Reduced Rank Regression using Generalized Method of Moments Estimators

with extensions to structural breaks in cointegration models*

Econometric Institute Reports, no. 9735/A

Frank Kleibergen†

June 2, 1997

Abstract

Generalized Method of Moments (GMM) Estimators are derived for Reduced Rank Regression Models, the Error Correction Cointegration Model (ECCM) and the Incomplete Simultaneous Equations Model (INSEM). The GMM (2SLS) estimators of the cointegrating vector in the ECCM are shown to have normal limiting distributions. Tests for the number of unit roots can be constructed straightforwardly and have Dickey-Fuller type limiting distributions. Two extensions of the ECCM, which are important in practice, are analyzed. First, cointegration estimators and tests allowing for structural shifts in the variance (heteroscedasticity) of the series are derived and analyzed using a Generalized Least Squares Estimator. Second, cointegrating vector estimators and tests are derived which allow for structural breaks in the cointegrating vector and/or multiplicator. The resulting cointegrating vectors estimators have again normal limiting distributions while the cointegration tests have limiting distributions which differ from the standard Dickey-Fuller type.

---

*Forthcoming in *Generalized Method of Moments Estimation*, eds. L. Mátyás, Kluwer Publishing Company.

†Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, Email:kleibergen@few.eur.nl. I thank Bas Werker for helpful comments and discussion. Part of this research was carried out while the author was affiliated with Tilburg University. The author thanks the Netherlands Organisation for Scientific Research (N.W.O.) for their financial support. Paper also available at http://www.eur.nl/few/ei/faculty/kleibergen.

1 Introduction

Cointegration has been an important research topic since its definition in [1] and already a large literature has evolved on it. An important part of this literature is devoted to the construction of estimators, test statistics and their limiting distributions, see e.g. [1], [4] and [12]. These contributions cover stylised models, which are constant over time and have a constant variance. Although models, which deviate from these stylised models, no longer suffice the condition of weak (covariance) stationarity, they can still show mean reversion so that they still possess properties of cointegration, see for example [7], where it is shown that cointegration can still be defined in periodic models although the model for the cointegrating relationships is not weakly stationary but still mean reverting. So, the cointegrating relationships do not suffice weak stationarity conditions in these cases but cointegration is still an important property of the series generated by these kind of models. In practice, there is a need for the construction of cointegration estimators and test statistics which can be applied in these kind of models as a large number of series possess properties resulting from these models, like heteroscedasticity and structural breaks, and still show mean reversion of linear combinations. Examples lie in areas like finance, where heteroscedasticity is a stylised fact, and macro-economics, where structural breaks are an important topic. Application of the cointegration estimators, which essentially assume that these properties are not present, to these kind of series can lead to inconsistent estimators and/or wrong expressions of the (asymptotic) variances of the estimators. There is, therefore, a need for the development of cointegration estimators and test statistics which can be applied in these kind of models. We develop a Generalised Method of Moments (GMM) framework, see [8], for cointegration models which allows for the incorporation of for example heteroscedasticity and/or structural breaks. Also the stylised models are covered by this framework and lead to estimators which are the 2SLS (two stage least squares) counterpart of the canonical correlation cointegration estimators, see [4].

The paper is organised as follows. In section 2, the relation between the GMM-2SLS estimators in cointegration and simultaneous equations models is discussed jointly with the limiting distributions of the cointegrating vector estimators for a few widely used specifications of the deterministic components. Section 3, shows the limiting distribution of the GMM objective function which can be used to test for the number of unit roots/cointegrating relationships. In section 4, the stylised model is extended to a model where a shift of variance occurs after a predefined fraction of time has evolved. A Generalised Least Squares approach, which assumes a priori knowledge of the variance shift moment, is used to construct the cointegration estimators and statistics that allow for heteroscedasticity. In section 5, cointegration estimators and statistics that account for a change in the cointegrating relationship and/or multiplicator, are constructed. Both extensions can be further generalised to more shifts and also other moment conditions can be added. Finally, the sixth section concludes.

