r_o columns of Π_o which is an $m \times r_o$ matrix with full rank and O_{r_o} is uniquely determined by the equation $\alpha_o O_{r_o} = \Pi_{o,2}$, where $\Pi_{o,2}$ denotes the last $m - r_o$ columns of Π_o. Correspondingly, for large enough n we can normalize $\widehat{\beta}_n$ as $\widehat{\beta}_n = [I_{r_o}, \widehat{O}_n]'$ where $\widehat{O}_n$ is some $r_o \times (m - r_o)$ matrix.

From Theorem 3.2 and Assumption 3.3.(ii), we have

$$O_p(1) = \left(\widehat{\Pi}_n - \Pi_o\right) Q^{-1} D_n = \left(\widehat{\Pi}_n - \Pi_o\right) \left[\sqrt{n}\alpha_o(\beta_o'\alpha_o)^{-1}, n\beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\right]. \tag{3.31}$$

From (3.31), we can deduce that

$$n\widehat{\alpha}_n \left(\widehat{\beta}_n - \beta_o\right)' \beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1} = O_p(1),$$

which implies

$$n\left(\widehat{O}_n - O_o\right) = O_p(1). \tag{3.32}$$

Similarly, we have

$$\sqrt{n}\left[\left(\widehat{\alpha}_n - \alpha_o\right)\widehat{\beta}_n'\alpha_o(\beta_o'\alpha_o)^{-1} - \alpha_o\left(\widehat{\beta}_n - \beta_o\right)'\alpha_o(\beta_o'\alpha_o)^{-1}\right] = O_p(1),$$

which combined with (3.32) implies

$$\sqrt{n}\left(\widehat{\alpha}_n - \alpha_o\right) = O_p(1). \tag{3.33}$$

To conclude this section, we give the centered limit distribution of the shrinkage estimator $\widehat{\Pi}_n$ in the following theorem.

Theorem 3.6 *Under Assumption WN, RR and PF, we have*

$$Q\left(\widehat{\Pi}_n - \Pi_o\right) Q^{-1} D_n^{-1} \to_d \begin{pmatrix} \beta_o' B_{1m} \Sigma_{z_1 z_1}^{-1} & \Gamma_o B_{2,m} \left(\int B_{w_2} B_{w_2}'\right)^{-1} \\ \alpha_{o,\perp}' B_{1m} \Sigma_{z_1 z_1}^{-1} & 0 \end{pmatrix} \tag{3.34}$$

where $B_{1m} = N(0, \Omega_u \otimes \Sigma_{z_1 z_1})$, $B_{2,m} = \left(\int B_{w_2} dB_u'\right)'$ *and* $\Gamma_o = \beta_o'\alpha_o(\alpha_o'\alpha_o)^{-1}\alpha_o'$.

Proof. By the sparsity of $\phi_k(\widehat{\Pi}_n)$ $(k=1,...,m)$, we can deduce that $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ minimize the following criterion function w.p.a.1,

$$V_{3,n}(\alpha,\beta)=\sum_{t=1}^{n}\left\|\Delta Y_t-\alpha\beta' Y_{t-1}\right\|_E^2+n\sum_{k\in\mathcal{S}_\phi}\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\alpha\beta')\right]. \tag{3.35}$$

Define $U_{1,n}^*=\sqrt{n}\left(\widehat{\alpha}_n-\alpha_o\right)$ and $U_{3,n}^*=n\left(\widehat{\beta}_n-\beta_o\right)'=\left[\mathbf{0}_{r_o},n\left(\widehat{O}_n-O_o\right)\right]:=\left[\mathbf{0}_{r_o},U_{2,n}^*\right]$, then

$$\begin{aligned}\left(\widehat{\Pi}_n-\Pi_o\right)Q^{-1}D_n^{-1} &= \left[\widehat{\alpha}_n\left(\widehat{\beta}_n-\beta_o\right)'+\left(\widehat{\alpha}_n-\alpha_o\right)\beta_o'\right]Q^{-1}D_n^{-1}\\ &= \left[n^{-\frac{1}{2}}\widehat{\alpha}_n U_{3,n}^*\alpha_o(\beta_o'\alpha_o)^{-1}+U_{1,n}^*,\ \widehat{\alpha}_n U_{3,n}^*\beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\right].\end{aligned}$$

Define

$$\Pi_n(U)=\left[n^{-\frac{1}{2}}\widehat{\alpha}_n U_3\alpha_o(\beta_o'\alpha_o)^{-1}+U_1,\widehat{\alpha}_n U_3\beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\right],$$

then by definition, $U_n^*=\left(U_{1,n}^*,U_{2,n}^*\right)$ minimizes the following criterion function

$$\begin{aligned}V_{4,n}(U) &= \sum_{t=1}^{n}\left(\|\Delta Y_t-\Pi_o Y_{t-1}-\Pi_n(U)D_nZ_{t-1}\|_E^2-\|\Delta Y_t-\Pi_o Y_{t-1}\|_E^2\right)\\ &\quad +n\sum_{k\in\mathcal{S}_\phi}\left\{\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_n(U)D_nQ+\Pi_o)\right]-\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right\}\end{aligned}$$

w.p.a.1.

For any compact set $K\in R^{m\times r_o}\times R^{r_o\times(m-r_o)}$ and any $U\in K$, we have

$$\Pi_n(U)D_nQ=O_p(n^{-\frac{1}{2}}).$$

Hence, from Assumption PF.(ii)-(iii) and the Bauer-Fiker Theorem, we can deduce that

$$\begin{aligned} & n\left|\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_n(U)D_nQ+\Pi_o)\right]-\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right|\\ \leq\ & Cn\max_{k\in\mathcal{S}_\phi}\left\{\widehat{P}_{\lambda_n}'\left[\bar{\phi}_k(\Pi_o)\right]\right\}\|\Pi_n(U)D_nQ\|_E\\ =\ & O_p(n^{\frac{1}{2}})\max_{k\in\mathcal{S}_\phi}\left\{\widehat{P}_{\lambda_n}'\left[\bar{\phi}_k(\Pi_o)\right]\right\}=o_p(1),\end{aligned} \tag{3.36}$$

uniformly over $U\in K$.

As $\beta_o = [I_{r_o}, O_{r_o}]'$, by definition we have $\beta_{o,\perp} = [-O'_{r_o}, I_{m-r_o}]'$ and hence

$$\begin{aligned}\Pi_n(U) &\to_p \left[U_1, (0_{m\times r_o}, \alpha_o U_2)\,\beta_{o,\perp}(\alpha'_{o,\perp}\beta_{o,\perp})^{-1}\right] \\ &= \left[U_1, \alpha_o U_2(\alpha'_{o,\perp}\beta_{o,\perp})^{-1}\right] =: \Pi_\infty(U) \qquad (3.37)\end{aligned}$$

uniformly over $U \in K$. By Lemma 10.1 and (3.37), we deduce that

$$\begin{aligned}&\sum_{t=1}^{n}\left(\|\Delta Y_t - \Pi_o Y_{t-1} - \Pi_n(U) D_n Z_{t-1}\|_E^2 - \|\Delta Y_t - \Pi_o Y_{t-1}\|_E^2\right) \\ =\;& vec\,[\Pi_n(U)]'\left(D_n \sum_{t=1}^{n} Z_{t-1}Z'_{t-1} D_n \otimes I_m\right) vec\,[\Pi_n(U)] \\ &-2vec\,[\Pi_n(U)]'\,vec\left(\sum_{t=1}^{n} u_t Z'_{t-1} D_n\right) \\ \to_d\;& vec\,[\Pi_\infty(U)]'\left[\begin{pmatrix} \Sigma_{z_1z_1} & 0 \\ 0 & \int B_{w_2}B'_{w_2}\alpha_{o,\perp}\end{pmatrix}\otimes I_m\right] vec\,[\Pi_\infty(U)] \\ &-2vec\,[\Pi_\infty(U)]'\,vec\,[(B_{1,m}, B_{2,m})] := V_2(U) \qquad (3.38)\end{aligned}$$

uniformly over $U \in K$.

Note that $vec\,[\Pi_\infty(U)] = \left[vec(U_1)', vec(\alpha_o U_2(\alpha'_{o,\perp}\beta_{o,\perp})^{-1})'\right]'$ and $vec(\alpha_o U_2(\alpha'_{o,\perp}\beta_{o,\perp})^{-1}) = \left[(\alpha'_{o,\perp}\beta_{o,\perp})^{-1}\otimes\alpha'_o\right]' vec(U_2)$. Hence $V_2(U)$ can be rewritten as

$$\begin{aligned}V_2(U) &= vec(U_1)'\left[\Sigma_{z_1z_1}\otimes I_m\right] vec(U_1) \\ &+vec(U_2)'\left[(\alpha'_{o,\perp}\beta_{o,\perp})^{-1}\int B_{w_2}B'_{w_2}(\beta'_{o,\perp}\alpha_{o,\perp})^{-1}\otimes\alpha'_o\alpha_o\right] vec(U_2) \\ &-2vec(U_1)'vec\,(B_{1,m}) - 2vec(U_2)'vec\left[\alpha'_o B_{2,m}(\beta'_{o,\perp}\alpha_{o,\perp})^{-1}\right]. \qquad (3.39)\end{aligned}$$

The expression in (3.39) makes it clear that $V_2(U)$ is uniquely minimized at (U_1^*, U_2^*), where

$$U_1^* = B_{1,m}\Sigma_{z_1z_1}^{-1} \text{ and } U_2^* = (\alpha'_o\alpha_o)^{-1}\left(\alpha'_o B_{2,m}\right)\left(\int B_{w_2}B'_{w_2}\right)^{-1}(\alpha'_{o,\perp}\beta_{o,\perp}). \qquad (3.40)$$

From (3.32) and (3.33), we can see that U_n^* is asymptotically tight. Invoking the Argmax Continuous Mapping Theorem (ACMT), we can deduce that

$$U_n^* = (U_{1,n}^*, U_{2,n}^*) \to_d (U_1^*, U_2^*).$$

Applying the CMT, we get

$$Q\left(\widehat{\Pi}_n-\Pi_o\right)Q^{-1}D_n^{-1}\rightarrow_d\left(\begin{array}{cc}\beta_o'U_1^* & \beta_o'\alpha_oU_2^*(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\\ \alpha_{o,\perp}'U_1^* & 0\end{array}\right),$$

which finishes the proof. ■

Remark 3.7 *The generalized shrinkage LS estimator is defined as*

$$\widehat{\Pi}_{g,n}=\arg\min_{\Pi}\sum_{t=1}^{n}\left(\Delta Y_t-\Pi Y_{t-1}\right)'\widehat{\Omega}_{u,n}^{-1}\left(\Delta Y_t-\Pi Y_{t-1}\right)+n\sum_{k=1}^{m}\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi)\right] \tag{3.41}$$

where $\widehat{\Omega}_{u,n}$ is some consistent estimator of Ω_u. The consistency and convergence rate of $\widehat{\Pi}_{g,n}$ and the sparsity of $\phi_k(\widehat{\Pi}_{g,n})$ $(k=1,...,m)$ can be established similarly. Following similar arguments to those of Theorem 3.6, we can deduce that

$$\begin{aligned}&\sum_{t=1}^{n}\left(u_t-\Pi_n(U)D_nZ_{t-1}\right)'\widehat{\Omega}_{u,n}^{-1}\left(u_t-\Pi_n(U)D_nZ_{t-1}\right)-\sum_{t=1}^{n}u_t'\widehat{\Omega}_{u,n}^{-1}u_t\\ \rightarrow_d\ & vec(U_1)'\left(\Sigma_{z_1z_1}\otimes\Omega_u^{-1}\right)vec(U_1)\\ &+vec(U_2)'\left[(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\int B_{w_2}B_{w_2}'(\beta_{o,\perp}'\alpha_{o,\perp})^{-1}\otimes\alpha_o'\Omega_u^{-1}\alpha_o\right]vec(U_2)\\ &-2vec(U_1)'vec\left(\Omega_u^{-1}B_{1,m}\right)-2vec(U_2)'vec\left[\alpha_o'\Omega_u^{-1}B_{2,m}(\beta_{o,\perp}'\alpha_{o,\perp})^{-1}\right]\\ :\ &=V_3(U),\end{aligned} \tag{3.42}$$

uniformly over $U\in K$. $V_3(U)$ is uniquely minimized at $\left(U_{g,1}^*,U_{g,2}^*\right)$, where $U_{g,1}^*=B_{1,m}\Sigma_{z_1z_1}^{-1}$ and

$$U_2^*=\left(\alpha_o'\Omega_u^{-1}\alpha_o\right)^{-1}\left(\alpha_o'\Omega_u^{-1}B_{2,m}\right)\left(\int B_{w_2}B_{w_2}'\right)^{-1}(\alpha_{o,\perp}'\beta_{o,\perp}).$$

Invoking the ACMT, we deduce that

$$\sqrt{n}\left(\widehat{\alpha}_{g,n}-\alpha_o\right)\rightarrow_d B_{1,m}\Sigma_{z_1z_1}^{-1}, \tag{3.43}$$

and

$$n\left(\widehat{O}_{g,n}-O_o\right)\rightarrow_d\left(\alpha_o'\Omega_u^{-1}\alpha_o\right)^{-1}\left(\alpha_o'\Omega_u^{-1}B_{2,m}\right)\left(\int B_{w_2}B_{w_2}'\right)^{-1}(\alpha_{o,\perp}'\beta_{o,\perp}). \tag{3.44}$$

In the triangular representation of a cointegration system studied in Phillips (1991), we have $\alpha_o = [I_{r_o}, 0_{r_o \times (m-r_o)}]'$ *and* $\beta_o = [I_{r_o}, O_o]'$. *Thus,* $\alpha'_{o,\perp}\beta_{o,\perp} = I_{m-r_o}$ *and*

$$\alpha'_o \Omega_u^{-1} \alpha_o = \Omega_{u,11\cdot 2}^{-1} \text{ and } \alpha'_o \Omega_u^{-1} = \Omega_{u,11\cdot 2}^{-1}[I_{r_o}, -\Omega_{u,12}\Omega_{u,22}^{-1}], \tag{3.45}$$

where $\Omega_{u,11\cdot 2} = \Omega_{u,11}^{-1} - \Omega_{u,12}\Omega_{u,22}^{-1}\Omega_{u,21}$ *and* $\Omega_{u,11}$, $\Omega_{u,12}$ *and* $\Omega_{u,22}$ *denote the leading* $r_o \times r_o$ *sub-matrix of* Ω_u, *the upper-right* $r_o \times (m-r_o)$ *sub-matrix of* Ω_u *and last* $(m-r_o)\times(m-r_o)$ *sub-matrix of* Ω_u *respectively. Using (3.44) and (3.45), we have*

$$\sqrt{n}\left(\widehat{O}_{g,n} - O_o\right) \to_d \int dB_{u_{1\cdot 2}} B'_{u_2} \left(\int B_{u_2} B'_{u_2}\right)^{-1}, \tag{3.46}$$

where B_{u_1} *and* B_{u_2} *denotes the first* r_o *and last* $m - r_o$ *vectors of* B_u *and* $B_{u_{1\cdot 2}} = B_{u_1} - \Omega_{u,12}\Omega_{u,22}^{-1} B_{u_2}$. *From the results in (3.46), we can see that the generalized shrinkage LS estimator* $\widehat{O}_{g,n}$ *of the cointegration matrix* O_o *is asymptotically equivalent to the maximum likelihood estimator studied in Phillips (1991).*

4 Extension I: VECM Estimation with Weakly Dependent Innovations

In this section, we study shrinkage reduced rank estimation in a scenario where the equation innovations $\{u_t\}_{t\geq 1}$ are weakly dependent. Specifically, we assume that $\{u_t\}_{t\geq 1}$ is generated by a linear process satisfying the following condition.

Assumption 4.1 (LP) *Let* $D(L) = \sum_{j=0}^{\infty} D_j L^j$, *where* $D_0 = I_m$ *and* $D(1)$ *has full rank. Let* u_t *have the Wold representation*

$$u_t = D(L)\varepsilon_t = \sum_{j=0}^{\infty} D_j \varepsilon_{t-j}, \text{ with } \sum_{j=0}^{\infty} j^{\frac{1}{2}} ||D_j||_E < \infty, \tag{4.1}$$

where ε_t *is i.i.d.*$(0, \Sigma_\varepsilon)$ *with* Σ_ε *positive definite and finite fourth moments.*

Denote the long-run variance of $\{u_t\}_{t\geq 1}$ as $\Omega_u = \sum_{h=-\infty}^{\infty} \Gamma_{uu}(h)$. From the Wold representation in (4.1), we have $\Omega_u = D(1)\Sigma_\varepsilon D(1)'$, which is positive definite because $D(1)$ has full rank and Σ_ε is positive definite. The fourth moment assumption is needed for the limit distribution of sample autocovariances in the case of misspecified transient dynamics.

The following lemma is useful in establishing the asymptotic properties of the shrinkage estimator with weakly dependent innovations.

Lemma 4.1 *Under Assumption 3.2 and 4.1, (a)-(c) and (e) of Lemma 10.1 are unchanged, while Lemma 10.1.(d) becomes*

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \left[u_t Z_{1,t-1}' - \Gamma_{uz_1}(1)\right] \to_d N(0, \Sigma_{uz_1}), \tag{4.2}$$

where $\Gamma_{uz_1}(1) = \sum_{j=0}^{\infty} \Gamma_{uu}(j)\beta_o \left(R^j\right)' < \infty$ *and* Σ_{uz_1} *is the long run variance matrix of* $u_t \otimes Z_{1,t-1}$.

Proof. From the partial sum expression in (3.7), we get $Z_{1,t-1} = \beta_o' Y_{t-1} = R(L)\beta_o' u_t$, which implies that $\{\beta_o' Y_{t-1}\}_{t\geq 1}$ is a stationary process. Note that

$$E\left[u_t Z_{1,t-1}'\right] = \sum_{j=0}^{\infty} E\left[u_t u_{t-j}'\right] \beta_o \left(R^j\right)' = \sum_{j=0}^{\infty} \Gamma_{uu}(j)\beta_o \left(R^j\right)' < \infty.$$

Using a CLT for linear process time series (e.g. the multivariate version of theorem 8 and Remark 3.9 of Phillips and Solo, 1992), we deduce that

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \left[u_t Z_{1,t-1}' - \Gamma_{uz_1}(1)\right] \to_d N(0, \Sigma_{uz_1}),$$

which establishes (4.2). The results of (a)-(c) and (e) can be proved using similar arguments to those of Lemma 10.1. ■

As expected, under general weak dependence assumptions on u_t, the simple reduced rank regression models (2.1) and (3.1) are susceptible to the effects of potential misspecification in the transient dynamics. These effects bear on the stationary components in the system. In particular, due to the centering term $\Gamma_{uz_1}(1)$ in (4.2), both the OLS estimator $\widehat{\Pi}_{ols}$ and the shrinkage estimator $\widehat{\Pi}_n$ are asymptotically biased. Specifically, we show that $\widehat{\Pi}_{ols}$ has the following probability limit,

$$\widehat{\Pi}_{ols} \to_p \Pi_1 := Q^{-1} H_o Q + \Pi_o, \tag{4.3}$$

where $H_o = \left[\Gamma_{wz_1}(1)\Sigma_{z_1z_1}^{-1}, 0_{m\times(m-r_o)}\right]$. Note that

$$Q^{-1}H_oQ + \Pi_o = \left[\alpha_o + \Gamma_{uz_1}(1)\Sigma_{z_1z_1}^{-1}\right]\beta_o' = \widetilde{\alpha}_o\beta_o', \tag{4.4}$$

which implies that the asymptotic bias of the OLS estimator $\widehat{\Pi}_{ols}$ is introduced via the bias in the pseudo true value limit $\widetilde{\alpha}_o$. Observe also that $\Pi_1 = \widetilde{\alpha}_o\beta_o'$ has rank at most equal to r_o, the number of rows in β_o'. Denote

$$\begin{aligned}
\widehat{S}_{12} &= \sum_{t=1}^{n}\frac{Z_{1,t-1}Z_{2,t-1}'}{n},\ S_{21} = \sum_{t=1}^{n}\frac{Z_{2,t-1}Z_{1,t-1}'}{n}, \\
\widehat{S}_{11} &= \sum_{t=1}^{n}\frac{Z_{1,t-1}Z_{1,t-1}'}{n} \text{ and } \widehat{S}_{22} = \sum_{t=1}^{n}\frac{Z_{2,t-1}Z_{2,t-1}'}{n}.
\end{aligned}$$

The next Lemma presents some asymptotic properties of the bias in the OLS estimator $\widehat{\Pi}_{ols}$.