Note that the following definitions are used throughout the paper; $\Rightarrow$ indicates weak convergence; integrals are taken over the unit interval unless indicated otherwise; when possible without confusion, integrals like $\int W(t)dt$ are shortly denoted as $\int W$. The theorems in the paper are derived assuming Gaussian disturbances, which assumption can be relaxed, see for example [15].

2 GMM-2SLS Estimators in Reduced Rank Regression Models

2.1 Reduced Rank Regression Models

Reduced rank regression models are characterized by the lower column or row rank of a parameter matrix. Two well known models which possess this property are the Error Correction Cointegration Model (ECCM) and the INcomplete Simultaneous Equations Model (INSEM). The ECCM is specified as

$$\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t, \tag{1}$$

where $x_t : k \times 1$, $t = 1, ..., T$; $\alpha, \beta : k \times r$; $\beta' = (I_r \ -\beta_2')$; and ε_t is Gaussian white noise with covariance matrix Σ. For simplicity higher order lags are left out. The INSEM reads

$$\begin{aligned} y_{1t} &= \beta_2' y_{2t} + \gamma_1' x_{1t} + \varepsilon_{1t} \\ y_{2t} &= \Pi_{21} x_{1t} + \Pi_{22} x_{2t} + \varepsilon_{2t} \end{aligned} \tag{2}$$

where $y_{1t} : m_1 \times 1$, $y_{2t} : m_2 \times 1$, $x_{1t} : k_1 \times 1$, $x_{2t} : k_2 \times 1$, $t = 1, ..., T$; $\beta_2 : m_2 \times m_1$; $\gamma_1 : k_1 \times m_1$; $\Pi_{21} : m_2 \times k_1$; $\Pi_{22} : m_2 \times k_2$. The disturbances ε_{1t} and ε_{2t} are assumed to be Gaussian white noise with covariance matrix Ω. The variables x_{1t} and x_{2t} are assumed to be (weakly) exogenous. The INSEM in (2) is identified when the number of excluded exogenous variables from the first set of equations, k_2, is at least as large as the number of equations in the second set, m_2, $k_2 \geq m_2$.

The reduced rank property of both of these models is obtained when we specify them as restrictions of the standard linear model,

$$y_t = \Pi z_t + u_t. \tag{3}$$

Both the ECCM and the INSEM are restricted versions of the model in (3). The ECCM is obtained by specifying $y_t = \Delta x_t$, $\Pi = \alpha\beta' = \begin{pmatrix} \alpha_{11} & -\alpha_{11}\beta_2' \\ \alpha_{21} & -\alpha_{21}\beta_2' \end{pmatrix}$, $u_t = \varepsilon_t$, $z_t = x_{t-1}$, while the INSEM is obtained when we substitute the equation of y_{2t} in the equation of y_{1t} which then results in (3) with $y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}$, $z_t =$

$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}$, $u_t = \begin{pmatrix} \varepsilon_{1t} - \beta_2' \varepsilon_{2t} \\ \varepsilon_{2t} \end{pmatrix}$, $\Pi = \begin{pmatrix} \beta_2' \Pi_{21} - \gamma_1 & \beta_2' \Pi_{22} \\ \Pi_{21} & \Pi_{22} \end{pmatrix}$. The reduced rank structure of the ECCM is obvious while the INSEM has a reduced rank structure when $\gamma_1 = 0$ since the first set of rows of Π is a linear function of the other rows in that case. The reduced rank properties of both models are different in nature, however, as in the ECCM the last set of columns is a linear combination of the first set while in the INSEM the first set of rows is a linear combination of the last set.

2.2 GMM-2SLS estimators

In the INSEM form (2), a consistent estimator of the parameters β_2 and γ_1 is obtained when we replace y_{2t} in the first set of equations by $\hat{y}_{2t} = \hat{\Pi}_{21} x_{1t} + \hat{\Pi}_{22} x_{2t}$, where $\hat{\Pi}_{21}$ and $\hat{\Pi}_{22}$ are the least squares estimators obtained from the second set of equations, and estimate the parameters of the resulting equation using least squares. The resulting estimator of (β_2, γ_1) is known as the two stage least squares (2SLS) estimator. A similar kind of 2SLS estimator can be constructed for the cointegrating vector β in the ECCM (1). An important difference between the cointegrating vector parameter β and the structural form parameters β_2 and γ_1 results, however, from the presence of the cointegrating vector in all equations of the ECCM while the structural form parameters of the INSEM only appear in the first set of equations. The 2SLS estimator for the ECCM has, therefore, a different specification than the 2SLS estimator for the INSEM. Both of these estimators are Generalized Method of Moments (GMM) estimators, see [3].