Lemma 4.2 *Let* $H_n = n\left[\Gamma_{wz_1}(1), 0_{m\times(m-r_o)}\right]\left(\sum_{t=1}^{n} Z_{t-1}Z_{t-1}'\right)^{-1}$ *and* $\Pi_{1,n} = Q^{-1}H_nQ + \Pi_o$. *Then*

(a) H_n *converges in probability to* H_o, *i.e.* $H_n \to_p H_o$;

(b) $nQ^{-1}H_nQ\beta_{o\perp}$ *has limit distribution* $\widetilde{\Pi}_1\beta_{o\perp}$, *where*

$$\widetilde{\Pi}_1 = \Gamma_{uz_1}(1)\Sigma_{z_1z_1}^{-1}(\beta_o'\alpha_o)^{-1}\left[\left(\int B_{w_2}dB_{w_1}'\right)' + \Sigma_{w_1w_2}\right]\left(\int B_{w_2}B_{w_2}'\right)^{-1}\alpha_{o\perp}'; \tag{4.5}$$

(c) $\sqrt{n}Q^{-1}\left(H_n - H_o\right)Q\beta_o$ *has the limit distribution* $\widetilde{\Pi}_2\beta_o$, *where*

$$\widetilde{\Pi}_2 = \Gamma_{uz_1}(1)\Sigma_{z_1z_1}^{-1}N(0, V_{z_1z_1})\Sigma_{z_1z_1}^{-1}\beta_o' \tag{4.6}$$

and $N(0, V_{z_1z_1})$ *denotes the matrix limit distribution of* $\sqrt{n}\left(\widehat{S}_{11} - \Sigma_{z_1z_1}\right)$.

Proof. (a). Denote $\Gamma_{wz_1}(1) = Q\Gamma_{uz_1}(1)$. By Lemma 4.1, we have

$$\begin{aligned} H_n &= \left[\Gamma_{wz_1}(1), 0_{m\times(m-r_o)}\right] n^{\frac{1}{2}} D_n^{-1} \left(D_n^{-1}\sum_{t=1}^{n} Z_{t-1}Z_{t-1}' D_n^{-1}\right)^{-1} D_n^{-1} n^{\frac{1}{2}} \\ &= \left[\Gamma_{wz_1}(1), 0_{m\times(m-r_o)}\right] \left(D_n^{-1}\sum_{t=1}^{n} Z_{t-1}Z_{t-1}' D_n^{-1}\right)^{-1} \left(\begin{array}{cc} I_{r_o} & 0 \\ 0 & n^{-\frac{1}{2}} I_{m-r_o} \end{array}\right) \\ &\to_p \left[\Gamma_{wz_1}\Sigma_{z_1z_1}^{-1} : 0\right] = H_o. \end{aligned} \tag{4.7}$$

(b). From the expression of H_n in the first line of (4.7), we get

$$\begin{aligned} nQ^{-1}H_nQ\beta_{o\perp} &= \left[\Gamma_{uz_1}(1), 0\right] \left(D_n^{-1}\sum_{t=1}^{n} Z_{t-1}Z_{t-1}' D_n^{-1}\right)^{-1} \left(\begin{array}{c} 0 \\ n^{\frac{1}{2}}\alpha_{o\perp}'\beta_{o\perp} \end{array}\right) \\ &= -\Gamma_{uz_1}(1)\widehat{S}_{11}^{-1}\widehat{S}_{12}\left(n^{-1}\widehat{S}_{22} - n^{-1}\widehat{S}_{21}\widehat{S}_{11}^{-1}\widehat{S}_{12}\right)^{-1} \alpha_{o\perp}'\beta_{o\perp}. \end{aligned} \tag{4.8}$$

By Lemma 4.1 and the CMT, we have

$$\left(\widehat{S}_{22} - n^{-1}\widehat{S}_{21}\widehat{S}_{11}^{-1}\widehat{S}_{12}\right)^{-1} \to_d \left(\int B_{w_2}B_{w_2}'\right)^{-1}. \tag{4.9}$$

Next note that

$$\widehat{S}_{12} \to_d -(\beta_o'\alpha_o)^{-1}\left[\left(\int B_{w_2}dB_{w_1}'\right)' + \Sigma_{w_1w_2}\right]. \tag{4.10}$$

The claimed result now follows by applying the results in (4.8)-(4.10), Lemma 4.1 and CMT into the expression in (4.8).

(c). From the expression of H_n in the first line of (4.7), we get

$$\begin{aligned} &\sqrt{n}Q^{-1}\left(H_n - H_o\right)Q\beta_o \\ =\ & \sqrt{n}\left[\Gamma_{uz_1}(1), 0\right]\left\{ nD_n^{-1}\left(\begin{array}{cc} \widehat{S}_{11} & n^{-\frac{1}{2}}\widehat{S}_{12} \\ n^{-\frac{1}{2}}\widehat{S}_{21} & n^{-1}\widehat{S}_{22} \end{array}\right)^{-1} D_n^{-1} - \left(\begin{array}{cc} \Sigma_{z_1z_1}^{-1} & 0 \\ 0 & 0 \end{array}\right)\right\} Q\beta_o \\ =\ & \sqrt{n}\Gamma_{uz_1}(1)\left(H_{n,11} - \Sigma_{z_1z_1}^{-1}\right)\beta_o'\beta_o, \end{aligned} \tag{4.11}$$

where $H_{n,11} = \left(\widehat{S}_{11} - \widehat{S}_{12}\widehat{S}_{22}^{-1}\widehat{S}_{21}\right)^{-1}$ and we use $\alpha'_{o\perp}\beta_o = 0$ in get the third equality. Invoking the CLT and Lemma 4.1, we get

$$\begin{aligned}
& \Gamma_{uz_1}(1)\sqrt{n}\left(H_{n,11} - \Sigma_{z_1z_1}^{-1}\right)\beta'_o\beta_o \\
= \quad & -\Gamma_{uz_1}(1)H_{n,11}\sqrt{n}\left(H_{n,11}^{-1} - \Sigma_{z_1z_1}\right)\Sigma_{z_1z_1}^{-1}\beta'_o\beta_o \\
= \quad & -\Gamma_{uz_1}(1)H_{11,n}\sqrt{n}\left(\widehat{S}_{11} - \Sigma_{z_1z_1}\right)\Sigma_{z_1z_1}^{-1}\beta'_o\beta_o + o_p(1) \\
\to_d \quad & \Gamma_{uz_1}(1)\Sigma_{z_1z_1}^{-1}N(0, V_{z_1z_1})\Sigma_{z_1z_1}^{-1}\beta'_o\beta_o, \qquad (4.12)
\end{aligned}$$

which finishes the proof. ■

Denote the rank of Π_1 by r_1. Then, by virtue of the expression $\Pi_1 = \widetilde{\alpha}_o\beta'_o$, we have $r_1 \leq r_o$ as indicated. Without loss of generality, we decompose Π_1 as $\Pi_1 = \widetilde{\alpha}_1\widetilde{\beta}_1'$ where $\widetilde{\alpha}_1$ and $\widetilde{\beta}_1$ are $m \times r_1$ matrixes with full rank. Denote the orthogonal complements of $\widetilde{\alpha}_1$ and $\widetilde{\beta}_1$ as $\widetilde{\alpha}_{1\perp}$ and $\widetilde{\beta}_{1\perp}$ respectively. Similarly, we decompose $\widetilde{\beta}_{1\perp}$ as $\widetilde{\beta}_{1\perp} = \left[\widetilde{\beta}_\perp, \beta_{o\perp}\right]$ where $\widetilde{\beta}_\perp$ is an $m \times (r_1 - r_o)$ matrix. Denote $\widetilde{\Pi}_0 = \left(\int B_{w_2}dB'_u\right)'\left(\int B_{w_2}B'_{w_2}\right)^{-1}\alpha'_{o\perp}$ and $\widetilde{\Pi}_3 = N(0, \Sigma_{uz_1})\Sigma_{z_1z_1}^{-1}\beta'_o$. Let $[\phi_1(\widehat{\Pi}_{ols}), ..., \phi_m(\widehat{\Pi}_{ols})]$ and $[\phi_1(\Pi_1), ..., \phi_m(\Pi_1)]$ be the ordered eigenvalues of $\widehat{\Pi}_{ols}$ and Π_1, respectively.

Lemma 4.3 *Under Assumption 3.2 and 4.1, we have the following results:*

(a) the OLS estimator $\widehat{\Pi}_{ols}$ satisfies

$$\left[Q\left(\widehat{\Pi}_{ols} - \Pi_o\right)Q^{-1} - H_n\right]D_n = O_p(1); \qquad (4.13)$$

(b) the eigenvalues of $\widehat{\Pi}_{ols}$ satisfy

$$[\phi_1(\widehat{\Pi}_{ols}), ..., \phi_{m-r_o}(\widehat{\Pi}_{ols})] \to_p (0, ..., 0)_{1\times(m-r_1)} \qquad (4.14)$$

and

$$[\phi_{m-r_1+1}(\widehat{\Pi}_{ols}), ..., \phi_m(\widehat{\Pi}_{ols})] \to_p \left[\phi_{m-r_o+1}(\Pi_1), ..., \phi_m(\Pi_1)\right]; \qquad (4.15)$$

(c) the first $m - r_o$ ordered eigenvalues of $\widehat{\Pi}_{ols}$ satisfy

$$n[\phi_1(\widehat{\Pi}_{ols}), ..., \phi_{m-r_o}(\widehat{\Pi}_{ols})] \to_d [\widetilde{\phi}_1', ..., \widetilde{\phi}_{m-r_o}'], \qquad (4.16)$$

where $\widetilde{\phi}_j'$ $(j = 1, ..., m-r_o)$ are the ordered solutions of $\left|uI_{m-r_o} - \beta'_{o\perp}\left(\widetilde{\Pi}_0 + \widetilde{\Pi}_1\right)\beta_{o\perp}\right| = 0$;

(d) $\widehat{\Pi}_{ols}$ *has* $r_o - r_1$ *eigenvalues satisfying*

$$\sqrt{n}[\phi_{m-r_o+1}(\widehat{\Pi}_{ols}), ..., \phi_{m-r_1}(\widehat{\Pi}_{ols})] \rightarrow_d [\widetilde{\phi}'_{m-r_o+1}, ..., \widetilde{\phi}'_{m-r_1}], \tag{4.17}$$

where $\widetilde{\phi}'_j$ *(*$j = m-r_o+1, ..., m-r_1$*) are the ordered solutions of* $\left|uI_{r_o-r_1} - \widetilde{\beta}'_{\perp}\left(\widetilde{\Pi}_2 + \widetilde{\Pi}_3\right)\widetilde{\beta}_{\perp}\right| = 0$*.*

Proof. (a). From the expression in (3.10) and (4.7),

$$\left[Q\left(\widehat{\Pi}_{ols} - \Pi_o\right)Q^{-1} - H_n\right]D_n = \left[\sum_{t=1}^{n}\left(w_t Z'_{t-1} - H_{1,o}\right)\right]D_n^{-1}\left(D_n^{-1}\sum_{t=1}^{n} Z_{t-1}Z'_{t-1}D_n^{-1}\right)^{-1}, \tag{4.18}$$

where $H_{1,o} = \left[\Gamma_{wz_1}(1), 0_{m\times(m-r_o)}\right]$. Now, the results in (a) are directly from Lemma 4.1 and CMT.

(b). Denote $P = \left[\widetilde{\beta}_1, \widetilde{\beta}_{1\perp}\right]$ and $S_n(\phi) = \phi I_m - \widehat{\Pi}_{ols}$, then by definition, the eigenvalues of $\widehat{\Pi}_{ols}$ are the solutions of the following determinantal equation,

$$0 = \left|P'S_n(\phi)P\right| = \left|\begin{array}{cc} \phi\widetilde{\beta}'_1\widetilde{\beta}_1 - \widetilde{\beta}'_1\widehat{\Pi}_{ols}\widetilde{\beta}_1 & -\widetilde{\beta}'_1\widehat{\Pi}_{ols}\widetilde{\beta}_{1\perp} \\ -\widetilde{\beta}'_{1\perp}\widehat{\Pi}_{ols}\widetilde{\beta}_1 & \phi I_{m-r_o} - \widetilde{\beta}'_{1\perp}\widehat{\Pi}_{ols}\widetilde{\beta}_{1\perp} \end{array}\right|. \tag{4.19}$$

As $\Pi_1\widetilde{\beta}_{1\perp} = 0$ and $\widehat{\Pi}_{ols} = \Pi_1 + o_p(1)$,

$$\widetilde{\beta}'_1\widehat{\Pi}_{ols}\widetilde{\beta}_{1\perp} = \widetilde{\beta}'_1\left(\widehat{\Pi}_{ols} - \Pi_1\right)\widetilde{\beta}_{1\perp} + \widetilde{\beta}'_1\Pi_1\widetilde{\beta}_{1\perp} = o_p(1). \tag{4.20}$$

Similarly, we have

$$\widetilde{\beta}'_{1\perp}\widehat{\Pi}_{ols}\widetilde{\beta}_{1\perp} = \widetilde{\beta}'_{1\perp}\left(\widehat{\Pi}_{ols} - \Pi_1\right)\widetilde{\beta}_{1\perp} = o_p(1) \tag{4.21}$$

and

$$\widetilde{\beta}'_{1\perp}\widehat{\Pi}_{ols}\widetilde{\beta}_1 \rightarrow_p \widetilde{\beta}'_{1\perp}\Pi_1\widetilde{\beta}_1 \text{ and } \widetilde{\beta}'_1\widehat{\Pi}_{ols}\widetilde{\beta}_1 \rightarrow_p \widetilde{\beta}'_1\Pi_1\widetilde{\beta}_1. \tag{4.22}$$

From the results in (4.19)-(4.22), we can invoke the Slutsky Theorem to deduce that

$$\left|\phi I_m - \widehat{\Pi}_{ols}\right| \rightarrow_p |\phi I_{m-r_1}| \times \left|\phi\widetilde{\beta}'_1\widetilde{\beta}_1 - \widetilde{\beta}'_1\Pi_1\widetilde{\beta}_1\right|, \tag{4.23}$$

uniformly over any compact set in R, where $\left|\phi\widetilde{\beta}'_1\widetilde{\beta}_1 -_p \widetilde{\beta}'_1\Pi_1\widetilde{\beta}_1\right| = 0$ can equivalently be written as $\left|\phi I_{r_1-r_o} - \widetilde{\beta}'_1\widetilde{\alpha}_1\right| = 0$. Hence the claimed results follow by (4.23) and the CMT.

(c). If we denote $u^*_{n,k} = n\phi_k(\widehat{\Pi}_{ols})$, then by definition, $u^*_{n,k}$ ($k \in \mathcal{S}^c_{ols,\phi}$) is the solution of the following determinantal equation

$$0 = \left|\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right| \times \left|\widetilde{\beta}_{1\perp}' \left\{S_n(u) - S_n(u)\widetilde{\beta}_1 \left[\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right]^{-1} \widetilde{\beta}_1' S_n(u)\right\} \widetilde{\beta}_{1\perp}\right|, \tag{4.24}$$

where $S_n(u) = \frac{u}{n} I_m - \widehat{\Pi}_{ols}$.

From the results in (4.3) and (4.4), we have

$$\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1 = n^{-\frac{1}{2}} u \widetilde{\beta}_1' \widetilde{\beta}_1 - \widetilde{\beta}_1' \widehat{\Pi}_{ols} \widetilde{\beta}_1 \to_p -\widetilde{\beta}_1' \widetilde{\alpha}_1 \widetilde{\beta}_1' \widetilde{\beta}_1, \tag{4.25}$$

where $\widetilde{\beta}_1' \widetilde{\alpha}_1 \widetilde{\beta}_1' \widetilde{\beta}_1$ is a $r_1 \times r_1$ nonsingular matrix. Hence $u^*_{n,k}$ is the solution of the following determinantal equation asymptotically

$$0 = \left|\widetilde{\beta}_{1\perp}' \left\{S_n(u) - S_n(u)\widetilde{\beta}_1 \left[\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right]^{-1} \widetilde{\beta}_1' S_n(u)\right\} \widetilde{\beta}_{1\perp}\right|. \tag{4.26}$$

Denote $T_n(u) = S_n(u) - S_n(u)\widetilde{\beta}_1 \left[\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right]^{-1} \widetilde{\beta}_1' S_n(u)$, then (4.26) can be equivalently written as

$$0 = \left|\widetilde{\beta}_\perp' T_n(u)\widetilde{\beta}_\perp\right| \times \left|\beta_{o\perp}' \left\{T_n(u) - T_n(u)\widetilde{\beta}_\perp \left[\widetilde{\beta}_\perp' T_n(u)\widetilde{\beta}_\perp\right]^{-1} \widetilde{\beta}_\perp' T_n(u)\right\} \beta_{o\perp}\right|. \tag{4.27}$$

Note that

$$\begin{aligned}
n^{\frac{1}{2}}\widetilde{\beta}_\perp' S_n(u)\widetilde{\beta}_\perp &= n^{-\frac{1}{2}} u \widetilde{\beta}_\perp' \widetilde{\beta}_\perp - n^{\frac{1}{2}}\widetilde{\beta}_\perp' \left(\widehat{\Pi}_{ols} - \Pi_{1,n}\right)\widetilde{\beta}_\perp - n^{\frac{1}{2}}\widetilde{\beta}_\perp' \Pi_{1,n}\widetilde{\beta}_\perp, \\
n^{\frac{1}{2}}\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_\perp &= -n^{\frac{1}{2}}\widetilde{\beta}_1' \left(\widehat{\Pi}_{ols} - \Pi_{1,n}\right)\widetilde{\beta}_\perp - n^{\frac{1}{2}}\widetilde{\beta}_1' \Pi_{1,n}\widetilde{\beta}_\perp, \\
\widetilde{\beta}_\perp' S_n(u)\widetilde{\beta}_1 &= -\widetilde{\beta}_\perp' \widehat{\Pi}_{ols}\widetilde{\beta}_1.
\end{aligned}$$

From above expressions, we can write

$$n^{\frac{1}{2}}\widetilde{\beta}_\perp' T_n(u)\widetilde{\beta}_\perp = n^{-\frac{1}{2}} u \widetilde{\beta}_\perp' \widetilde{\beta}_\perp - \widetilde{\beta}_\perp' \left[I_m + \widehat{\Pi}_{ols}\widetilde{\beta}_1 \left[\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right]^{-1} \widetilde{\beta}_1'\right] M_{1,n}, \tag{4.28}$$

where $M_{1,n} = n^{\frac{1}{2}} \left(\widehat{\Pi}_{ols} - \Pi_{1,n} \right) \widetilde{\beta}_\perp + n^{\frac{1}{2}} \Pi_{1,n} \widetilde{\beta}_\perp$. Using Lemma 4.1 and the results in (a), we can deduce that

$$\begin{aligned} n^{\frac{1}{2}} \left(\widehat{\Pi}_{ols} - \Pi_{1,n} \right) \widetilde{\beta}_\perp &= Q^{-1} \left[Q \left(\widehat{\Pi}_{ols} - \Pi_{1,n} \right) Q^{-1} D_n \right] n^{\frac{1}{2}} D_n^{-1} Q \widetilde{\beta}_\perp \\ &\to_d N(0, \Sigma_{uz_1}) \Sigma_{z_1 z_1}^{-1} \beta_o' \widetilde{\beta}_\perp := \mathcal{N}_1 \widetilde{\beta}_\perp. \end{aligned} \tag{4.29}$$

Using the results in Lemma 4.2, we get

$$\begin{aligned} n^{\frac{1}{2}} \Pi_{1,n} \widetilde{\beta}_\perp &= n^{\frac{1}{2}} \left(\Pi_{1,n} - \Pi_1 \right) \widetilde{\beta}_\perp = n^{\frac{1}{2}} Q^{-1} \left(H_n - H_o \right) Q \widetilde{\beta}_\perp \\ &\to_d \Gamma_{uz_1}(1) \Sigma_{z_1 z_1}^{-1} N(0, V_{z_1 z_1}) \Sigma_{z_1 z_1}^{-1} \beta_o' \widetilde{\beta}_\perp := \mathcal{N}_2 \widetilde{\beta}_\perp. \end{aligned} \tag{4.30}$$

From (4.28)-(4.30), we can deduce that

$$\left| \sqrt{n} \widetilde{\beta}_\perp' T_n(u) \widetilde{\beta}_\perp \right| \to_d \left| \widetilde{\beta}_\perp' \left(\mathcal{N}_1 + \mathcal{N}_2 \right) \widetilde{\beta}_\perp \right| \neq 0, \; a.e. \tag{4.31}$$

Next note that

$$\beta_{o\perp}' T_n(u) \beta_{o\perp} = n^{-1} u \beta_{o\perp}' \beta_{o\perp} - \beta_{o\perp}' \left[I_m + \widehat{\Pi}_{ols} \widetilde{\beta}_1 \left(\widetilde{\beta}_1' S_n(u) \widetilde{\beta}_1 \right)^{-1} \widetilde{\beta}_1' \right] M_{2,n}, \tag{4.32}$$

$$\widetilde{\beta}_\perp' T_n(u) \beta_{o\perp} = -\widetilde{\beta}_\perp' \left[I_m + \widehat{\Pi}_{ols} \widetilde{\beta}_1 \left(\widetilde{\beta}_1' S_n(u) \widetilde{\beta}_1 \right)^{-1} \widetilde{\beta}_1' \right] M_{2,n}, \tag{4.33}$$

where $M_{2,n} = \left(\widehat{\Pi}_{ols} - \Pi_{1,n} \right) \beta_{o\perp} + \Pi_{1,n} \beta_{o\perp}$. By (4.22) and (4.25), we have

$$\beta_{o\perp}' T_n(u) \widetilde{\beta}_\perp = \beta_{o\perp}' \widehat{\Pi}_{ols} \widetilde{\beta}_\perp - \beta_{o\perp}' \widehat{\Pi}_{ols} \widetilde{\beta}_1 \left(\widetilde{\beta}_1' S_n(u) \widetilde{\beta}_1 \right)^{-1} \widetilde{\beta}_1' \widehat{\Pi}_{ols} \widetilde{\beta}_\perp = o_p(1). \tag{4.34}$$

Using Lemma 4.2, we can deduce that

$$\begin{aligned} n \left(\widehat{\Pi}_{ols} - \Pi_{1,n} \right) \beta_{o\perp} &= n Q^{-1} \left[Q \left(\widehat{\Pi}_{ols} - \Pi_{1,n} \right) Q^{-1} D_n \right] D_n^{-1} Q \beta_{o\perp} \\ &= Q^{-1} \left[Q \left(\widehat{\Pi}_{ols} - \Pi_{1,n} \right) Q^{-1} D_n \right] \begin{pmatrix} 0 \\ \alpha_{o\perp}' \beta_{o\perp} \end{pmatrix} \\ &\to_d \left(\int B_{w_2} dB_u' \right)' \left(\int B_{w_2} B_{w_2}' \right)^{-1} \alpha_{o\perp}' \beta_{o\perp} := \widetilde{\Pi}_0 \beta_{o\perp} \end{aligned} \tag{4.35}$$

Using the result in (4.5), we get

$$
\begin{aligned}
n\Pi_{1n}\beta_{o\perp} &= n\left[\Pi_o + Q^{-1}H_nQ\right]\beta_{o\perp} \\
&= nQ^{-1}H_nQ\beta_{o\perp} \to_d \widetilde{\Pi}_1\beta_{o\perp}. \qquad (4.36)
\end{aligned}
$$

From (4.32)-(4.36), we deduce that

$$
\begin{aligned}
&\left|n\beta_{o\perp}'\left\{T_n(u) - T_n(u)\widetilde{\beta}_\perp\left[\widetilde{\beta}_\perp' T_n(u)\widetilde{\beta}_\perp\right]^{-1}\widetilde{\beta}_\perp' T_n(u)\right\}\beta_{o\perp}\right| \\
\to_d\ &\left|uI_{m-r_o} - \beta_{o\perp}'\left(\widetilde{\Pi}_0 + \widetilde{\Pi}_1\right)\beta_{o\perp}\right|, \qquad (4.37)
\end{aligned}
$$

uniformly over any compact set in R. Now, the results in (c) follow from (4.37) and the CMT.

(d) If we denote $u_{n,k}^* = \sqrt{n}\phi_k(\widehat{\Pi}_{ols})$, then by definition, $u_{n,k}^*$ ($k \in \widetilde{\mathcal{S}}_{ols,\phi}^c$) is the solution of the following determinantal equation

$$
0 = \left|\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right| \times \left|\widetilde{\beta}_{1\perp}'\left\{S_n(u) - S_n(u)\widetilde{\beta}_1\left[\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right]^{-1}\widetilde{\beta}_1' S_n(u)\right\}\widetilde{\beta}_{1\perp}\right|, \qquad (4.38)
$$

where $S_n(u) = \frac{u}{\sqrt{n}}I_m - \widehat{\Pi}_{ols}$.

Note that

$$
\begin{aligned}
\widetilde{\beta}_{1\perp}' S_n(u)\widetilde{\beta}_{1\perp} &= n^{-\frac{1}{2}}uI_{m-r_1} - \widetilde{\beta}_{1\perp}'\widehat{\Pi}_{ols}\widetilde{\beta}_{1\perp}, \qquad (4.39) \\
\widetilde{\beta}_{1\perp}' S_n(u)\widetilde{\beta}_1 &= -\widetilde{\beta}_{1\perp}'\widehat{\Pi}_{ols}\widetilde{\beta}_1 \text{ and } \widetilde{\beta}_1' S_n(u)\widetilde{\beta}_{1\perp} = -\widetilde{\beta}_1'\widehat{\Pi}_{ols}\widetilde{\beta}_{1\perp}. \qquad (4.40)
\end{aligned}
$$

Using expressions (4.39) and (4.40), we have

$$
\begin{aligned}
&\widetilde{\beta}_{1\perp}'\left\{S_n(u) - S_n(u)\widetilde{\beta}_1\left[\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right]^{-1}\widetilde{\beta}_1' S_n(u)\right\}\widetilde{\beta}_{1\perp} \\
=\ &n^{-\frac{1}{2}}uI_{m-r_1} - \widetilde{\beta}_{1\perp}'\left\{I_m + \widehat{\Pi}_{ols}\widetilde{\beta}_1\left[\widetilde{\beta}_1' S_n(u)\widetilde{\beta}_1\right]^{-1}\widetilde{\beta}_1'\right\}\widehat{\Pi}_{ols}\widetilde{\beta}_{1\perp}. \qquad (4.41)
\end{aligned}
$$

From (4.35), we get

$$
\sqrt{n}\left(\widehat{\Pi}_{ols} - \Pi_{1n}\right)\beta_{o,\perp} = o_p(1). \qquad (4.42)
$$

Using (4.29), (4.5) and (4.6), we have

$$\sqrt{n}\left(\widehat{\Pi}_{ols}-\Pi_{1n}\right)\widetilde{\beta}_{\perp} \to_d \widetilde{\Pi}_3\widetilde{\beta}_{\perp}, \tag{4.43}$$

$$\sqrt{n}\Pi_{1n}\beta_{o\perp} = o_p(1), \tag{4.44}$$

$$\sqrt{n}\Pi_{1n}\widetilde{\beta}_{\perp} \to_d \widetilde{\Pi}_2\widetilde{\beta}_{\perp}. \tag{4.45}$$

Now, using the results in (4.42)-(4.45), we get

$$\begin{aligned}\sqrt{n}\widehat{\Pi}_{ols}\widetilde{\beta}_{1\perp} &= \sqrt{n}\left[\widehat{\Pi}_{ols}-\Pi_{1n}\right]\widetilde{\beta}_{1\perp}+\sqrt{n}\Pi_{1n}\widetilde{\beta}_{1\perp} \\ &\to_d \left[0_{m\times(m-r_o)},\ \widetilde{\Pi}_3\widetilde{\beta}_{\perp}\right]+\left[0_{m\times(m-r_o)},\ \widetilde{\Pi}_2\widetilde{\beta}_{\perp}\right] \\ &= \left[0_{m\times(m-r_o)},\ \left(\widetilde{\Pi}_2+\widetilde{\Pi}_3\right)\widetilde{\beta}_{\perp}\right].\end{aligned} \tag{4.46}$$

From the results in (4.41)-(4.46), it follows that

$$\begin{aligned}&\left|\sqrt{n}\widetilde{\beta}_{1\perp}'\left\{S_n(u)-S_n(u)\widetilde{\beta}_1\left[\widetilde{\beta}_1'S_n(u)\widetilde{\beta}_1\right]^{-1}\widetilde{\beta}_1'S_n(u)\right\}\widetilde{\beta}_{1\perp}\right| \\ \to_d &\left|uI_{m-r_1}-\widetilde{\beta}_{1\perp}'\left[I_m+\widetilde{\alpha}_1\left(\widetilde{\beta}_1'\widetilde{\alpha}_1\right)^{-1}\widetilde{\beta}_1'\right]\left[0_{m\times(m-r_o)},\ \left(\mathcal{N}_1+\mathcal{N}_2\right)\widetilde{\beta}_{\perp}\right]\right| \\ = &\left|uI_{m-r_0}\right|\times\left|uI_{r_o-r_1}-\widetilde{\beta}_{\perp}'\left(\widetilde{\Pi}_2+\widetilde{\Pi}_3\right)\widetilde{\beta}_{\perp}\right|.\end{aligned} \tag{4.47}$$

Note that the determinantal equation

$$\left|uI_{m-r_0}\right|\times\left|uI_{r_o-r_1}-\widetilde{\beta}_{\perp}'\left(\widetilde{\Pi}_2+\widetilde{\Pi}_3\right)\widetilde{\beta}_{\perp}\right|=0 \tag{4.48}$$

has $m-r_0$ zero eigenvalues, which correspond to the probability limit of $\sqrt{n}\phi_{\mathcal{S}^c_{ols,\phi}}(\widehat{\Pi}_{ols})$, as illustrated in (c). Equation (4.48) also has r_o-r_1 non-trivial eigenvalues as solutions of the stochastic determinantal equation $\left|uI_{r_o-r_1}-\widetilde{\beta}_{\perp}'\left(\widetilde{\Pi}_2+\widetilde{\Pi}_3\right)\widetilde{\beta}_{\perp}\right|=0$, which finishes the proof. ■

We next derive the asymptotic properties of the shrinkage estimator $\widehat{\Pi}_n$ with weakly dependent innovations. Let $\widetilde{\mathcal{S}}_\phi=\{k:\phi_k(\Pi_1)\neq 0, k=1,...,m\}$ be the index set of non-zero eigenvalues of Π_1. From Lemma 4.3, it is clear that $\widetilde{\mathcal{S}}_\phi\subset\mathcal{S}_\phi$.

Corollary 4.1 *Under Assumption 3.1, 4.1 and 3.3.(i), the shrinkage estimator* $\widehat{\Pi}_n$ *satis-*

fies

$$\widehat{\Pi}_n \to_p \Pi_1, \tag{4.49}$$

where Π_1 *is defined in (4.3).*

Proof. First, when $r_o = 0$, then $\Pi_1 = \widetilde{\alpha}_o \beta_o' = 0 = \Pi_o$. Hence, (4.49) follows by the similar arguments to those in the proof of Theorem 3.1. To finish the proof, we only need to consider scenarios where $r_o = m$ and $r_o \in (0, m)$.

Using the same notation for $V_{1,n}(\cdot)$ defined in the proof of Theorem 3.1, by definition we have $V_{1,n}(\widehat{\Pi}_n) \leq V_{1,n}(\Pi_1)$, which implies

$$\begin{aligned} & \left[vec(\widehat{\Pi}_n - \Pi_1)\right]' \left(\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' \otimes I_m\right) \left[vec(\widehat{\Pi}_n - \Pi_1)\right] \\ & +2\left[vec(\widehat{\Pi}_n - \Pi_1)\right]' vec\left[\sum_{t=1}^{n} u_t Y_{t-1}' - (\Pi_1 - \Pi_o)\sum_{t=1}^{n} Y_{t-1}Y_{t-1}'\right] \\ \leq \; & n\left\{\sum_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_1)\right] - \sum_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\widehat{\Pi}_n)\right]\right\}. \end{aligned} \tag{4.50}$$

When $r_o = m$, Y_t is stationary and we have

$$\frac{1}{n}\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' \to_p \Gamma_{YY}(0) = R(1)\Omega_u R(1)'. \tag{4.51}$$

From the results in (4.50) and (4.51), we get w.p.a.1,

$$\left\|\widehat{\Pi}_n - \Pi_1\right\|_E^2 \mu_{\min} - \left\|\widehat{\Pi}_n - \Pi_1\right\|_E c_n - d_n \leq 0, \tag{4.52}$$

where $\mu_{\min}$ denotes the smallest eigenvalue of $\Gamma_{YY}(0)$, which is positive with probability 1,

$$\begin{aligned} c_n &= \left\|\frac{\sum_{t=1}^{n} u_t Y_{t-1}'}{n} - (\Pi_1 - \Pi_o)\frac{\sum_{t=1}^{n} Y_{t-1}Y_{t-1}'}{n}\right\|_E \\ &\to_p \left\|\Gamma_{uY}(1) - \Gamma_{uY}(1)\Sigma_{z_1z_1}^{-1}\Sigma_{z_1z_1}\right\|_E = 0, \end{aligned}$$

and $d_n = \sum_{k\in\widetilde{\mathcal{S}}_\phi} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_1)\right] = o_p(1)$. So result in (4.49) follows directly from the inequality in (4.52).

When $0 < r_o < m$, we get

$$\begin{aligned}
& \left[vec(\widehat{\Pi}_n - \Pi_1)\right]' \left(\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' \otimes I_m\right) \left[vec(\widehat{\Pi}_n - \Pi_1)\right] \\
= \;& vec(\widehat{\Pi}_n - \Pi_1)' \left(B_n D_n \sum_{t=1}^{n} Z_{t-1}Z_{t-1}' D_n B_n' \otimes I_m\right) vec(\widehat{\Pi}_n - \Pi_1) \\
\geq \;& \mu_{\min} \left\| \left(\widehat{\Pi}_n - \Pi_1\right) B_n \right\|_E^2 . \qquad (4.53)
\end{aligned}$$

Next, note that

$$\begin{aligned}
& \left\{\sum_{t=1}^{n} u_t Z_{t-1}' - \left[(\Pi_1 - \Pi_o)Q^{-1}\right] \sum_{t=1}^{n} Z_{t-1}Z_{t-1}'\right\} D_n \\
= \;& \begin{bmatrix} n^{-\frac{1}{2}} \sum_{t=1}^{n} Z_{1,t-1}u_t' \\ n^{-1} \sum_{t=1}^{n} Z_{2,t-1}u_t' \end{bmatrix}' - \begin{bmatrix} n^{-\frac{1}{2}} \sum_{t=1}^{n} Z_{1,t-1}Z_{1,t-1}' \Sigma_{z_1 z_1}^{-1} \Gamma_{uz_1}'(1) \\ n^{-1} \sum_{t=1}^{n} Z_{2,t-1}Z_{1,t-1}' \Sigma_{z_1 z_1}^{-1} \Gamma_{uz_1}'(1) \end{bmatrix}' . \qquad (4.54)
\end{aligned}$$

From Lemma 10.1, we can deduce that

$$n^{-1} \sum_{t=1}^{n} Z_{2,t-1}u_t' = O_p(1) \text{ and } n^{-1} \sum_{t=1}^{n} Z_{2,t-1}Z_{1,t-1}' \Sigma_{\beta\beta}^{-1} \Gamma_{uZ_1}'(1) = O_p(1). \qquad (4.55a)$$

Similarly, we get

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \left[Z_{1,t-1}u_t' - \Gamma_{uz_1}'(1)\right] - n^{\frac{1}{2}} \left[S_{n,11} - \Sigma_{z_1 z_1}\right] \Sigma_{z_1 z_1}^{-1} \Gamma_{uz_1}'(1) = O_p(1). \qquad (4.56)$$

Define $e_n = \left\| \left\{ \sum_{t=1}^{n} u_t Z_{t-1}' - \left[(\Pi_1 - \Pi_o)Q^{-1}\right] \sum_{t=1}^{n} Z_{t-1}Z_{t-1}' \right\} D_n \right\|_E$, then from (4.54)-(4.56) we can deduce that $e_n = O_p(1)$. By Cauchy-Schwarz, we have

$$\begin{aligned}
& \left| \left[vec(\widehat{\Pi}_n - \Pi_1)\right]' vec\left[\sum_{t=1}^{n} u_t Y_{t-1}' - (\Pi_1 - \Pi_o) \sum_{t=1}^{n} Y_{t-1}Y_{t-1}'\right] \right| \\
= \;& \left| \left[vec(\widehat{\Pi}_n - \Pi_1)\right]' vec\left[\left\{\sum_{t=1}^{n} u_t Z_{t-1}' - \left[(\Pi_1 - \Pi_o)Q^{-1}\right] \sum_{t=1}^{n} Z_{t-1}Z_{t-1}'\right\} D_n B_n'\right] \right| \\
\leq \;& \left\| \left(\widehat{\Pi}_n - \Pi_1\right) B_n \right\|_E e_n . \qquad (4.57)
\end{aligned}$$

Define $d_n = n \sum_{k \in \widetilde{\mathcal{S}}_\phi} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_1)\right]$, from results in (4.50), (4.53) and (4.57), we get

$$\mu_{\min} \left\| \left(\widehat{\Pi}_n - \Pi_1 \right) B_n \right\|_E^2 - 2 \left\| \left(\widehat{\Pi}_n - \Pi_1 \right) B_n \right\|_E e_n - d_n \leq 0 \tag{4.58}$$

Now, result in (4.49) follows by the same arguments in Theorem 3.1. ■

Corollary 4.1 implies that the shrinkage estimator $\widehat{\Pi}_n$ has the same probability limit as that of the OLS estimator $\widehat{\Pi}_{ols}$. We next derive the convergence rate of $\widehat{\Pi}_n$.

Corollary 4.2 *Denote $a_n = \max_{k \in \widetilde{\mathcal{S}}_\phi} \left| \widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_1)\right] \right|$. Under Assumption RR, 3.3.(i)-(ii) and 4.1, the shrinkage LS estimator $\widehat{\Pi}_n$ satisfies*
(a) if $r_o = 0$, then $\widehat{\Pi}_n - \Pi_1 = O_p(n^{-1} + n^{-1} a_n)$;
(b) if $0 < r_o \leq m$, then $\left(\widehat{\Pi}_n - \Pi_1 \right) Q^{-1} D_n^{-1} = O_p(n^{-\frac{1}{2}} + a_n)$.

Proof. By the results of Corollary 4.1 and the CMT, we deduce that $\phi_k(\widehat{\Pi}_n)$ is a consistent estimator of $\phi_k(\Pi_1)$ for all $k = 1, ..., m$. Using the same arguments as those of Theorem 3.2, we deduce that

$$\begin{aligned} \sum_{k \in \widetilde{\mathcal{S}}_\phi} \left(\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_1)\right] - \widehat{P}_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)] \right) \quad \leq \quad & C \max_{k \in \widetilde{\mathcal{S}}_\phi} \left| \widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_1)\right] \right| \left\| \widehat{\Pi}_n - \Pi_1 \right\|_E \\ & + o_p(1) \left\| \widehat{\Pi}_n - \Pi_1 \right\|_E^2 . \end{aligned} \tag{4.59}$$

Using the inequality (4.59) in (4.50) gives

$$\begin{aligned} & \left[vec(\widehat{\Pi}_n - \Pi_1) \right]' \left(\sum_{t=1}^n Y_{t-1} Y'_{t-1} \otimes I_m \right) \left[vec(\widehat{\Pi}_n - \Pi_1) \right] \\ & +2 \left[vec(\widehat{\Pi}_n - \Pi_1) \right]' vec \left[\sum_{t=1}^n u_t Y'_{t-1} - (\Pi_1 - \Pi_o) \sum_{t=1}^n Y_{t-1} Y'_{t-1} \right] \\ \leq \quad & Cn \max_{k \in \widetilde{\mathcal{S}}_\phi} \left| \widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_1)\right] \right| \left\| \widehat{\Pi}_n - \Pi_1 \right\|_E + o_p(1) \left\| \widehat{\Pi}_n - \Pi_1 \right\|_E^2 , \end{aligned} \tag{4.60}$$

where $C > 0$ denotes a generic constant.