To derive the expressions of the GMM-2SLS estimators both in the INSEM and the ECCM, we use the first order conditions for a maximum of the likelihood. The derivatives of the log likelihood, when assuming Gaussian white noise disturbances with covariance matrix Σ, of the model in (3), read

$$\begin{aligned} \frac{\partial \ln l(\theta)}{\partial \theta'} &= vec(\Sigma^{-1})' \sum_{t=1}^{T} (u_t \otimes I_k) \frac{\partial u_t}{\partial \theta'} \qquad (4) \\ &= \sum_{t=1}^{T} vec(u_t x_t')' (I_k \otimes \Sigma^{-1}) \frac{\partial vec(\Pi)}{\partial \theta'}. \end{aligned}$$

In the GMM objective function we only use the $\sum_{t=1}^{T} vec(u_t x_t')$ part of the derivatives in (4). When we substitute the expression of θ in $\frac{\partial vec(\Pi)}{\partial \theta'}$, the first order derivatives of the different parameters are obtained. The resulting expressions read, for the ECCM,

$$\begin{aligned} \frac{\partial vec(\Pi)}{\partial vec(\beta')'} &= -(I_k \otimes \alpha), \qquad (5) \\ \frac{\partial vec(\Pi)}{\partial vec(\alpha)'} &= -(\beta \otimes I_k), \end{aligned}$$

and for the INSEM,

$$\begin{aligned}
\frac{\partial vec(\Pi)}{\partial vec(\Phi_2')'} &= -(\begin{pmatrix} \Pi_{21} & \Pi_{22} \end{pmatrix}' \otimes \begin{pmatrix} I_{m_1} \\ 0 \end{pmatrix}), \\
\frac{\partial vec(\Pi)}{\partial vec(\gamma_1')'} &= -(\begin{pmatrix} I_{k_1} & 0 \end{pmatrix}' \otimes \begin{pmatrix} I_{m_1} \\ 0 \end{pmatrix}), \\
\frac{\partial vec(\Pi)}{\partial vec(\Pi_{21})'} &= -(\begin{pmatrix} I_{k_1} & 0 \end{pmatrix}' \otimes \begin{pmatrix} \Phi_2' \\ I_{m_2} \end{pmatrix}), \\
\frac{\partial vec(\Pi)}{\partial vec(\Pi_{22})'} &= -(\begin{pmatrix} 0 & I_{k_2} \end{pmatrix}' \otimes \begin{pmatrix} \Phi_2' \\ I_{m_2} \end{pmatrix}).
\end{aligned} \tag{6}$$

The expressions of the derivatives of the individual parameters are substituted in the first order derivative of the objective function which is minimized in the GMM framework. As we cannot exactly solve the normal equation, $\sum_{t=1}^T u_t x_t' = 0$, in case of reduced rank parameter matrices, we take a quadratic form containing these normal equations as objective function to be minimized in the GMM framework, see also [2],

$$Q(\theta) = vec(\sum_{t=1}^T u_t x_t')'((\sum_{t=1}^T x_t x_t')^{-1} \otimes \Sigma^{-1})vec(\sum_{t=1}^T u_t x_t'). \tag{7}$$

The first order condition of the GMM objective function then becomes

$$\begin{aligned}
\frac{\partial Q(\theta)}{\partial \theta'} &= 0 \Leftrightarrow \\
\sum_{t=1}^T (\frac{\partial u_t}{\partial \theta'})(x_t' \otimes I_k)((\sum_{t=1}^T x_t x_t')^{-1} \otimes \Sigma^{-1})vec(\sum_{t=1}^T u_t x_t') &= 0 \Leftrightarrow \\
(\frac{\partial vec(\Pi)}{\partial \theta'})' vec(\Sigma^{-1} \sum_{t=1}^T u_t x_t') &= 0
\end{aligned} \tag{8}$$

The first order condition of the GMM objective function in (8) equals the first order condition for a maximum likelihood value, see (4).