When $r_o = 0$, the convergence rate of $\widehat{\Pi}_n$ could be derived using the same arguments in Theorem 3.2. Hence, to finish the proof, we only need to consider scenarios where $r_o = m$ and $0 < r_o < m$.

Consider the case $r_o = m$. Note that

$$
\begin{aligned}
& n^{-\frac{1}{2}} \sum_{t=1}^{n} u_t Y_{t-1}' - n^{-\frac{1}{2}} (\Pi_1 - \Pi_o) \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' \\
= \; & n^{-\frac{1}{2}} \sum_{t=1}^{n} \left[u_t Y_{t-1}' - \Gamma_{uY}(1) \right] - \Gamma_{uY}(1) \Sigma_{z_1 z_1}^{-1} \left[n^{\frac{1}{2}} \left(\widehat{S}_{11} - \Sigma_{z_1 z_1} \right) \right] = O_p(1).
\end{aligned}
$$

Following similar arguments to those of Theorem 3.2, we get

$$
\left\| \widehat{\Pi}_n - \Pi_1 \right\|_E^2 \mu_{\min} - \left\| \widehat{\Pi}_n - \Pi_1 \right\|_E (c_n + a_n C) \leq 0, \tag{4.61}
$$

where $c_n = \left\| n^{-1} \left[\sum_{t=1}^{n} u_t Y_{t-1}' - (\Pi_1 - \Pi_o) \sum_{t=1}^{n} Y_{t-1} Y_{t-1}' \right] \right\|_E = O_p(n^{-\frac{1}{2}})$. From the inequality in (10.32), we deduce that

$$
\widehat{\Pi}_n - \Pi_1 = O_p(n^{-\frac{1}{2}} + a_n). \tag{4.62}
$$

When $0 < r_o < m$, we can use (4.50), (4.53) and (4.57) in the proof of Corollary 4.1 and (10.29) in the proof of Corollary 3.2 to get

$$
\mu_{\min} \left\| \left(\widehat{\Pi}_n - \Pi_1 \right) B_n \right\|_E^2 - 2 \left\| \left(\widehat{\Pi}_n - \Pi_1 \right) B_n \right\|_E e_n \leq n a_n C_V \left\| \widehat{\Pi}_n - \Pi_1 \right\|_E, \tag{4.63}
$$

where $e_n = \left\| \left\{ \sum_{t=1}^{n} u_t Z_{t-1}' - \left[(\Pi_1 - \Pi_o) Q^{-1} \right] \sum_{t=1}^{n} Z_{t-1} Z_{t-1}' \right\} D_n \right\|_E = O_p(1)$. Note that

$$
\left\| \widehat{\Pi}_n - \Pi_1 \right\|_E = \left\| \left(\widehat{\Pi}_n - \Pi_1 \right) B_n B_n^{-1} \right\|_E \leq C n^{-\frac{1}{2}} \left\| \left(\widehat{\Pi}_n - \Pi_1 \right) B_n \right\|_E. \tag{4.64}
$$

From the inequalities in (4.63) and (4.64), we obtain

$$
\left(\widehat{\Pi}_n - \Pi_1 \right) B_n = O_p(n^{-\frac{1}{2}} + a_n), \tag{4.65}
$$

which finishes the proof. ■

Corollary 4.1 implies that the nonzero eigenvalues of Π_1 are estimated as nonzeros w.p.a.1. However, the matrix Π_1 may have more zero eigenvalues than Π_o. To ensure consistent cointegration rank selection, we need to show that Π_1 has $m - r_o$ zero eigenvalues which are estimated as zero w.p.a.1 and $r_o - r_1$ zero eigenvalues which are estimated as non-zeros w.p.a.1. From Lemma (4.2), we see that $\widehat{\Pi}_{ols}$ has $m - r_o$ eigenvalues which converge to zero at the rate n and $r_o - r_1$ eigenvalues which converge to zero at the rate $\sqrt{n}$. The different convergence rates of the estimates of the zero eigenvalues of Π_1 enable us to empirically distinguish the estimates of the $m - r_o$ zero eigenvalues of Π_1 from the estimates of the $r_o - r_1$ zero eigenvalues of Π_1. For example, we can define thresholding estimators of the eigenvalues of Π_1 in the following way

$$\widehat{\phi}_{th,k} = \begin{cases} 0 & \text{if } \left|\widehat{\phi}_k\right| \le n^{-\frac{5}{8}} \\ \widehat{\phi}_k & \text{otherwise} \end{cases},$$

where $\widehat{\phi}_k = \phi_k(\widehat{\Pi}_{ols})$. Let $\widetilde{\mathcal{S}}_{ols,\phi} = \{k : \phi_k(\widehat{\Pi}_{ols}) \neq 0, k = 1, ..., m\}$ be the index set of nonzero eigenvalues of $\widehat{\Pi}_{ols}$. As $n^{\frac{5}{8}}\left|\widehat{\phi}_k\right| = n^{\frac{1}{8}}\left|n^{\frac{1}{2}}\widehat{\phi}_k\right| \to_p \infty$ for all $k \in \widetilde{\mathcal{S}}_{ols,\phi}$, it follows that $\widehat{\phi}_{th,k} = \widehat{\phi}_k \ \forall k \in \widetilde{\mathcal{S}}_{ols,\phi}$ w.p.a.1. On the other hand, $n^{\frac{5}{8}}\left|\widehat{\phi}_k\right| = n^{-\frac{3}{8}}\left|n\widehat{\phi}_k\right| \to_p 0$ for all $k \in \widetilde{\mathcal{S}}^c_{ols,\phi}$, and hence $\widehat{\phi}_{th,k} = 0 \ \forall k \in \widetilde{\mathcal{S}}^c_{ols,\phi}$ w.p.a.1. From the arguments above, we see that consistent cointegration rank is achieved. We next show that the LS shrinkage estimator $\widehat{\Pi}_n$ has only $m - r_o$ zero eigenvalues w.p.a.1.

Assumption 4.2 *The penalty function* $\widehat{P}_{\lambda_n}(\cdot)$ *satisfies*

$$n^{\frac{1}{2}+v}\widehat{P}'_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)] = o_p(1)$$

for all $k \in \mathcal{S}_\phi$ *and some positive constant* v.

It is easy to see that if $n^{\frac{1+\omega}{2}}\lambda_n = o(1)$, then by Lemma (4.3) the adaptive Lasso penalty satisfies Assumption 4.2. For any $m \times m$ matrix Π, we have by definition

$$\Pi v_k = \phi_k v_k \text{ and } \Pi' u_k = \phi_k u_k,$$

where ϕ_k is the k-th largest eigenvalue of Π, and v_k and u_k are the related normalized right and left eigenvectors. Applying the chain rule, we get

$$d\Pi v_k + \Pi d v_k = v_k d\phi_k + \phi_k d v_k,$$

so that

$$u_k' d\Pi v_k = u_k' v_k d\phi_k, \tag{4.66}$$

where we use the fact that $u_k'\Pi = u_k'\phi_k$. It is easy to see that $\widetilde{\beta}_{1\perp}$ and $\widetilde{\alpha}_{1\perp}$ contain the right and left eigenvectors of the zero eigenvalues of Π_1 respectively. When $r_1 = m$, then the limit results $\phi_k(\widehat{\Pi}_n) \to_p \phi_k(\Pi_1)$, $k = 1, ..., m$ imply consistent cointegration rank selection immediately. When $r_1 < m$, both $\widetilde{\beta}_{1\perp}$ and $\widetilde{\alpha}_{1\perp}$ are non-zero. Suppose the (i, j)-th element of $\widetilde{\beta}_{1\perp}$ and (l, m)-*th* element of $\widetilde{\alpha}_{1\perp}$ are nonzero, then from the differentials in (4.66), we deduce that

$$\left.\frac{d\Pi_{jl}}{d\phi_k}\right|_{\Pi=\Pi_1} = \frac{(\widetilde{\beta}_{1\perp})'_{\cdot j}(\widetilde{\alpha}_{1\perp})_{l\cdot}}{(\widetilde{\beta}_{1\perp})_{ij}(\widetilde{\alpha}_{1\perp})_{lm}}. \tag{4.67}$$

The result in (4.67) is important for establishing the following result.

Corollary 4.3 *Under Assumption LP, RR, Assumption 3.3 and Assumption 4.2, we have*

$$\Pr\left(\phi_k(\widehat{\Pi}_n) \neq 0\right) \to 1 \tag{4.68}$$

for all $k \in \mathcal{S}_\phi$ *as* $n \to \infty$, *and*

$$\Pr\left(\phi_k(\widehat{\Pi}_n) = 0\right) \to 1 \tag{4.69}$$

for all $k \in \mathcal{S}_\phi^c$ *as* $n \to \infty$.

Proof. On the event $\{\phi_k(\widehat{\Pi}_n) = 0\}$, for some $k \in \mathcal{S}_\phi$, we have the following Karush-Kuhn-Tucker (KKT) optimality condition

$$\left| n^{-\frac{1}{2}+v} vec(\partial\widehat{\Pi}_n)' \left(Y_{-1} \otimes I_m\right) \left[\Delta y - \left(Y_{-1}' \otimes I_m\right) vec(\widehat{\Pi}_n)\right] \right| < \frac{n^{\frac{1}{2}+v}\widehat{P}'_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)]}{2}, \tag{4.70}$$

where $\partial\widehat{\Pi}_n$ is a matrix whose jk-th element equals $\frac{(\widehat{\beta}_{1\perp})'_{\cdot j}(\widehat{\alpha}_{1\perp})_{k\cdot}}{(\widehat{\beta}_{1\perp})_{ij}(\widehat{\alpha}_{1\perp})_{kl}}$ and whose other elements are zero. By Corollary 4.1 and continuous mapping, we have

$$\partial\widehat{\Pi}_n \to_p \frac{(\widetilde{\beta}_{1\perp})'_{\cdot j}(\widetilde{\alpha}_{1\perp})_{l\cdot}}{(\widetilde{\beta}_{1\perp})_{ij}(\widetilde{\alpha}_{1\perp})_{lm}} := \partial\Pi_1. \tag{4.71}$$

Denote

$$\begin{aligned}
\widehat{J} &= \frac{\sum_{t=1}^n Y_{t-1}u_t'}{n} - \frac{\sum_{t=1}^n Y_{t-1}Y_{t-1}'}{n}(\widehat{\Pi}_n - \Pi_o) \\
&= Q^{-1}\left[\frac{\sum_{t=1}^n Z_{t-1}w_t'}{n} - \frac{\sum_{t=1}^n Z_{t-1}Z_{t-1}'}{n}Q'^{-1}(\widehat{\Pi}_n - \Pi_o)'Q'\right]Q'^{-1},
\end{aligned}$$

where

$$\begin{aligned}
& \frac{\sum_{t=1}^{n} Z_{t-1} Z_{t-1}'}{n} Q'^{-1} (\widehat{\Pi}_n - \Pi_o)' Q' \\
= & \left(\begin{array}{cc} \widehat{S}_{11} & \widehat{S}_{12} \\ \widehat{S}_{21} & \widehat{S}_{22} \end{array} \right) \left(\begin{array}{cc} \beta_o'(\widehat{\Pi}_n - \Pi_o)\alpha_o(\beta_o'\alpha_o)^{-1} & \beta_o'\widehat{\Pi}_n(\alpha_{o\perp}'\beta_{o\perp})^{-1} \\ \alpha_{o\perp}'(\widehat{\Pi}_n - \Pi_o)\alpha_o(\beta_o'\alpha_o)^{-1} & \alpha_{o\perp}'\widehat{\Pi}_n\beta_{o\perp}(\alpha_{o\perp}'\beta_{o\perp})^{-1} \end{array} \right)' \\
= & \left(\begin{array}{cc} \widehat{S}_{11} & \widehat{S}_{12} \\ \widehat{S}_{21} & \widehat{S}_{22} \end{array} \right) \left(\begin{array}{cc} \Sigma_{z_1 z_1}^{-1} \Gamma_{z_1 w_1}(1) + o_p(1) & \Sigma_{z_1 z_1}^{-1} \Gamma_{z_1 w_2}(1) + o_p(1) \\ (\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\widehat{\Pi}_n'\beta_o & (\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\widehat{\Pi}_n'\alpha_{o\perp} \end{array} \right).
\end{aligned}$$

Hence

$$\begin{aligned}
\widehat{J}_{11} &= \frac{\sum_{t=1}^{n} Z_{1,t-1} w_{1t}'}{n} - \left[\widehat{S}_{11} \Sigma_{z_1 z_1}^{-1} \Gamma_{z_1 w_1}(1) + \widehat{S}_{12} (\beta_{o\perp}'\alpha_{o\perp})^{-1} \beta_{o\perp}' \widehat{\Pi}_n' \beta_o \right] + o_p(1) \\
&= \frac{\sum_{t=1}^{n} Z_{1,t-1} w_{1t}'}{n} - \widehat{S}_{11} \Sigma_{z_1 z_1}^{-1} \Gamma_{z_1 w_1}(1) + o_p(1) \\
&= \widehat{S}_{11} \left(\widehat{S}_{11}^{-1} - \Sigma_{z_1 z_1}^{-1} \right) \frac{\sum_{t=1}^{n} Z_{1,t-1} w_{1t}'}{n} + \widehat{S}_{11} \Sigma_{z_1 z_1}^{-1} \left(\frac{\sum_{t=1}^{n} Z_{1,t-1} w_{1t}'}{n} - \Gamma_{z_1 w_1}(1) \right) + o_p(1),
\end{aligned}$$

which implies that

$$n^{\frac{1}{2}+v} \widehat{J}_{11} \to_p \infty. \tag{4.72}$$

for any $v > 0$. Next note that

$$\begin{aligned}
\widehat{J}_{12} &= \frac{\sum_{t=1}^{n} Z_{1,t-1} w_{2t}'}{n} - \left[\widehat{S}_{11} \Sigma_{z_1 z_1}^{-1} \Gamma_{z_1 w_2}(1) + \widehat{S}_{12} (\beta_{o\perp}'\alpha_{o\perp})^{-1} \beta_{o\perp}' \widehat{\Pi}_n' \alpha_{o\perp} \right] + o_p(1) \\
&= \frac{\sum_{t=1}^{n} Z_{1,t-1} w_{2t}'}{n} - \widehat{S}_{11} \Sigma_{z_1 z_1}^{-1} \Gamma_{z_1 w_2}(1) + o_p(1) \\
&= \widehat{S}_{11} \left(\widehat{S}_{11}^{-1} - \Sigma_{z_1 z_1}^{-1} \right) \frac{\sum_{t=1}^{n} Z_{1,t-1} w_{2t}'}{n} + \widehat{S}_{11} \Sigma_{z_1 z_1}^{-1} \left(\frac{\sum_{t=1}^{n} Z_{1,t-1} w_{2t}'}{n} - \Gamma_{z_1 w_2}(1) \right) + o_p(1),
\end{aligned}$$

which implies that

$$n^{\frac{1}{2}+v} \widehat{J}_{12} \to_p \infty. \tag{4.73}$$

For the third term,

$$\begin{aligned}
\widehat{J}_{21} &= \frac{\sum_{t=1}^{n} Z_{2,t-1}w_{1t}'}{n} - \left[\widehat{S}_{21}\Sigma_{z_1z_1}^{-1}\Gamma_{z_1w_1}(1) + \widehat{S}_{22}(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\widehat{\Pi}_n'\beta_o\right] + o_p(1) \\
&= \frac{\sum_{t=1}^{n} Z_{2,t-1}w_{1t}'}{n} - \widehat{S}_{22}(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\left(\widehat{\Pi}_n - \Pi_{1,n}\right)'\beta_o \\
&\quad - \left[\widehat{S}_{21}\Sigma_{z_1z_1}^{-1}\Gamma_{z_1w_1}(1) + \widehat{S}_{22}(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\Pi_{1,n}'\beta_o\right] + o_p(1). \\
&= \left(\int B_{w_2}dB_{w_1}'\right) - \left(\int B_{w_2}dB_{w_2}'\right)(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\left(\widehat{\Pi}_n - \Pi_{1,n}\right)'\beta_o + o_p(1).
\end{aligned}$$

If $\widehat{\Pi}_n$ is the OLS estimator, then it is easy to see that

$$\left(\int B_{w_2}dB_{w_2}'\right)(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\left(\widehat{\Pi}_{ols} - \Pi_{1,n}\right)'\beta_o \to_d \left(\int B_{w_2}dB_{w_1}'\right).$$

However, as shown in Liao and Phillips (2010), when there are rank constraints imposed on $\widehat{\Pi}_n$, the limiting distribution of $\left(\int B_{w_2}dB_{w_2}'\right)(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\left(\widehat{\Pi}_n - \Pi_{1,n}\right)'\beta_o$ will be different from the above. Hence, we can deduce that

$$n^{\frac{1}{2}+v}\widehat{J}_{21} \to_p \infty. \tag{4.74}$$

For the fourth term,

$$\begin{aligned}
\widehat{J}_{22} &= \frac{\sum_{t=1}^{n} Z_{2,t-1}w_{2t}'}{n} - \left[\widehat{S}_{21}\Sigma_{z_1z_1}^{-1}\Gamma_{z_1w_2}(1) + \widehat{S}_{22}(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\widehat{\Pi}_n'\alpha_{o\perp}\right] + o_p(1) \\
&= \frac{\sum_{t=1}^{n} Z_{2,t-1}w_{2t}'}{n} - \widehat{S}_{22}(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\left(\widehat{\Pi}_n - \Pi_{1,n}\right)'\alpha_{o\perp} \\
&\quad - \left[\widehat{S}_{21}\Sigma_{\beta\beta}^{-1}\Gamma_{z_1w_2}(1) + \widehat{S}_{22}(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\Pi_{1,n}'\alpha_{o\perp}\right] + o_p(1). \\
&= \left(\int B_{w_2}B_{w_2}'\right) - \left(\int B_{w_2}B_{w_2}'\right)(\beta_{o\perp}'\alpha_{o\perp})^{-1}\beta_{o\perp}'\left(\widehat{\Pi}_n - \Pi_{1,n}\right)'\alpha_{o\perp} + o_p(1).
\end{aligned}$$

Using similar arguments to those in deriving (4.74), we have

$$n^{\frac{1}{2}+v}\widehat{J}_{22} \to_p \infty. \tag{4.75}$$

From the results in (4.71) and (4.72)-(4.75), we can deduce that

$$\left|n^{-\frac{1}{2}+v}vec(\partial\widehat{\Pi}_n)'\left(Y_{-1}\otimes I_m\right)\left[\Delta y - \left(Y_{-1}'\otimes I_m\right)vec(\widehat{\Pi}_n)\right]\right| \to_p \infty. \tag{4.76}$$

However, from Assumption 4.2, we have

$$\frac{n^{\frac{1}{2}+v}\widehat{P}'_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)]}{2} = o_p(1) \tag{4.77}$$

Results in (4.70), (4.76) and (4.77) imply that $\Pr\left(\bar{\phi}_k(\widehat{\Pi}_n) = 0\right) \to 0$, for any $k \in \mathcal{S}_\phi$, giving (4.68). The proof of (4.69) is similar to that of Theorem 3.3 and is therefore omitted. ■

Corollary 4.3 implies that $\widehat{\Pi}_n$ has $m - r_o$ eigenvalues estimated as zero w.p.a.1 and the other eigenvalues are estimated as nonzero w.p.a.1. Hence Corollary 4.3 gives us the following result immediately.

Theorem 4.4 (Sparsity-II) *Under the conditions of Corollary 4.3, we have*

$$\Pr\left(r(\widehat{\Pi}_n) = r_o\right) \to 1 \tag{4.78}$$

as $n \to \infty$*, where* $r(\widehat{\Pi}_n)$ *denotes the rank of* $\widehat{\Pi}_n$*.*

5 Extension II: VECM Estimation with Explicit Transient Dynamics

In this section, we are interested in the estimation of the VECM model

$$\Delta Y_t = \Pi_o Y_{t-1} + \sum_{j=1}^{p} B_{o,j}\Delta Y_{t-j} + u_t \tag{5.1}$$

with simultaneous cointegration rank selection and lag order selection. Denoting $B_o = (B_{o,1}, ..., B_{o,p})$, the unknown parameter matrices (Π_o, B_o) are estimated via the following penalized LS estimation

$$\begin{aligned}(\widehat{\Pi}_n, \widehat{B}_n) \quad = \quad & \underset{\Pi \in R^{m\times m}, B_j \in R^{m\times m}}{\arg\min} \sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum_{j=1}^{p} B_j \Delta Y_{t-j} \right\|_E^2 \\ & + n\sum_{j=1}^{p} \widehat{P}_{\lambda_n}(B_j) + n\sum_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi)\right]. \end{aligned} \tag{5.2}$$

In (5.2), $\widehat{P}_{\lambda_n}(\cdot)$ denotes the adaptive group Lasso penalty function, i.e.

$$\widehat{P}_{\lambda_n}(A) = \lambda_n \|A\|_E / ||\widehat{A}_n||_E^{\omega},$$

for any $m \times m$ matrix, where $\omega > 0$ and $\widehat{A}_n$ is some first-step estimator of A.

The adaptive group Lasso delivers the penalty to the shrinkage estimator groupwisely. For example, if $\|A\|_E \neq 0$ for some $m \times m$ matrix A and $\widehat{A}_n$ is a first step consistent estimator of A, then under condition $\lambda_n \to 0$, the penalty on the estimate of each component of A will go to zero asymptotically because

$$||\widehat{A}_n||_E^{\omega} \to_p \|A\|_E^{\omega} \neq 0,$$

even though the matrix A may have some zero elements. On the other hand, if $\|A\|_E = 0$, then the penalty on the estimate of each component of A will be nontrivial and it is possible to consistently recover the zero matrix A in the LS shrinkage estimation. Also note that when A is a scalar, then the adaptive group Lasso penalty function reduces to the adaptive Lasso penalty function.

To consistently select the true order of the lagged differences, B_o should be at least consistently estimable. Thus we assume that the error term u_t satisfies Assumption 3.1. Assumption 3.2 needs revision to accommodate the general structure in (5.1). Denote

$$C(\phi) = \Pi_o + \sum_{j=0}^{p} B_{o,j}(1-\phi)\phi^i$$

where $B_{o,0} = -I_m$.

Assumption 5.1 (RR) *(i) The determinantal equation $|C(\phi)| = 0$ has roots on or outside the unit circle; (ii) the matrix Π_o has rank r_o, with $0 \leq r_o \leq m$; (iii) the following $(m - r_o) \times (m - r_o)$ matrix*

$$\alpha'_{o,\perp} \left(I_m - \sum_{j=1}^{p} B_{o,j} \right) \beta_{o,\perp} \tag{5.3}$$

is nonsingular.

Under Assumption 5.1, the time series Y_t has following partial sum representation,

$$Y_t = C_B \sum_{s=1}^{t} u_s + \Xi(L)u_t + C_B Y_0 \tag{5.4}$$

where $C_B = \beta_{o,\perp}\left[\alpha'_{o,\perp}\left(I_m - \sum_{j=1}^{p} B_{o,j}\right)\beta_{o,\perp}\right]^{-1}\alpha'_{o,\perp}$ and $\Xi(L)u_t = \sum_{s=0}^{\infty}\Xi_s u_{t-s}$ is a stationary process. From the partial sum representation in (5.4), we deduce that $\beta'_o Y_t = \beta'_o \Xi(L)u_t$ and ΔY_{t-j} $(j = 0, ..., p)$ are stationary.

Define

$$Q_B := \begin{pmatrix} \beta'_o & 0 \\ 0 & I_{mp\times mp} \\ \alpha'_{o,\perp} & 0 \end{pmatrix}_{m(p+1)\times m(p+1)}$$

and denote $\Delta X_{t-1} = \left[\Delta Y'_{t-1}, ..., \Delta Y'_{t-p}\right]'$, then the model in (5.1) can be rewritten as

$$\Delta Y_t = \begin{bmatrix} \Pi_o & B_o \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ \Delta X_{t-1} \end{bmatrix} + u_t, \tag{5.5}$$

Denote

$$Z_{t-1} = Q_B \begin{bmatrix} Y_{t-1} \\ \Delta X_{t-1} \end{bmatrix} = \begin{bmatrix} Z_{1,t-1} \\ Z_{2,t-1} \end{bmatrix}, \tag{5.6}$$

where $Z_{1,t-1} = \begin{bmatrix} \beta'_o Y_{t-1} \\ \Delta X_{t-1} \end{bmatrix}$ is a stationary process and $Z_{2,t-1} = \alpha'_{o,\perp} Y_{t-1}$ consists the $I(1)$ components.

Lemma 5.1 *Under Assumption 3.1 and 5.1, the results in Lemma 10.1 are satisfied for* $Z_{1,t}$ *and* $Z_{2,t}$ *defined in (5.6).*

We first establish the asymptotic properties of the OLS estimator $\left(\widehat{\Pi}_{ols}, \widehat{B}_{ols}\right)$ of (Π_o, B_o) and the asymptotic properties of the eigenvalues of $\widehat{\Pi}_{ols}$. The estimate $\left(\widehat{\Pi}_{ols}, \widehat{B}_{ols}\right)$ has the following closed-form solution

$$\left(\widehat{\Pi}_{ols}, \widehat{B}_{ols}\right) = \begin{pmatrix} \widehat{S}_{y_0 y_1} & \widehat{S}_{y_0 x_0} \end{pmatrix} \begin{pmatrix} \widehat{S}_{y_1 y_1} & \widehat{S}_{y_1 x_0} \\ \widehat{S}_{x_0 y_1} & \widehat{S}_{x_0 x_0} \end{pmatrix}^{-1}, \tag{5.7}$$

where

$$\widehat{S}_{y_0y_1} = \frac{1}{n}\sum_{t=1}^{n}\Delta Y_t Y'_{t-1},\ \widehat{S}_{y_0x_0} = \frac{1}{n}\sum_{t=1}^{n}\Delta Y_t\Delta X'_{t-1},\ \widehat{S}_{y_1y_1} = \frac{1}{n}\sum_{t=1}^{n}Y_{t-1}Y'_{t-1},$$

$$\widehat{S}_{y_1x_0} = \frac{1}{n}\sum_{t=1}^{n}Y_{t-1}\Delta X'_{t-1},\ \widehat{S}_{x_0y_1} = \widehat{S}'_{y_1x_0} \text{ and } \widehat{S}_{x_0x_0} = \frac{1}{n}\sum_{t=1}^{n}\Delta X_{t-1}\Delta X'_{t-1}. \quad (5.8)$$

Denote $Y_- = (Y_0, ..., Y_{n-1})_{m\times n}$, $\Delta Y = (\Delta Y_1, ..., \Delta Y_n)_{m\times n}$ and $\widehat{M}_0 = I_m - \Delta X'\widehat{S}^{-1}_{x_0x_0}\Delta X$, where $\Delta X = (\Delta X_0, ..., \Delta X_{n-1})_{mp\times n}$, then $\widehat{\Pi}_{ols}$ has the explicit representation

$$\begin{aligned}\widehat{\Pi}_{ols} &= \left(\Delta Y\widehat{M}_0 Y'_-\right)\left(Y_-\widehat{M}_0 Y'_-\right)^{-1} \\ &= \Pi_o + \left(U\widehat{M}_0 Y'_-\right)\left(Y_-\widehat{M}_0 Y'_-\right)^{-1}, \end{aligned} \quad (5.9)$$

where $U = (u_1, ..., u_n)_{m\times n}$. Recall that $[\phi_1(\widehat{\Pi}_{ols}), ..., \phi_m(\widehat{\Pi}_{ols})]$ and $[\phi_1(\Pi_o), ..., \phi_m(\Pi_o)]$ are the ordered eigenvalues of $\widehat{\Pi}_{ols}$ and Π_o respectively, where $\phi_j(\Pi_o) = 0$ $(j = 1, ..., m - r_o)$.

Lemma 5.2 *Suppose that Assumption 3.1 and 5.1 hold.*

(a) Define $D_{n,B} = diag(n^{\frac{1}{2}}I_{r_o+mp}, nI_{m-r_o})$*, then* $\left[(\widehat{\Pi}_{ols}, \widehat{B}_{ols}) - (\Pi_o, B_o)\right]Q_B^{-1}D_{n,B}$ *has the following limit distribution*

$$\left(\ N(0, \Omega_u\otimes\Sigma_{z_1z_1})\Sigma^{-1}_{z_1z_1}, \quad \left(\int B_{w_2}dB'_u\right)'\left(\int B_{w_2}B'_{w_2}\right)^{-1}\ \right); \quad (5.10)$$

(b) the eigenvalues of $\widehat{\Pi}_{ols}$ *satisfy Lemma 3.1.(b);*

(c) the first $m - r_o$ *ordered eigenvalues of* $\widehat{\Pi}_{ols}$ *satisfy Lemma 3.1.(c).*

Proof. To prove (a), start by defining $\widehat{S}_{uy_1} = \frac{1}{n}\sum_{t=1}^{n}u_tY'_{t-1}$ and $\widehat{S}_{ux_0} = \frac{1}{n}\sum_{t=1}^{n}u_t\Delta X'_{t-1}$. From the expression in (5.8), we get

$$\begin{aligned}&\left[(\widehat{\Pi}_{ols}, \widehat{B}_{ols}) - (\Pi_o, B_o)\right]Q_B^{-1}D_{n,B} \\ = &\left(\begin{array}{cc}\widehat{S}_{uy_1} & \widehat{S}_{ux_0}\end{array}\right)Q'_BD^{-1}_{n,B}\left[D^{-1}_{n,B}Q_B\left(\begin{array}{cc}\widehat{S}_{y_1y_1} & \widehat{S}_{y_1x_0} \\ \widehat{S}_{x_0y_1} & \widehat{S}_{x_0x_0}\end{array}\right)Q'_BD^{-1}_{n,B}\right]^{-1}. \end{aligned} \quad (5.11)$$

Note that

$$\begin{pmatrix} \widehat{S}_{uy_1} & \widehat{S}_{ux_0} \end{pmatrix} Q_B' D_{n,B}^{-1} = U \left[Q_B \begin{pmatrix} Y_- \\ \Delta X \end{pmatrix} \right]' D_{n,B}^{-1} = \begin{pmatrix} n^{-\frac{1}{2}} U Z_1' & n^{-1} U Z_2' \end{pmatrix} \tag{5.12}$$

and

$$D_{n,B}^{-1} Q_B \begin{pmatrix} \widehat{S}_{y_1y_1} & \widehat{S}_{y_1x_0} \\ \widehat{S}_{x_0y_1} & \widehat{S}_{x_0x_0} \end{pmatrix} Q_B' D_{n,B}^{-1} = \begin{pmatrix} n^{-1} \sum_{t=1}^{n} Z_{1,t} Z_{1,t}' & n^{-\frac{3}{2}} \sum_{t=1}^{n} Z_{1,t} Z_{2,t}' \\ n^{-\frac{3}{2}} \sum_{t=1}^{n} Z_{2,t} Z_{1,t}' & n^{-2} \sum_{t=1}^{n} Z_{2,t} Z_{2,t}' \end{pmatrix}, \tag{5.13}$$

where $Z_1 = (Z_{1,0}, ..., Z_{1,n-1})$ and $Z_2 = (Z_{2,0}, ..., Z_{2,n-1})$. Now the result in (5.10) follows by applying the Lemma 5.1.

The proofs of (b) and (c) are similar to those of Lemma 3.1.(b)-(c) and are therefore omitted. ■

Denote the index set of the zero components in B_o as $\mathcal{S}_B^c$ such that $\|B_{o,j}\| = 0$ for all $j \in \mathcal{S}_B^c$ and $\|B_{o,j}\| \neq 0$ otherwise. We next derive the asymptotic properties of the LS shrinkage estimator defined in (5.2).

Lemma 5.3 *If Assumption 3.1 and 5.1 are satisfied and $\lambda_n = o(1)$, then the LS shrinkage estimator $(\widehat{\Pi}_n, \widehat{B}_n)$ satisfies*

$$\left[(\widehat{\Pi}_n, \widehat{B}_n) - (\Pi_o, B_o) \right] Q_B^{-1} D_{n,B} = O_p(1 + n^{\frac{1}{2}} \lambda_n). \tag{5.14}$$

Proof. Let $\Theta = (\Pi, B)$ and

$$\begin{aligned} V_{3,n}(\Theta) &= \sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum_{j=1}^{p} B_j \Delta Y_{t-j} \right\|_E^2 \\ &\quad + n \sum_{k=1}^{p} \widehat{P}_{\lambda_n}(B_j) + n \sum_{k=1}^{m} \widehat{P}_{\lambda_n} \left[\bar{\phi}_k(\Pi) \right], \\ &= V_{4,n}(\Theta) + n \sum_{k=1}^{p} \widehat{P}_{\lambda_n}(B_j) + n \sum_{k=1}^{m} \widehat{P}_{\lambda_n} \left[\bar{\phi}_k(\Pi) \right]. \end{aligned}$$

Set $\widehat{\Theta}_n = (\widehat{\Pi}_n, \widehat{B}_n)$, and then by definition $V_{3,n}(\widehat{\Theta}_n) \leq V_{3,n}(\Theta_o)$, so that

$$\left\{vec\left[n^{-\frac{1}{2}}(\widehat{\Theta}_n - \Theta_o)Q_B^{-1}D_{n,B}\right]\right\}' W_n \left\{vec\left[n^{-\frac{1}{2}}(\widehat{\Theta}_n - \Theta_o)Q_B^{-1}D_{n,B}\right]\right\}$$
$$: \quad +2\left\{vec\left[n^{-\frac{1}{2}}(\widehat{\Theta}_n - \Theta_o)Q_B^{-1}D_{n,B}\right]\right\}' d_n \leq (e_{1,n} + e_{2,n}) \tag{5.15}$$

where $W_n = D_{n,B}^{-1}\sum_{t=1}^{n} Z_{t-1}Z_{t-1}' D_{n,B}^{-1} \otimes I_{m(p+1)}$, $d_n = vec\left(n^{-\frac{1}{2}}D_{n,B}^{-1}\sum_{t=1}^{n} Z_{t-1}u_t'\right)$, $e_{1,n} = \sum_{k\in\mathcal{S}_B^c}\left[\widehat{P}_{\lambda_n}(B_{o,j}) - \widehat{P}_{\lambda_n}(\widehat{B}_{n,j})\right]$ and $e_{2,n} = \sum_{k\in\mathcal{S}_\phi^c}\left\{\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right] - \widehat{P}_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)]\right\}$. Applying the Cauchy-Schwarz inequality to (5.15), we deduce that

$$\mu_{n,\min}\left\|n^{-\frac{1}{2}}(\widehat{\Theta}_n - \Theta_o)Q_B^{-1}D_{n,B}\right\|_E^2 - \left\|n^{-\frac{1}{2}}(\widehat{\Theta}_n - \Theta_o)Q_B^{-1}D_{n,B}\right\|_E d_n - (e_{1,n} + e_{2,n}) \leq 0, \tag{5.16}$$

where $\mu_{n,\min}$ denotes the smallest eigenvalue of W_n which is bounded away from zero w.p.a.1. By the definition of the adaptive group Lasso penalty function, the consistency of $\widehat{\Theta}_{ols}$ and the Slutsky Theorem, we find that

$$e_{1,n} \leq \sum_{k\in\mathcal{S}_B^c}\widehat{P}_{\lambda_n}(B_{o,j}) = o_p(1) \text{ and } e_{2,n} \leq \sum_{k\in\mathcal{S}_\phi^c}\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right] = o_p(1). \tag{5.17}$$

If $r_o = m$, then $D_{n,B} = n^{\frac{1}{2}}I_{m(p+1)}$ and hence $n^{-\frac{1}{2}}D_{n,B} = I_{m(p+1)}$ and $n^{-\frac{1}{2}}D_{n,B}^{-1} = n^{-1}I_{m(p+1)}$. As $Z_{t-1}u_t'$ is a stationary martingale difference, we have by the LLN,

$$d_n = vec\left(n^{-1}\sum_{t=1}^{n} Z_{t-1}u_t'\right) = o_p(1). \tag{5.18}$$

The consistency of $\widehat{\Theta}_n$ is implied by the inequality in (5.16) and the results in (5.17)-(5.18). If $r_o \in [0, m)$, then by Lemma 5.1, we can deduce that

$$d_n = vec\left(n^{-\frac{1}{2}}D_{n,B}^{-1}\sum_{t=1}^{n} Z_{t-1}u_t'\right) = o_p(1). \tag{5.19}$$

The consistency of $\widehat{\Theta}_n$ is implied by the inequality in (5.16) and the results in (5.17) and (5.19).

We next derive the convergence rate of the LS shrinkage estimator $\widehat{\Theta}_n$. Using the similar argument in the proof of Theorem 3.2, we get

$$\sum_{k\in\mathcal{S}_\phi}\left(\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]-\widehat{P}_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)]\right)\leq n^{-\frac{1}{2}}C\max_{k\in\mathcal{S}_\phi}\left\{\widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right\}\left\|\left(\widehat{\Theta}_n-\Theta_o\right)Q_B^{-1}D_{n,B}\right\|, \tag{5.20}$$

where C is some generic positive constant. Similarly

$$\sum_{k\in\mathcal{S}_B}\left[\widehat{P}_{\lambda_n}(B_{o,j})-\widehat{P}_{\lambda_n}(\widehat{B}_{n,j})\right]\leq n^{-\frac{1}{2}}C\max_{k\in\mathcal{S}_B}\left\{\widehat{P}'_{\lambda_n}(B_{o,j})\right\}\left\|\left(\widehat{\Theta}_n-\Theta_o\right)Q_B^{-1}D_{n,B}\right\|. \tag{5.21}$$

Combining the results in (5.20)-(5.21), we get

$$n(e_{1,n}+e_{2,n})\leq n^{\frac{1}{2}}a_n\left\|\left(\widehat{\Theta}_n-\Theta_o\right)Q_B^{-1}D_{n,B}\right\|, \tag{5.22}$$

where $a_n=\max_{k\in\mathcal{S}_\phi}\left\{\widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right\}\vee\max_{k\in\mathcal{S}_B}\left\{\widehat{P}'_{\lambda_n}(B_{o,j})\right\}=O_p(\lambda_n)$. From Lemma 5.1, we have

$$n^{\frac{1}{2}}d_n=vec\left(D_{n,B}^{-1}\sum_{t=1}^{n}Z_{t-1}u'_t\right)=O_P(1). \tag{5.23}$$

Using the inequality in (5.22), we can rewrite the inequality in (5.16) as

$$\mu_{n,\min}\left\|(\widehat{\Theta}_n-\Theta_o)Q_B^{-1}D_{n,B}\right\|_E^2-\left\|(\widehat{\Theta}_n-\Theta_o)Q_B^{-1}D_{n,B}\right\|_E(d_n+n^{\frac{1}{2}}a_n)\leq 0, \tag{5.24}$$

which combined with the result in (5.23) imply the rate claimed in (5.14). ■

The results in Lemma 5.3 imply that the LS shrinkage estimator $(\widehat{\Pi}_n,\widehat{B}_n)$ may have the same convergence rate of the OLS estimator $(\widehat{\Pi}_{ols},\widehat{B}_{ols})$. We next show that if the tuning parameter λ_n converges to zero not too fast, then the zero eigenvalues of Π_o and zero matrices in B_o are estimated as zero w.p.a.1. Let the smallest $m-r_o$ eigenvalues of $\widehat{\Pi}_n$ be indexed by $\mathcal{S}_{n,\phi}^c$ and the zero matrix in $\widehat{B}_n$ by $\mathcal{S}_{n,B}^c$.

Theorem 5.1 *Suppose that Assumption 3.1 and 5.1 hold. If the tuning parameter λ_n and ω satisfy the conditions $\omega>\frac{1}{2}$, $n^{\frac{1}{2}}\lambda_n=o(1)$, $\lambda_n n^{\omega}\to\infty$ and $n^{\frac{1+\omega}{2}}\lambda_n\to\infty$, then we have*

$$\Pr\left(\phi_k(\widehat{\Pi}_n)=0\right)\to 1 \tag{5.25}$$

for all $k\in\mathcal{S}_\phi^c$, and

$$\Pr\left(\widehat{B}_{n,j}=\mathbf{0}_{m\times m}\right)\to 1 \tag{5.26}$$

for all $j \in \mathcal{S}_B^c$.