For the parameters of the ECM these first order conditions read,

$$\begin{aligned}
(I_k \otimes \alpha)' vec(\Sigma^{-1} \sum_{t=1}^T u_t z_{t-1}') &= 0 \Leftrightarrow \\
(\sum_{t=1}^T z_{t-1} z_{t-1}')^{-1} (\sum_{t=1}^T z_{t-1} \Delta y_t') \Sigma^{-1} \alpha (\alpha' \Sigma^{-1} \alpha)^{-1} &= \beta,
\end{aligned} \tag{9}$$

$$\begin{aligned}
(\beta \otimes I_k)' vec(\Sigma^{-1} \sum_{t=1}^T u_t z_{t-1}') &= 0 \Leftrightarrow \\
(\sum_{t=1}^T \Delta y_t z_{t-1}' \beta)(\beta' \sum_{t=1}^T z_{t-1} z_{t-1}' \beta)^{-1} &= \alpha,
\end{aligned} \tag{10}$$

and for the parameters of the INSEM these first order conditions read,

$$((\Pi_{21}\ \ \Pi_{22})' \otimes \begin{pmatrix} I_{m_1} \\ 0 \end{pmatrix})' vec(\Sigma^{-1}\sum_{t=1}^{T} u_t x_t') = 0 \Leftrightarrow \quad (11)$$

$$((\Pi_{21}\ \ \Pi_{22})\sum_{t=1}^{T} x_t y_{2t}')^{-1}(\Pi_{21}\ \ \Pi_{22})\sum_{t=1}^{T} x_t(y_{1t} - \gamma_1' x_{1t})' = \beta_2,$$

$$((I_{k_1}\ \ 0)' \otimes \begin{pmatrix} I_{m_1} \\ 0 \end{pmatrix})' vec(\Sigma^{-1}\sum_{t=1}^{T} u_t x_t') = 0 \Leftrightarrow \quad (12)$$

$$(\sum_{t=1}^{T} x_{1t}x_{1t}')^{-1}\sum_{t=1}^{T} x_{1t}(y_{1t} - \beta_2' y_{2t})' = \gamma_1,$$

$$(I_k \otimes \begin{pmatrix} \beta_2' \\ I_{m_2} \end{pmatrix})' vec(\Sigma^{-1}\sum_{t=1}^{T} u_t x_t') = 0 \Leftrightarrow \quad (13)$$

$$(\sum_{t=1}^{T} y_{2t}x_t')(\sum_{t=1}^{T} x_t x_t')^{-1} = \Pi_2.$$

The normal equations for the INSEM directly lead to the 2SLS estimator as the estimator of Π_2 is independent of the parameters β_2 and γ_1 such that it can be estimated independently. The resulting estimator of Π_2 is then used to construct estimators for β_2 and γ_1 (2SLS estimators). The estimators of α and β in the ECM both depend on one another. As we did not restrict α and β, they are also not identified. If we specify β as, $\beta = (I_r\ \ -\beta_2')'$, both α and β_2 are properly identified. If this specification of β is used, a consistent estimator of α is,

$$\hat{\alpha} = (\sum_{t=1}^{T} \Delta x_t(1 - x_{2t-1}'(\sum_{t=1}^{T} x_{2t-1}x_{2t-1}')^{-1}x_{2t-1})x_{1t-1}') \quad (14)$$
$$(\sum_{t=1}^{T} x_{1t-1}(1 - x_{2t-1}'(\sum_{t=1}^{T} x_{2t-1}x_{2t-1}')^{-1}x_{2t-1})x_{1t-1}')^{-1}$$

where $x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}$, $x_{1t} : r \times 1$, $x_{2t} : (k-r) \times 1$, which is just the first r columns of the least squares estimator of Π in (5).