Proof. The first result can be proved using similar arguments to those in the proof of Theorem 3.3 and hence is omitted. We next show the second result. Note that

$$
\begin{aligned}
& \sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum_{j=1}^{p} B_j \Delta Y_{t-j} \right\|_E^2 - \sum_{t=1}^{n} \|u_t\|_E^2 \\
= \;& vec(\Theta - \Theta_o)' \left(Q_B^{-1} \sum_{t=1}^{n} Z_{t-1} Z_{t-1}' Q_B^{-1\prime} \otimes I_m \right) vec(\Theta - \Theta_o) \\
& -2 vec(\Theta - \Theta_o)' vec \left(Q_B^{-1} \sum_{t=1}^{n} \Delta Z_t u_t' \right). \qquad (5.27)
\end{aligned}
$$

On the event $\{\widehat{B}_{n,j} \neq \mathbf{0}_{m\times m}\}$ for some $j \in \mathcal{S}_B^c$, we have the following KKT optimality condition,

$$
\frac{n\lambda_n}{2||\widehat{B}_{ols,j}||_E^{\omega}||\widehat{B}_{n,j}||_E} \widehat{B}_{n,j} = \overline{V} \left(\widehat{W}_B \otimes I_m \right) vec(\widehat{\Theta}_n - \Theta_o) - \overline{V} vec \left(Q_B^{-1} \sum_{t=1}^{n} Z_{t-1} u_t' \right) \quad (5.28)
$$

where $\overline{V} = diag(V_1, ..., V_{p+1})$, $V_{j+1} = I_m$ and $V_{-(j+1)} = \mathbf{0}_m$ and $\widehat{W}_B = Q_B^{-1} \sum_{t=1}^{n} Z_{t-1} Z_{t-1}' Q_B^{-1\prime}$. By the definition of $\overline{V}$, we have $\overline{V} = V^* \otimes I_m$ where V^* is a $(p+1) \times (p+1)$ diagonal matrix whose $j+1$-th diagonal element is 1 and zero otherwise. Hence, results in (5.28) imply that

$$
\frac{n^{\frac{1+\omega}{2}} \lambda_n}{2||n^{\frac{1}{2}} \widehat{B}_{ols,j}||_E^{\omega}} = \left\| \left(\frac{V^* \widehat{W}_B}{n^{\frac{1}{2}}} \otimes I_m \right) vec(\widehat{\Theta}_n - \Theta_o) - vec \left(\frac{V^* Q_B^{-1} \sum_{t=1}^{n} Z_{t-1} u_t'}{n^{\frac{1}{2}}} \right) \right\|_E . \quad (5.29)
$$

By Lemma 5.1 and Lemma 5.3, we deduce that

$$
\begin{aligned}
& \left(\frac{V^* \widehat{W}_B}{n^{\frac{1}{2}}} \otimes I_m \right) vec(\widehat{\Theta}_n - \Theta_o) \\
= \;& \left(\frac{V^* \widehat{W}_B Q_B' D_{n,B}^{-1}}{n^{\frac{1}{2}}} \otimes I_m \right) vec \left[\left(\widehat{\Theta}_n - \Theta_o \right) Q_B^{-1} D_{n,B} \right] = O_p(1), \qquad (5.30)
\end{aligned}
$$

and

$$vec\left(\frac{V^{*}Q_{B}^{-1}\sum_{t=1}^{n}Z_{t-1}u_{t}'}{n^{\frac{1}{2}}}\right)=O_{p}(1). \tag{5.31}$$

However, as $n^{\frac{1+\omega}{2}}\lambda_n\rightarrow\infty$, it follows by Lemma 5.2 and the Slutsky Theorem that

$$\frac{n^{\frac{1+\omega}{2}}\lambda_n}{2||n^{\frac{1}{2}}\widehat{B}_{ols,j}||_E^{\omega}}\rightarrow_p\infty. \tag{5.32}$$

Now, using the results in (5.29)-(5.31), we can deduce that the event $\{\widehat{B}_{n,j}\neq\mathbf{0}_{m\times m}\}$ $\forall j\in\mathcal{S}_B^c$ has zero probability with $n\rightarrow\infty$, which finishes the proof. ■

Theorem 5.1 indicates that the zero eigenvalues of Π_o and the zero matrices in B_o are estimated as zeros w.p.a.1. Hence Lemma 5.3 and Theorem 5.1 imply consistent cointegration rank selection and consistent lag order selection.

Theorem 5.2 *Under the conditions of Theorem 5.1, we have*

$$\Pr\left(\mathcal{S}_{n,\phi}=\mathcal{S}_{\phi}\ and\ \mathcal{S}_{n,B}=\mathcal{S}_{B}\right)\rightarrow1, \tag{5.33}$$

as $n\rightarrow\infty$.

As

$$Q_B=\begin{pmatrix} I_{r_o} & 0 & 0\\ 0 & 0 & I_{mp}\\ 0 & I_{m-r_o} & 0\end{pmatrix}diag(Q,I_{mp}),$$

it is easy to see that

$$Q_B^{-1}=\begin{pmatrix}\alpha_o(\beta_o'\alpha_o)^{-1} & 0 & \beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\\ 0 & I_{mp} & 0\end{pmatrix}.$$

To ensure identification, we normalize β_o as $\beta_o=[I_{r_o},O_{r_o}]'$, where O_{r_o} is some $r_o\times(m-r_o)$ matrix such that $\Pi_o=\alpha_o\beta_o'=[\alpha_o,\alpha_oO_{r_o}]$. From Lemma 5.3, we have

$$\begin{pmatrix} n^{\frac{1}{2}}(\widehat{\Pi}_n-\Pi_o)\alpha_o(\beta_o'\alpha_o)^{-1} & n^{\frac{1}{2}}(\widehat{B}_n-B_o) & n(\widehat{\Pi}_n-\Pi_o)\beta_{o,\perp}(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\end{pmatrix}=O_p(1),$$

which implies that

$$n\left(\widehat{O}_n - O_o\right) = O_p(1), \tag{5.34}$$

$$n^{\frac{1}{2}}(\widehat{B}_n - B_o) = O_p(1), \tag{5.35}$$

$$n^{\frac{1}{2}}(\widehat{\alpha}_n - \alpha_o) = O_p(1). \tag{5.36}$$

To conclude this section, we derive the centered limit distribution of the shrinkage estimator $\widehat{\Theta}_{\mathcal{S}} = \left(\widehat{\Pi}_n, \widehat{B}_{\mathcal{S}_B}\right)$ in the following theorem, where $\widehat{B}_{\mathcal{S}_B}$ denotes the LS shrinkage estimator of the nonzero matrices in B_o. Denote $I_{\mathcal{S}_B} = diag(I_{1,m}, ..., I_{d_{\mathcal{S}_B},m})$ where the $I_{j,m}$ ($j = 1, ..., d_{\mathcal{S}_B}$) are $m \times m$ identity matrices and $d_{\mathcal{S}_B}$ is the dimensionality of the index set S_B. Define

$$Q_{\mathcal{S}} := \begin{pmatrix} \beta_o' & 0 \\ 0 & I_{\mathcal{S}_B} \\ \alpha_{o,\perp}' & 0 \end{pmatrix} \text{ and } D_{\mathcal{S}} = diag(n^{\frac{1}{2}} I_{r_o}, n^{\frac{1}{2}} I_{\mathcal{S}_B}, n I_{m-r_o}),$$

where the identity matrix $I_{\mathcal{S}_B} = I_{md_{\mathcal{S}_B}}$ in Q_S serves to accommodate the nonzero matrices in B_o.

Theorem 5.3 *Under conditions of Theorem 5.1, we have*

$$\left(\widehat{\Theta}_{\mathcal{S}} - \Theta_{o,\mathcal{S}}\right) Q_{\mathcal{S}}^{-1} D_{\mathcal{S}}^{-1} \to_d \left(\begin{array}{cc} N(0, \Omega_u \otimes \Sigma_{z_1 z_1})\Sigma_{z_1 z_1}^{-1} & \alpha_o U_2^* (\alpha_{o,\perp}' \beta_{o,\perp})^{-1} \end{array} \right), \tag{5.37}$$

where

$$U_2^* = (\alpha_o' \alpha_o)^{-1} \alpha_o' \left(\int dB_u B_{w_2}' \right) \left(\int B_{w_2} B_{w_2}' \right)^{-1} (\alpha_{o,\perp}' \beta_{o,\perp}). \tag{5.38}$$

Proof. From the results of Theorem 5.1, we deduce that $\widehat{\alpha}_n$, $\widehat{\beta}_n$ and $\widehat{B}_{\mathcal{S}_B}$ minimize the following criterion function w.p.a.1,

$$V_{3,n}(\Theta_{\mathcal{S}}) = \sum_{t=1}^{n} \left\| \Delta Y_t - \alpha\beta' Y_{t-1} - \sum_{j \in \mathcal{S}_B} B_j \Delta Y_{t-j} \right\|_E^2 + n \sum_{k \in \mathcal{S}_\phi} \widehat{P}_{\lambda_n} \left[\bar{\phi}_k(\alpha\beta') \right] + n \sum_{j \in \mathcal{S}_B} \widehat{P}_{\lambda_n}(B_j).$$

Define $U_{1,n}^* = \sqrt{n}\,(\widehat{\alpha}_n - \alpha_o)$, $U_{2,n} = \left[\mathbf{0}_{r_o}, U_{2,n}^*\right]'$, where $U_{2,n}^* = n\left(\widehat{O}_n - O_o\right)$ and $U_{3,n}^* =$

$\sqrt{n}\left(\widehat{B}_{\mathcal{S}_B} - B_{o,\mathcal{S}_B}\right)$ then

$$\begin{aligned} & \left[\left(\widehat{\Pi}_n - \Pi_o\right), \left(\widehat{B}_{\mathcal{S}_B} - B_{o,\mathcal{S}_B}\right)\right] Q_{\mathcal{S}}^{-1} D_{\mathcal{S}}^{-1} \\ = & \left[n^{-\frac{1}{2}} \widehat{\alpha}_n U_{2,n} \alpha_o (\beta_o' \alpha_o)^{-1} + U_{1,n}^*, U_{3,n}^*, \ \widehat{\alpha}_n U_{2,n} \beta_{o,\perp} (\alpha_{o,\perp}' \beta_{o,\perp})^{-1}\right]. \end{aligned}$$

Denote

$$\Pi_n(U) = \left[n^{-\frac{1}{2}} \widehat{\alpha}_n U_2 \alpha_o (\beta_o' \alpha_o)^{-1} + U_1, U_3, \widehat{\alpha}_n U_2 \beta_{o,\perp} (\alpha_{o,\perp}' \beta_{o,\perp})^{-1}\right],$$

then by definition, $U_n^* = \left(U_{1,n}^*, U_{2,n}^*, U_{3,n}^*\right)$ minimizes the following criterion function

$$\begin{aligned} V_{4,n}(U) = & \sum_{t=1}^{n} \left(\|u_t - \Pi_n(U) D_{\mathcal{S}} Z_{\mathcal{S},t-1}\|_E^2 - \|u_t\|_E^2\right) \\ & + n \sum_{k \in \mathcal{S}_\phi} \left\{\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_n(U) D_{\mathcal{S}} Q_{\mathcal{S}} L_1 + \Pi_o)\right] - \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right\} \\ & + n \sum_{j \in \mathcal{S}_B} \left\{\widehat{P}_{\lambda_n}\left[\Pi_n(U) D_{\mathcal{S}} Q_{\mathcal{S}} L_{j+1} + B_{o,j}\right] - \widehat{P}_{\lambda_n}(B_{o,j})\right\}, \end{aligned}$$

where $L_j = diag(A_1, ..., A_{p+1})$ with $A_j = I_m$ and $A_{-j} = 0$ otherwise and $Z_{\mathcal{S},t-1} = Q_{\mathcal{S}} \begin{pmatrix} Y_{t-1} \\ \Delta X_{\mathcal{S}_B,t-1} \end{pmatrix}$.

For any compact set $K \in R^{m \times r_o} \times R^{r_o \times (m-r_o)} \times R^{m \times md_{\mathcal{S}_B}}$ and any $U \in K$, there is

$$\Pi_n(U) D_{\mathcal{S}} Q_{\mathcal{S}} = O_p(n^{-\frac{1}{2}}).$$

Hence using similar arguments in the proof of Theorem 3.6, we can deduce that

$$n \sum_{k \in \mathcal{S}_\phi} \left\{\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_n(U) D_{\mathcal{S}} Q_{\mathcal{S}} L_1 + \Pi_o)\right] - \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right\} = o_p(1) \tag{5.39}$$

and

$$n \sum_{j \in \mathcal{S}_B} \left\{\widehat{P}_{\lambda_n}\left[\Pi_n(U) D_{\mathcal{S}} Q_{\mathcal{S}} L_{j+1} + B_{o,j}\right] - \widehat{P}_{\lambda_n}(B_{o,j})\right\} = o_p(1) \tag{5.40}$$

uniformly over $U \in K$.

Next, note that

$$\Pi_n(U) \to_p \left[U_1, U_3, \alpha_o U_2 (\alpha_{o,\perp}' \beta_{o,\perp})^{-1}\right] := \Pi_\infty(U) \tag{5.41}$$

uniformly over $U \in K$. By Lemma 5.1 and (5.41), we can deduce that

$$\begin{aligned} &\sum_{t=1}^{n}\left(\|u_t - \Pi_n(U)D_{\mathcal{S}}Z_{\mathcal{S},t-1}\|_E^2 - \|u_t\|_E^2\right) \\ \rightarrow_d\ & vec\left[\Pi_\infty(U)\right]'\left[\begin{pmatrix} \Sigma_{z_1z_1} & 0 \\ 0 & \int B_{w_2}B_{w_2}' \end{pmatrix} \otimes I_m\right] vec\left[\Pi_\infty(U)\right] \\ & -2vec\left[\Pi_\infty(U)\right]' vec\left[(B_{1,m}, B_{2,m})\right] := V_2(U) \end{aligned} \tag{5.42}$$

uniformly over $U \in K$, where $B_{1,m} = N(0, \Omega_u \otimes \Sigma_{z_1z_1})$ and $B_{2,m} = \left(\int B_{w_2}dB_u'\right)'$.

Note that $V_2(U)$ can be rewritten as

$$\begin{aligned} V_2(U) &= vec(U_1, U_3)'\left(\Sigma_{z_1z_1} \otimes I_m\right) vec(U_1, U_3) \\ &\quad + vec(U_2)'\left[(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\int B_{w_2}B_{w_2}'(\beta_{o,\perp}'\alpha_{o,\perp})^{-1} \otimes \alpha_o'\alpha_o\right] vec(U_2) \\ &\quad - 2vec(U_1, U_3)'vec\left(B_{1,m}\right) - 2vec(U_2)'vec\left[B_{2,m}(\beta_{o,\perp}'\alpha_{o,\perp})^{-1}\right]. \end{aligned} \tag{5.43}$$

The expression in (5.43) makes it clear that $V_2(U)$ is uniquely minimized at (U_1^*, U_2^*, U_3^*), where

$$(U_1^*, U_3^*) = B_{1,m}\Sigma_{z_1z_1}^{-1} \text{ and } U_2^* = (\alpha_o'\alpha_o)^{-1}\left(\alpha_o'B_{2,m}\right)\left(\int B_{w_2}B_{w_2}'\right)^{-1}(\alpha_{o,\perp}'\beta_{o,\perp}). \tag{5.44}$$

From (5.34) and (5.36), we can see that U_n^* is asymptotically tight. Invoking the ACMT, we deduce that $U_n^* \rightarrow_d U^*$. The results in (5.37) follow by applying the CMT. ■

6 Adaptive Selection of the Tuning Parameters

This section develops a data-driven procedure of selecting the tuning parameter λ_n. Given λ, we denote the rank of the shrinkage estimator $\widehat{\Pi}_n(\lambda)$ as $\widehat{r}_n(\lambda)$ and let $\mathcal{S}_{B,\lambda}$ be the index set of the nonzero matrices in the shrinkage estimator $\widehat{B}_n$. Define

$$IC_n(\lambda) = \min_{\substack{\Pi \in R^{m\times m}, r(\Pi) = \widehat{r}_n(\lambda) \\ B_j \in R^{m\times m}}}\left\{\sum_{t=1}^{n}\left\|\Delta Y_t - \Pi Y_{t-1} - \sum_{j\in\mathcal{S}_{B,\lambda}} B_j\Delta Y_{t-j}\right\|_E^2\right\} - C_n\mathcal{K}\left[\widehat{r}_n(\lambda), \widehat{p}_n(\lambda)\right], \tag{6.1}$$

where $\widehat{p}_n(\lambda)$ denotes the dimensionality of the index set $\mathcal{S}_{B,\lambda}$, $\mathcal{K}(r,p)$ is a strictly decreasing function in r and p and C_n is a positive sequence. We propose to select the tuning parameter λ by minimizing the information criterion $IC_n(\lambda)$ defined in (6.1), i.e.

$$\widehat{\lambda}_n^* = \arg\min_{\lambda\in\Omega_n} IC_n(\lambda), \tag{6.2}$$

where $\Omega_n = [0, \lambda_{\max,n}]$ and $\lambda_{\max,n}$ is a positive sequence such that $\lambda_{\max,n} = o(1)$.

Note that the first term in (6.1) is used to measure the fit of the selected model and the second term gives extra bonus to select smaller cointegration rank and smaller number of lagged differences. When an overfitted model is selected, then we can show that

$$\begin{aligned} & \min_{\substack{\Pi\in R^{m\times m}, r(\Pi)=\widehat{r}_n(\lambda) \\ B_j\in R^{m\times m}}} \left\{ \sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum_{j\in\mathcal{S}_{B,\lambda}} B_j \Delta Y_{t-j} \right\|_E^2 \right\} \\ & - \min_{\substack{\Pi\in R^{m\times m}, r(\Pi)=r_o \\ B_j\in R^{m\times m}}} \left\{ \sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum_{j\in\mathcal{S}_{B}} B_j \Delta Y_{t-j} \right\|_E^2 \right\} \\ = \;& O_p(1), \end{aligned} \tag{6.3}$$

while if $C_n \to \infty$, then

$$C_n \left\{ \mathcal{K}(r_o, p_o) - \mathcal{K}\left[\widehat{r}_n(\lambda), \widehat{p}_n(\lambda)\right] \right\} \to \infty \tag{6.4}$$

Results in (6.3)-(6.4) imply that overfitted models will not be selected if the tuning parameter $\widehat{\lambda}_n^*$ is selected by minimizing the information criterion in (6.1). On the other hand, if an underfitted model is selected, then we can show that

$$\begin{aligned} & \min_{\substack{\Pi\in R^{m\times m}, r(\Pi)=\widehat{r}_n(\lambda) \\ B_j\in R^{m\times m}}} \left\{ \frac{1}{n}\sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum_{j\in\mathcal{S}_{B,\lambda}} B_j \Delta Y_{t-j} \right\|_E^2 \right\} \\ & - \min_{\substack{\Pi\in R^{m\times m}, r(\Pi)=r_o \\ B_j\in R^{m\times m}}} \left\{ \frac{1}{n}\sum_{t=1}^{n} \left\| \Delta Y_t - \Pi Y_{t-1} - \sum_{j\in\mathcal{S}_{B}} B_j \Delta Y_{t-j} \right\|_E^2 \right\} \\ \to \;& {}_pC, \end{aligned} \tag{6.5}$$

where C is some generic positive constant, while if $C_n/n \to 0$, then

$$\frac{C_n}{n}\left\{\mathcal{K}(r_o, p_o) - \mathcal{K}\left[\widehat{r}_n(\lambda), \widehat{p}_n(\lambda)\right]\right\} \to 0. \tag{6.6}$$

Results in (6.3)-(6.4) imply that underfitted models will not be selected if the tuning parameter $\widehat{\lambda}^*_n$ is selected by minimizing the information criterion in (6.1). From above discussion, we see that the following assumption is important for the LS shrinkage estimator based on $\widehat{\lambda}^*_n$ to enjoy oracle-like properties.

Assumption 6.1 *(i) $\mathcal{K}(r, p)$ is a bounded and strictly decreasing function in r and p; (ii) $C_n \to \infty$ as the sample size $n \to \infty$ and $C_n = o(n)$.*

We next show that the LS shrinkage estimation based on $\widehat{\lambda}^*_n$ selected by minimizing the information criterion defined in (6.1) is consistent in cointegration rank and lag order selection.

Theorem 6.1 *Under Assumption 3.1, 5.1 and 6.1, there is*

$$\Pr\left(\mathcal{S}_{\phi,\widehat{\lambda}^*_n} = \mathcal{S}_{\phi} \text{ and } \mathcal{S}_{B,\widehat{\lambda}^*_n} = \mathcal{S}_{B}\right) \to 1 \tag{6.7}$$

where $\mathcal{S}_{\phi,\widehat{\lambda}^*_n}$ denotes the index set of nonzero eigenvalues of $\widehat{\Pi}_n(\widehat{\lambda}^*_n)$ and $\mathcal{S}_{B,\widehat{\lambda}^*_n}$ denotes the nonzero matrices in $\widehat{B}_n(\widehat{\lambda}^*_n)$.

For empirical implementation, the function $\mathcal{K}(r, p)$ can be specified as $\mathcal{K}(r, p) = r^2 - 2mr - p$ and the sequence C_n can be $\log n$ or $\log\log(n)$, which give BIC and HQIC type information criteria, respectively. On the other hand, if $C_n = o(n)$ and $C_n < M$ for some finite constant M and all n, then we can show that the LS shrinkage estimation based on $\widehat{\lambda}^*_n$ selects underfitted models with probability approaching zero, while it has nontrivial probability of selecting overfitted models. One such example is the AIC type of information criterion where $\mathcal{K}(r, p) = r^2 - 2mr - p$ and $C_n = 2$.

7 Simulation Study

We conduct simulation analysis to assess the finite sample performance of the shrinkage estimates in terms of cointegration rank selection and efficient estimation. Three models are investigated in this section. In the first model, the simulated data are generated from

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \Pi_o \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix},$$
$$\text{where } u_t \equiv \text{i.i.d. } N(0, \Omega_u) \text{ with } \Omega_u = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.75 \end{pmatrix}. \tag{7.1}$$

The initial observation Y_0 is set to be zero for simplicity. Π_o will be specified as

$$\begin{pmatrix} \pi_{11,o} & \pi_{12,o} \\ \pi_{21,o} & \pi_{22,o} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -1 & -0.5 \\ 1 & 0.5 \end{pmatrix} \text{ and } \begin{pmatrix} -0.5 & 0.1 \\ 0.2 & -0.4 \end{pmatrix} \tag{7.2}$$

to allow for the cointegration rank to be 2, 1 and 0 respectively. Y_0 is set to be zero for simplicity.

In the second model, the simulated data $\{Y_t\}_{t=1}^n$ are generated from equation (7.1)-(7.2), while the innovation term u_t is generated by

$$\begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.75 \end{pmatrix} \begin{pmatrix} u_{1,t-1} \\ u_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix},$$
$$\text{where } \varepsilon_t \equiv \text{i.i.d. } N(0, \Omega_\varepsilon) \text{ with } \Omega_\varepsilon = diag(1.25, 0.75).$$

The initial values Y_0 and ε_0 are set to be zero.

The third model has the following form

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \Pi_o \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + B_{1,o} \begin{pmatrix} \Delta Y_{1,t-1} \\ \Delta Y_{2,t-1} \end{pmatrix} + B_{3,o} \begin{pmatrix} \Delta Y_{1,t-3} \\ \Delta Y_{2,t-3} \end{pmatrix} + u_t, \tag{7.3}$$

where u_t is generated under the same condition in (7.1), Π_o is specified similarly in (7.2), $B_{1,o}$ and $B_{2,o}$ is taken to be $\begin{pmatrix} 0.4 & 0 \\ 0 & 0.4 \end{pmatrix}$ such that Assumption 5.1 is satisfied. The initial values (Y_t, ε_t) $(t = -3, ..., 0)$ are set to be zero. In the above three cases, we include 50 additional observations to the simulated sample with sample size n to eliminate start-up effects from the initialization.

.

Table 7.1 Cointegration Rank Selection with Adaptive Lasso Penalty

	Model 1					
	r_o=0, λ_o=(0 0)		r_o=1, λ_o=(0 -0.5)		r_o=2, λ_o=(-0.6 -0.5)	
	$n = 100$	$n = 400$	$n = 100$	$n = 400$	$n = 100$	$n = 400$
$\widehat{r}_n = 0$	**0.8985**	**0.9475**	0.0000	0.0000	0.0000	0.0000
$\widehat{r}_n = 1$	0.1005	0.0525	**0.9875**	**0.9925**	0.1175	0.0000
$\widehat{r}_n = 2$	0.0010	0.0000	0.0125	0.0075	**0.8825**	**1.0000**
	Model 2					
	r_o=0, λ_o=(0 0)		r_o=1, λ_o^*=(0 -0.25)		r_o=2, λ_o^*=(-0.30 -0.15)	
	$n = 100$	$n = 400$	$n = 100$	$n = 400$	$n = 100$	$n = 400$
$\widehat{r}_n = 0$	**0.9445**	**0.9790**	0.0060	0.0000	0.0000	0.0000
$\widehat{r}_n = 1$	0.0550	0.0205	**0.9090**	**0.9870**	0.2515	0.0000
$\widehat{r}_n = 2$	0.0005	0.0005	0.0850	0.0130	**0.7485**	**1.0000**

Table 7.1: Replication=2000. The tuning parameter equals to 2klog(n)/n, where k denotes the number of the total elements which will enter the penalty function. In model 1 and 2 k=2.

In the first two models, we assume that the econometrician specifies the following model

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \Pi_o \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + u_t, \tag{7.4}$$

where u_t is i.i.d.$(0, \Omega_u)$ with some unknown positive definite matrix Ω_u. The above empirical model is correctly specified under the data generating assumption (7.1), but is misspecified under (7.2). We are interested in investigating the performance of the shrinkage method in selecting the correct rank of Π_o under both data generating assumptions and efficient estimation of Π_o under Assumption (7.1).

In the third model, we assume that the econometrician specifies the following model

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \Pi_o \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \sum_{j=1}^{3} B_{j,o} \begin{pmatrix} \Delta Y_{1,t-j} \\ \Delta Y_{2,t-j} \end{pmatrix} + u_t, \tag{7.5}$$

where u_t is i.i.d.$(0, \Omega_u)$ with some unknown positive definite matrix Ω_u. The above empirical model is over-parameterized according to (7.3). We are interested in investigating the performance of the shrinkage method in selecting the correct rank of Π_o and the order of the lagged differences, and efficient estimation of Π_o and $B_{2,o}$.

Table 7.2 Cointegration Rank Selection and Lagged Difference Selection with Adaptive Lasso Penalty

	Cointegration Rank Selection								
	r_o=0, λ_o=(0 0)			r_o=1, λ_o=(0 -0.5)			r_o=2, λ_o=(-0.6 -0.5)		
	$n = 100$	$n = 400$	$n = 600$	$n = 100$	$n = 400$	$n = 600$	$n = 100$	$n = 400$	$n = 600$
$\widehat{r}_n = 0$	**0.8640**	**0.9010**	**0.9100**	0.0015	0.0000	0.0000	0.0260	0.0000	0.0000
$\widehat{r}_n = 1$	0.1345	0.0990	0.0895	**0.9850**	**1.0000**	**1.0000**	0.8380	0.0000	0.0000
$\widehat{r}_n = 2$	0.0015	0.0000	0.0005	0.0135	0.0000	0.0000	**0.1360**	**1.0000**	**1.0000**
	Lagged Difference Selection								
	r_o=0, λ_o=(0 0)			r_o=1, λ_o=(0 -0.5)			r_o=2, λ_o=(-0.6 -0.5)		
	$n = 100$	$n = 400$	$n = 600$	$n = 100$	$n = 400$	$n = 600$	$n = 100$	$n = 400$	$n = 600$
$\widehat{p}_n \in T$	**0.8810**	**0.9605**	**0.9590**	**0.8340**	**0.9595**	**0.9945**	**0.8255**	**0.9845**	**0.9945**
$\widehat{p}_n \in C$	0.1180	0.0395	0.0410	0.1660	0.0405	0.0055	0.0920	0.0155	0.0055
$\widehat{p}_n \in I$	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0825	0.0000	0.0000
	Model Selection								
	r_o=0, λ_o=(0 0)			r_o=1, λ_o=(0 -0.5)			r_o=2, λ_o=(-0.6 -0.5)		
	$n = 100$	$n = 400$	$n = 600$	$n = 100$	$n = 400$	$n = 600$	$n = 100$	$n = 400$	$n = 600$
$\widehat{m}_n \in T$	**0.7590**	**0.8635**	**0.8720**	**0.8225**	**0.9595**	**0.9945**	**0.1095**	**0.9845**	**0.9945**
$\widehat{m}_n \in C$	0.2400	0.1365	0.1280	0.1640	0.0405	0.0055	0.0205	0.0155	0.0055
$\widehat{m}_n \in I$	0.0010	0.0000	0.0000	0.0135	0.0000	0.0000	0.8705	0.0000	0.0000

Table 7.2: Replication=2000 and the tuning parameter equals to 2klog(n)/n, where k=5. "T", "C" and "I" denote the true lags/model, consistent lags/model and inconsistent lags/model selected by the shrinkage estimation.

Table 7.3 Finite Sample Properties of the Shrinkage Estimates

	Model 1 with r_o=0, λ_o=(0.0 0.0) and n=100								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	0.0006	0.0319	0.0319	0.0132	0.0482	0.0515	—	—	—
$B_{o,1}$	0.2301	0.2458	0.3506	0.0451	0.3989	0.4039	0.0255	0.3542	0.3559
$B_{o,2}$	0.0019	0.0246	0.0246	0.0291	0.4200	0.4218	—	—	—
$B_{o,3}$	0.2391	0.2354	0.3519	0.0487	0.3848	0.3902	0.0401	0.3423	0.3464
	Model 1 with r_o=0, λ_o=(0.0 0.0) and n=400								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	0.0005	0.0069	0.0069	0.0037	0.0116	0.0127	—	—	—
$B_{o,1}$	0.0909	0.1214	0.1594	0.0166	0.2171	0.2180	0.0102	0.1956	0.1960
$B_{o,2}$	0.0002	0.0059	0.0059	0.0116	0.2314	0.2319	—	—	—
$B_{o,3}$	0.0916	0.1216	0.1600	0.0177	0.2162	0.2173	0.0151	0.1923	0.1931
	Model 1 with r_o=0, λ_o=(0.0 0.0) and n=600								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	0.0004	0.0043	0.0043	0.0026	0.0077	0.0085	—	—	—
$B_{o,1}$	0.0672	0.1025	0.1279	0.0090	0.1781	0.1786	0.0070	0.1606	0.1609
$B_{o,2}$	0.0000	0.0013	0.0013	0.0103	0.1926	0.1929	—	—	—
$B_{o,3}$	0.0688	0.099	0.1266	0.0113	0.1781	0.1787	0.0100	0.1594	0.1600

Table 7.3: Replication=2000. The tuning parameter equals to 2klog(n)/n, where k denotes the number of the total elements which will enter the penalty function. In model 1 and 2 k=2.

Table 7.4 Finite Sample Properties of the Shrinkage Estimates

	Model 1 with r_o=1, λ_o=(0.0 -0.5) and n=100								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	0.0525	0.3619	0.3657	0.1137	0.3181	0.3822	0.0127	0.3141	0.3145
$B_{o,1}$	0.1686	0.3417	0.3841	0.0293	0.4367	0.4379	0.0286	0.4061	0.4073
$B_{o,2}$	0.0068	0.0384	0.0391	0.0119	0.4122	0.4124	—	—	—
$B_{o,3}$	0.1190	0.1915	0.2290	0.0280	0.3698	0.3712	0.0209	0.2907	0.2917
	Model 1 with r_o=1, λ_o=(0.0 -0.5) and n=400								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	0.0062	0.1938	0.1939	0.1057	0.1838	0.2575	0.0066	0.1820	0.1822
$B_{o,1}$	0.0641	0.1871	0.2001	0.0036	0.2487	0.2487	0.0032	0.2340	0.2340
$B_{o,2}$	0.0007	0.0081	0.0082	0.0017	0.2281	0.2281	—	—	—
$B_{o,3}$	0.0657	0.1138	0.1329	0.0142	0.2139	0.2144	0.0103	0.1712	0.1716
	Model 1 with r_o=1, λ_o=(0.0 -0.5) and n=600								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	0.0060	0.1514	0.1516	0.1012	0.1477	0.2277	0.0024	0.1467	0.1467
$B_{o,1}$	0.0541	0.1484	0.1597	0.0071	0.2013	0.2015	0.0072	0.1896	0.1898
$B_{o,2}$	0.0002	0.0033	0.0033	0.0108	0.1876	0.1879	—	—	—
$B_{o,3}$	0.0506	0.0914	0.1055	0.0050	0.1739	0.1741	0.0048	0.1370	0.1371

Table 7.4: Replication=2000. The tuning parameter equals to 2klog(n)/n, where k denotes the number of the total elements which will enter the penalty function. In model 1 and 2 k=2.

Table 7.5 Finite Sample Properties of the Shrinkage Estimates

	Model 1 with r_o=2, λ_o=(-0.6 -0.5) and n=100								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	1.2314	0.3037	1.2698	0.0225	0.3691	0.3698	0.0205	0.3077	0.3084
$B_{o,1}$	0.7527	0.2762	0.8294	0.0181	0.4044	0.4049	0.0128	0.3701	0.3704
$B_{o,2}$	0.0028	0.0303	0.0304	0.0077	0.3798	0.3799	—	—	—
$B_{o,3}$	0.6224	0.2351	0.6993	0.0264	0.3599	0.3616	0.0223	0.3495	0.3506
	Model 1 with r_o=2, λ_o=(-0.6 -0.5) and n=400								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	0.4078	0.1768	0.4491	0.0034	0.2151	0.2151	0.0067	0.1744	0.1746
$B_{o,1}$	0.3250	0.1560	0.3778	0.0076	0.2278	0.2279	0.0048	0.2079	0.2079
$B_{o,2}$	0.0002	0.0072	0.0082	0.0054	0.2167	0.2168	—	—	—
$B_{o,3}$	0.2986	0.1376	0.3394	0.0115	0.2078	0.2084	0.0092	0.2009	0.2013
	Model 1 with r_o=2, λ_o=(-0.6 -0.5) and n=600								
	Lasso Estimates			OLS			Oracle Estimates		
	Bias	Std	RMSE	Bias	Std	RMSE	Bias	Std	RMSE
Π_o	0.2967	0.1327	0.3292	0.0050	0.1754	0.1756	0.0044	0.1448	0.1449
$B_{o,1}$	0.2414	0.1242	0.2858	0.0063	0.1905	0.1907	0.0034	0.1734	0.1735
$B_{o,2}$	0.0001	0.0022	0.0022	0.0040	0.1777	0.1778	—	—	—
$B_{o,3}$	0.2205	0.1100	0.2562	0.0089	0.1709	0.1712	0.0073	0.1663	0.1665

Table 7.5: Replication=2000. The tuning parameter equals to 2klog(n)/n, where k denotes the number of the total elements which will enter the penalty function. In model 1 and 2 k=2.

8 An Empirical Example

9 Conclusion

One of the main challenges in any applied econometric work is the selection of a good model for practical implementation. The conduct of inference and model use in forecasting and policy analysis are inevitably conditioned on the empirical process of model selection, which typically leads to issues of post-model selection inference. Adaptive lasso and bridge estimation methods provide a methodology where these difficulties may be partly attenuated by simultaneous model selection and estimation to facilitate empirical research in complex models like reduced rank regressions where many selection decisions need to be made to construct a satisfactory empirical model. On the other hand, as indicated in the Introduction, the methods certainly do not eliminate post-shrinkage selection inference issues in finite samples because the estimators carrry the effects of the in-built selections.

This paper shows how to use the methodology of shrinkage in a multivariate system to develop an automated approach to cointegrated system modeling that enables simultaneous estimation of the cointegrating rank and autoregressive order in conjunction with oracle-like efficient estimation of the cointegrating matrix and the transient dynamics. As such the methods offer practical advantages to the empirical researcher by avoiding sequential techniques where cointegrating rank and transient dynamics are estimated prior to model fitting.

Various extensions of the methods developed here are possible. One rather obvious extension is to allow for parametric restrictions on the cointegrating matrix which may relate to theory-induced specifications. Lasso type procedures have so far been confined to parametric models, whereas cointegrated systems are often formulated with some non-parametric elements relating to unknown features of the model. A second extension of the present methodology, therefore, is to semiparametric formulations in which the error process in the VECM is weakly dependent, which is partly considered already in Section 4. The effects of post-shrinkage inference issues also merit detailed investigation. These and other generalizations of the framework will be explored in future work.

10 Appendix

We start with the following standard asymptotic result.

Lemma 10.1 *Under Assumptions 3.1 and 3.2, we have*

(a) $n^{-1}\sum_{t=1}^{n} Z_{1,t-1}Z_{1,t-1}' \to_p \Sigma_{z_1z_1}$*;*

(b) $n^{-\frac{3}{2}}\sum_{t=1}^{n} Z_{1,t-1}Z_{2,t-1}' \to_p 0$*;*

(c) $n^{-2}\sum_{t=1}^{n} Z_{2,t-1}Z_{2,t-1}' \to_d \int B_{w_2}B_{w_2}'$*;*

(d) $n^{-\frac{1}{2}}\sum_{t=1}^{n} u_t Z_{1,t-1}' \to_d N(0, \Omega_u \otimes \Sigma_{z_1z_1})$*;*

(e) $n^{-1}\sum_{t=1}^{n} u_t Z_{2,t-1}' \to_d \left(\int B_{w_2} dB_u'\right)'$*.*

The quantities in (c), (d), and (e) converge jointly.

Proof of Lemma 10.1. See Johansen (1995) and Cheng and Phillips (2009). ■

Proof of Lemma 3.1. (a) From (3.10)

$$
\begin{aligned}
Q\left(\widehat{\Pi}_{ols} - \Pi_o\right) Q^{-1} D_n &= \left(\sum_{t=1}^{n} Q u_t Y_{t-1}' Q'\right)\left(\sum_{t=1}^{n} Q Y_{t-1} Y_{t-1}' Q'\right)^{-1} D_n \\
&= \left(\sum_{t=1}^{n} v_t Z_{t-1}' D_n^{-1}\right)\left(D_n^{-1}\sum_{t=1}^{n} Z_{t-1}Z_{t-1}' D_n^{-1}\right)^{-1}, \quad (10.1)
\end{aligned}
$$

where $v_t = Qu_t$ and $Z_{t-1} = QY_{t-1}$. Result (a) follows directly from Lemma 10.1.

(b) Denote $P = [\beta_o, \beta_{o\perp}]$ and $S_n(\phi) = \phi I_m - \widehat{\Pi}_{ols}$. Then, by definition, the elements of $\phi(\widehat{\Pi}_{ols})$ are the solutions of the determinantal equation,

$$
0 = \left|\phi I_m - \widehat{\Pi}_{ols}\right| = \left|P' S_n(\phi) P\right| = \begin{vmatrix} \phi\beta_o'\beta_o - \beta_o'\widehat{\Pi}_{ols}\beta_o & -\beta_o'\widehat{\Pi}_{ols}\beta_{o\perp} \\ -\beta_{o\perp}'\widehat{\Pi}_{ols}\beta_o & \phi I_{m-r_o} - \beta_{o\perp}'\widehat{\Pi}_{ols}\beta_{o\perp} \end{vmatrix}. \quad (10.2)
$$

Using (a) we deduce that

$$
\begin{aligned}
\beta_o'\widehat{\Pi}_{ols}\beta_{o\perp} &= \beta_o'\left(\widehat{\Pi}_{ols} - \Pi_o\right)\beta_{o\perp} = o_p(1), \quad (10.3) \\
\beta_{o\perp}'\widehat{\Pi}_{ols}\beta_{o\perp} &= \beta_{o\perp}'\left(\widehat{\Pi}_{ols} - \Pi_o\right)\beta_{o\perp} = o_p(1), \quad (10.4)
\end{aligned}
$$

and, similarly,

$$
\beta_{o\perp}'\widehat{\Pi}_{ols}\beta_o \to_p \beta_{o\perp}'\Pi_o\beta_o \text{ and } \beta_o'\widehat{\Pi}_{ols}\beta_o \to_p \beta_o'\Pi_o\beta_o. \quad (10.5)
$$

Using (10.2)-(10.5), we deduce that

$$
\left|\phi I_m - \widehat{\Pi}_{ols}\right| \to_p \left|\phi I_{m-r_o}\right| \times \left|\phi\beta_o'\beta_o - \beta_o'\Pi_o\beta_o\right|, \quad (10.6)
$$

uniformly over any compact set in R^m. By Assumption 3.2.(i), $\phi(\Pi_o) \in \overline{U}_1 := \{\phi \in R^m, ||\phi||_E \leq 1\}$ and $\overline{U}_1$ is a compact set in R^m. Thus, by continuous mapping, we have

$$\phi_{\mathcal{S}^c_{\phi,n}}(\widehat{\Pi}_{ols}) \rightarrow_p 0 \text{ and } \phi_{\mathcal{S}_{\phi,n}}(\widehat{\Pi}_{ols}) \rightarrow_p \phi_{\mathcal{S}_\phi}(\Pi_o), \tag{10.7}$$

where $\phi_{\mathcal{S}_\phi}(\Pi_o)$ denotes the solutions of the equation $\left|\phi\beta_o'\beta_o - \beta_o'\Pi_o\beta_o\right| = 0$. The determinantal equation $\left|\phi\beta_o'\beta_o - \beta_o'\Pi_o\beta_o\right| = 0$ is equivalent to $\left|\phi I_{r_o} - \beta_o'\alpha_o\right| = 0$, so result (b) follows.

(c) Using the notation from (b), we have

$$|S_n(\phi)| = \left|\beta_o'S_n(\phi)\beta_o\right| \times \left|\beta_{o\perp}'\left\{S_n(\phi) - S_n(\phi)\beta_o\left[\beta_o'S_n(\phi)\beta_o\right]^{-1}\beta_o'S_n(\phi)\right\}\beta_{o\perp}\right|. \tag{10.8}$$

Let $\mu_k^* = n\phi_k(\widehat{\Pi}_{ols})$ ($k \in \mathcal{S}^c_{n,\phi}$), so that μ_k^* is by definition a solution of the equation

$$0 = \left|\beta_o'S_n(\mu)\beta_o\right| \times \left|\beta_{o\perp}'\left\{S_n(\mu) - S_n(\mu)\beta_o\left[\beta_o'S_n(\mu)\beta_o\right]^{-1}\beta_o'S_n(\mu)\right\}\beta_{o\perp}\right|, \tag{10.9}$$

where $S_n(\mu) = \frac{\mu}{n}I_m - \widehat{\Pi}_{ols}$.

For any compact subset $K \subset R$, we can invoke the results in (a) to show

$$\beta_o'S_n(\mu)\beta_o = \frac{\mu}{n}\beta_o'\beta_o - \beta_o'\left(\widehat{\Pi}_{ols} - \Pi_o\right)\beta_o + \beta_o'\Pi_o\beta_o \rightarrow_p \beta_o'\Pi_o\beta_o, \tag{10.10}$$

uniformly over K. From Assumption 3.2.(iii), we have

$$\left|\beta_o'\Pi_o\beta_o\right| = \left|\beta_o'\alpha_o\beta_o'\beta_o\right| = \left|\beta_o'\alpha_o\right| \times \left|\beta_o'\beta_o\right| \neq 0.$$

Thus, the normalized $m - r_o$ smallest eigenvalues μ_k^* ($k \in \mathcal{S}^c_{n,\phi}$) of $\widehat{\Pi}_{ols}$ are asymptotically the solutions of the following determinantal equation,

$$0 = \left|\beta_{o\perp}'\left\{S_n(\mu) - S_n(\mu)\beta_o\left[\beta_o'S_n(\mu)\beta_o\right]^{-1}\beta_o'S_n(\mu)\right\}\beta_{o\perp}\right|, \tag{10.11}$$

where

$$\beta_o'S_n(\mu)\beta_{o\perp} = \beta_o'\left(\widehat{\Pi}_{ols} - \Pi_o\right)\beta_{o\perp}, \tag{10.12}$$

$$\beta_{o\perp}'S_n(\mu)\beta_{o\perp} = \frac{\mu}{n}I_{m-r_o} - \beta_{o\perp}'\left(\widehat{\Pi}_{ols} - \Pi_o\right)\beta_{o\perp}, \tag{10.13}$$

$$\beta_{o\perp}'S_n(\mu)\beta_o = \beta_{o\perp}'\widehat{\Pi}_{ols}\beta_o \rightarrow_p \beta_{o\perp}'\alpha_o\beta_o'\beta_o. \tag{10.14}$$

Using the results in (10.10) and (10.12)-(10.14), we get

$$\begin{aligned} & \beta_{o\perp}'\left\{S_n(\mu)-S_n(\mu)\beta_o\left[\beta_o' S_n(\mu)\beta_o\right]^{-1}\beta_o' S_n(\mu)\right\}\beta_{o\perp} \\ = & \frac{\mu}{n}I_{m-r_o}-\beta_{o\perp}'\left[I_m-\alpha_o\left(\beta_o'\alpha_o\right)^{-1}\beta_o'+o_p(1)\right]\left(\widehat{\Pi}_{ols}-\Pi_o\right)\beta_{o\perp}. \end{aligned} \tag{10.15}$$

Note that

$$\beta_{o\perp}'\left[I_{m-r_o}-\alpha_o\left(\beta_o'\alpha_o\right)^{-1}\beta_o'\right]Q^{-1}=\left[\mathbf{0}_{(m-r_o)\times r_o},(\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\right]:=H_1, \tag{10.16}$$

and

$$Q\beta_{o\perp}=\left[\beta_o,\alpha_{o\perp}\right]'\beta_{o\perp}=\left[\mathbf{0}_{(m-r_o)\times r_o},\beta_{o\perp}'\alpha_{o\perp}\right]':=H_2'. \tag{10.17}$$

Using (10.1), we deduce that

$$\begin{aligned} & nH_1Q\left(\widehat{\Pi}_{ols}-\Pi_o\right)Q^{-1}H_2' \\ = & \left(H_1\sum_{t=1}^n w_t Z_{t-1}'D_n^{-1}\right)\left(D_n^{-1}\sum_{t=1}^n Z_t Z_t' D_n^{-1}\right)^{-1}H_2' \\ \to_d & (\alpha_{o,\perp}'\beta_{o,\perp})^{-1}\left(\int B_{w_2}dB_{w_2}'\right)'\left(\int B_{w_2}B_{w_2}'\right)^{-1}(\alpha_{o,\perp}'\beta_{o,\perp}). \end{aligned} \tag{10.18}$$

Then, from (10.11)-(10.18), we obtain

$$\begin{aligned} & \left|n\beta_{o\perp}'\left\{S_n(\mu)-S_n(\mu)\beta_o\left[\beta_o' S_n(\mu)\beta_o\right]^{-1}\beta_o' S_n(\mu)\right\}\beta_{o\perp}\right| \\ \to_d & \left|\mu I_{m-r_o}-\left(\int B_{w_2}dB_{w_2}'\right)'\left(\int B_{w_2}B_{w_2}'\right)^{-1}\right|, \end{aligned} \tag{10.19}$$

uniformly over K. The result in (c) follows from (10.19) and by continuous mapping. ■

Proof of Theorem 3.1. Define

$$V_{1,n}(\Pi)=\sum_{t=1}^n\|\Delta Y_t-\Pi Y_{t-1}\|_E^2+n\sum_{k=1}^m\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi)\right]=V_{2,n}(\Pi)+n\sum_{k=1}^m\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi)\right].$$

We can write

$$\sum_{t=1}^n\|\Delta Y_t-\Pi Y_{t-1}\|_E^2=\left[\Delta y-\left(Y_{-1}'\otimes I_m\right)vec(\Pi)\right]'\left[\Delta y-\left(Y_{-1}'\otimes I_m\right)vec(\Pi)\right]$$

where $\Delta y = vec\,(\Delta Y)$, $\Delta Y = (\Delta Y_1, ..., \Delta Y_n)_{m\times n}$ and $Y_{-1} = (Y_0, ..., Y_{T-1})_{m\times n}$.

By definition, $V_{1,n}(\widehat{\Pi}_n) \leq V_{1,n}(\Pi_o)$ and thus

$$\begin{aligned} & \left[vec(\widehat{\Pi}_n - \Pi_o)\right]' \left(\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' \otimes I_m\right) \left[vec(\widehat{\Pi}_n - \Pi_o)\right] \\ & +2\left[vec(\widehat{\Pi}_n - \Pi_o)\right]' vec\left(\sum_{t=1}^{n} Y_{t-1}u_t'\right) \\ \leq \;\; & n\left\{\sum_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right] - \sum_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\widehat{\Pi}_n)\right]\right\}. \end{aligned} \tag{10.20}$$

When $r_o = 0$, ΔY_t is stationary and Y_t is full rank $I\,(1)$, so that

$$n^{-2}\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' \to_d \int_0^1 B_u(a)B_u'(a)da, \tag{10.21}$$

and $n^{-2}\sum_{t=1}^{T} Y_{t-1}u_t' = o_p(1)$. From the results in (10.20) and (10.21), we get w.p.a.1,

$$\left\|\widehat{\Pi}_n - \Pi_o\right\|_E^2 \mu_{\min} - 2\left\|\widehat{\Pi}_n - \Pi_o\right\|_E c_n - d_n \leq 0, \tag{10.22}$$

where $\mu_{\min}$ denotes the smallest eigenvalue of the random matrix $\int_0^1 B_u(a)B_u'(a)da$, which is positive with probability 1, $c_n = \left\|n^{-2}\sum_{t=1}^{n} Y_{t-1}u_t'\right\|_E = o_p(1)$ and $d_n = n^{-1}\sum_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right] = o_p(1)$. From (10.22), it is straightforward to deduce that $\left\|\widehat{\Pi}_n - \Pi_o\right\|_E = o_p(1)$.

When $r_o = m$, Y_t is stationary and we have

$$n^{-1}\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' \to_p \Gamma_{YY}(0) = R(1)\Omega_u R(1)', \tag{10.23}$$

and $n^{-1}\sum_{t=1}^{T} Y_{t-1}u_t' = o_p(1)$. From the results in (10.20) and (10.21), we get w.p.a.1,

$$\left\|\widehat{\Pi}_n - \Pi_o\right\|_E^2 \mu_{\min} - 2\left\|\widehat{\Pi}_n - \Pi_o\right\|_E c_n - d_n \leq 0 \tag{10.24}$$

where $\mu_{\min}$ denotes the smallest eigenvalue of $\Gamma_{YY}(0)$, which is positive with probability 1, $c_n = \left\|n^{-1}\sum_{t=1}^{n} Y_{t-1}u_t'\right\|_E = o_p(1)$ and $d_n = \sum_{k=1}^{m} \widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right] = o_p(1)$. So, consistency of $\widehat{\Pi}_n$ follows directly from the inequality in (10.24).

Denote $B_n = (D_n Q)^{-1}$, then when $0 < r_o < m$, we can use the results in Lemma 1 to deduce that

$$\begin{aligned}\sum_{t=1}^{n} Y_{t-1}Y_{t-1}' &= Q^{-1}D_n^{-1}D_n\sum_{t=1}^{n} Z_{t-1}Z_{t-1}'D_nD_n^{-1}Q'^{-1} \\ &= B_n\left[\begin{pmatrix} \Sigma_{z_1z_1} & 0 \\ 0 & \int B_{w_2}B_{w_2}' \end{pmatrix} + o_p(1)\right]B_n',\end{aligned}$$

and thus

$$vec(\widehat{\Pi}_n - \Pi_o)'\left(B_nD_n\sum_{t=1}^{n} Z_{t-1}Z_{t-1}'D_nB_n' \otimes I_m\right)vec(\widehat{\Pi}_n - \Pi_o) \geq \mu_{\min}\left\|\left(\widehat{\Pi}_n - \Pi_o\right)B_n\right\|_E^2, \tag{10.25}$$

where $\mu_{\min}$ is the smallest eigenvalue of $diag\left(\Sigma_{z_1z_1}, \int B_{w_2}B_{w_2}'\right)$ and is strictly positive with probability 1. Next observe that

$$\left|\left[vec(\widehat{\Pi}_n - \Pi_o)\right]' vec\left(B_nD_n\sum_{t=1}^{n} Z_{t-1}u_t'\right)\right| \leq \left\|\left(\widehat{\Pi}_n - \Pi_o\right)B_n\right\|_E e_n, \tag{10.26}$$

where $e_n = \|D_n\sum_{t=1}^{n} Y_{t-1}u_t'\|_E = O_p(1)$ by Lemma 10.1. Denoting $d_n = n\sum_{k=1}^{m}\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]$, we have the inequality

$$\mu_{\min}\left\|\left(\widehat{\Pi}_n - \Pi_o\right)B_n\right\|_E^2 - 2\left\|\left(\widehat{\Pi}_n - \Pi_o\right)B_n\right\|_E e_n - d_n \leq 0, \tag{10.27}$$

which implies

$$\left(\widehat{\Pi}_n - \Pi_o\right)B_n = O_p(1 + d_n^{\frac{1}{2}}). \tag{10.28}$$

By the definition of B_n and (10.28), we deduce that

$$\widehat{\Pi}_n - \Pi_o = O_p(n^{-\frac{1}{2}} + n^{-\frac{1}{2}}d_n^{\frac{1}{2}}) = o_p(1),$$

which implies the consistency of $\widehat{\Pi}_n$. ■

Proof of Theorem 3.2. By the consistency of $\widehat{\Pi}_n$ and the CMT $\phi_k(\widehat{\Pi}_n)$ is consistent for $\phi_k(\Pi_o)$. Consistency of $\phi_k(\widehat{\Pi}_n)$ and continuous twice differentiability of $\widehat{P}_{\lambda_n}(\cdot)$ imply that

$$\begin{aligned}\left|\sum_{k\in\mathcal{S}_\phi}\left(\widehat{P}_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]-\widehat{P}_{\lambda_n}[\bar{\phi}_k(\widehat{\Pi}_n)]\right)\right| &\leq \max_{k\in\mathcal{S}_\phi}\left|\widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right|\sum_{k\in\mathcal{S}_\phi}\left|\bar{\phi}_k(\widehat{\Pi}_n)-\bar{\phi}_k(\Pi_o)\right| \\ &\quad+\max_{k\in\mathcal{S}_\phi}\left|\widehat{P}''_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]+o_p(1)\right|\sum_{k\in\mathcal{S}_\phi}\left|\bar{\phi}_k(\widehat{\Pi}_n)-\bar{\phi}_k(\Pi_o)\right|^2 \quad (10.29)\end{aligned}$$

By the triangle inequality and Bauer-Fiker Theorem on eigenvalue sensitivity we have the inequality

$$\left|\bar{\phi}_k(\widehat{\Pi}_n)-\bar{\phi}_k(\Pi_o)\right|\leq\left|\phi_k(\widehat{\Pi}_n)-\phi_k(\Pi_o)\right|\leq C_V\left\|\widehat{\Pi}_n-\Pi_o\right\|_E, \quad (10.30)$$

for all $k\in S_\phi$, where the constant $C_V=\|V_o\|_E/\left\|V_o^{-1}\right\|_E\in(0,\infty)$ and V_o is the eigenvector matrix of Π_o.

Using (10.29) and (10.30) and invoking the inequality in (10.20) we get

$$\begin{aligned}&\left[vec(\widehat{\Pi}_n-\Pi_o)\right]'\left(\sum_{t=1}^n Y_{t-1}Y'_{t-1}\otimes I_m\right)\left[vec(\widehat{\Pi}_n-\Pi_o)\right] \\ &+2\left[vec(\widehat{\Pi}_n-\Pi_o)\right]'vec\left(\sum_{t=1}^n Y_{t-1}u'_t\right) \\ \leq\ & Cn\max_{k\in\mathcal{S}_\phi}\left|\widehat{P}'_{\lambda_n}\left[\bar{\phi}_k(\Pi_o)\right]\right|\left\|\widehat{\Pi}_n-\Pi_o\right\|_E+o_p(1)\left\|\widehat{\Pi}_n-\Pi_o\right\|_E^2, \quad (10.31)\end{aligned}$$

where $C>0$ denotes a generic constant.

When $r_o=0$, we use (10.21) and (10.31) to deduce that

$$\left\|\widehat{\Pi}_n-\Pi_o\right\|_E^2\mu_{\min}-\left\|\widehat{\Pi}_n-\Pi_o\right\|_E\left(c_n+n^{-1}a_nC\right)\leq 0, \quad (10.32)$$

where $c_n=\left\|n^{-2}\sum_{t=1}^n Y_{t-1}u'_t\right\|_E=O_p(n^{-1})$. We deduce from the inequality (10.32) that

$$\widehat{\Pi}_n-\Pi_o=O_p(n^{-1}+n^{-1}a_n). \quad (10.33)$$

When $r_o=m$, we use (10.23) and (10.31) to deduce that

$$\left\|\widehat{\Pi}_n-\Pi_o\right\|_E^2\mu_{\min}-\left\|\widehat{\Pi}_n-\Pi_o\right\|_E(c_n+a_nC)\leq 0, \quad (10.34)$$

where $c_n = \left\| \frac{1}{n} \sum_{t=1}^{n} Y_{t-1}u_t' \right\|_E = O_p(n^{-\frac{1}{2}})$. The inequality (10.32) leads to

$$\widehat{\Pi}_n - \Pi_o = O_p(n^{-\frac{1}{2}} + a_n). \tag{10.35}$$

When $0 < r_o < m$, we can use the results in (10.25), (10.26) and (10.31) to deduce that

$$\mu_{\min} \left\| \left(\widehat{\Pi}_n - \Pi_o \right) B_n \right\|_E^2 - 2 \left\| \left(\widehat{\Pi}_n - \Pi_o \right) B_n \right\|_E e_n \leq n a_n C_V \left\| \left(\widehat{\Pi}_n - \Pi_o \right) B_n B_n^{-1} \right\|_E, \tag{10.36}$$

where $e_n = \| D_n Q \sum_{t=1}^{n} Y_{t-1} u_t' \|_E = O_p(1)$. Note that

$$\left\| \left(\widehat{\Pi}_n - \Pi_o \right) B_n B_n^{-1} \right\|_E \leq C n^{-\frac{1}{2}} \left\| \left(\widehat{\Pi}_n - \Pi_o \right) B_n \right\|_E. \tag{10.37}$$

Using (10.36) and (10.37), we get

$$\left(\widehat{\Pi}_n - \Pi_o \right) B_n = O_p(1 + n^{\frac{1}{2}} a_n) \tag{10.38}$$

which finishes the proof. ■

References

[1] M. Caner and K. Knight, "No country for old unit root tests: bridge estimators differentiate between nonstationary versus stationary models and select optimal lag," *unpublished manuscript*, 2009.

[2] J. Chao and P.C.B. Phillips, "Model selection in partially nonstationary vector autoregressive processes with reduced rank structure," *Journal of Econometrics*, vol. 91, no. 2, pp. 227.271, 1999.

[3] X. Cheng and P.C.B. Phillips, "Semiparametric cointegrating rank selection," *Econometrics Journal*, vol. 12, pp. S83.S104, 2009.

[4] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348.1360, 2001.

[5] S. Johansen, "Statistical analysis of cointegration vectors," Journal of economic dynamics and control, vol. 12, no. 2-3, pp. 231.254, 1988.

[6] S. Johansen, *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press, USA, 1995.

[7] K. Knight and W. Fu, "Asymptotics for lasso-type estimators," *Annals of Statistics*, vol. 28, no. 5, pp. 1356.1378, 2000.

[8] H. Leeb and B. Potscher, "Model selection and inference: facts and fiction," *Econometric Theory*, vol. 21, no. 01, pp. 21.59, 2005.

[9] H. Leeb and B. M. Potscher, "Sparse estimators and the oracle property, or the return of the Hodges estimator", *Journal of Econometrics*, vol. 142, no. 1, pp. 201-211, 2008.

[10] Z. Liao, "Adaptive GMM shrinkage estimation with consistent moment selection," *unpublished manuscript*, 2010.

[11] Z. Liao and P.C.B. Phillips, "Reduced rank regression of partially non-stationary vector autoregressive processes under misspecification," *unpublished manuscript*, 2010.

[12] P.C.B. Phillips, "Optimal inference in cointegrated systems," *Econometrica*, vol. 59, no. 2, pp. 283.306, 1991.

[13] ———, "Fully modified least squares and vector autoregression," *Econometrica*, vol. 63, no. 5, pp. 1023.1078, 1995.

[14] ———, "Econometric model determination," *Econometrica*, vol. 64, no. 4, pp. 763.812, 1996.

[15] P.C.B. Phillips and V. Solo, "Asymptotics for linear processes," *Annals of Statistics*, vol. 20, no. 2, pp. 971.1001, 1992.

[16] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418.1429, 2006.