If the estimator of α (14) is used in the estimation of the cointegrating vector β, (8), the identifying restrictions on β are automatically fulfilled. The resulting estimator of β is then the 2SLS estimator of the cointegrating vector β. In a Bayesian analysis this 2SLS estimator equals the mean of the conditional posterior of β given α when a diffuse prior is used, see [6]. The estimators of α and β in (8) and (10) also allow for the construction of an iterative estimation scheme for which the resulting estimators converge to the maximum likelihood estimators.

Asymptotically the 2SLS least squares cointegrating vector estimator possesses the same kind of properties as the maximum likelihood estimator, i.e. superconsistency and normal limiting distribution. This is proved in the theorems in the following (sub)sections.

2.3 Limiting distributions ECM-2SLS cointegration estimators

As the limiting distribution of the 2SLS estimator in the INSEM model is discussed at length in the literature, see for example [10], we concentrate on the limiting distribution of the 2SLS estimator for the cointegration case, which is only sparsely discussed in the literature, see for example [16], where the case that w_t in (7) is uncorrelated with x_{t-1} is discussed. Theorem 1 states the limiting distribution of the multiplicator estimator, $\hat{\alpha}$, and the 2SLS cointegrating vector estimator, $\hat{\beta}$.

Theorem 1 *When the Data Generating Process (DGP) in (7) is such that the number of cointegrating vectors equals r (k − r unit roots), the estimators*

$$\begin{aligned}\hat{\alpha} \;=\; & (\sum_{t=1}^{T}\Delta x_t(I - x'_{2t-1}(\sum_{t=1}^{T}x_{2t-1}x'_{2t-1})^{-1}x_{2t-1})x'_{1t-1}) \\ & (\sum_{t=1}^{T}x_{1t-1}(I - x'_{2t-1}(\sum_{t=1}^{T}x_{2t-1}x'_{2t-1})^{-1}x_{2t-1})x'_{1t-1})^{-1}\end{aligned} \qquad (15)$$

and

$$\hat{\beta} = (\sum_{t=1}^{T}x_{t-1}x'_{t-1})^{-1}(\sum_{t=1}^{T}x_{t-1}\Delta x'_t)\hat{\Omega}^{-1}\hat{\alpha}(\hat{\alpha}'\hat{\Omega}^{-1}\hat{\alpha})^{-1} \qquad (16)$$

have a limiting behavior which can be characterized by

$$\sqrt{T}(\hat{\alpha}-\alpha) \Rightarrow n(0,\ cov(\beta' x)^{-1} \otimes \Omega)) \qquad (17)$$

$$\begin{aligned}T(\hat{\beta}-\beta) \;\Rightarrow\; & \begin{pmatrix} 0 \\ (\beta'_{\perp}\beta_{\perp})^{-1}\beta'_{\perp}\alpha_{\perp}\sigma_1^{-1\prime}(\int B_1B'_1)^{-1}\int B_1 dB'_2\sigma_2 \end{pmatrix} \\ \Rightarrow\; & \begin{pmatrix} 0 \\ n(0, \alpha'\Omega^{-1}\alpha \otimes \Psi \end{pmatrix},\end{aligned} \qquad (18)$$

where B_1, resp. B_2 are $(k-r)$, resp. r dimensional stochastically independent Brownian motions defined on the unit interval, $\sigma_1 = (\alpha'_{\perp}\Omega\alpha_{\perp})^{\frac{1}{2}}$, $\sigma_2 = (\alpha'\Omega^{-1}\alpha)^{\frac{1}{2}}$, $\Psi = (\beta'_{\perp}\beta_{\perp})^{-1}\beta'_{\perp}\alpha_{\perp}\sigma_1^{-1\prime}(\int B_1B'_1)^{-1}\sigma_1^{-1}\alpha'_{\perp}\beta_{\perp}(\beta'_{\perp}\beta_{\perp})^{-1}$ and Ω is estimated by the sum of squared residuals, $\hat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}(\Delta x_t(I - x'_{t-1}(\sum_{t=1}^{T}x_{t-1}x'_{t-1})^{-1}x_{t-1})\Delta x'_t)$.

Proof: see appendix.

Theorem 1 discusses the limiting distribution of the cointegrating vector estimator for the most straightforward case, i.e. no further lags in the VAR polynomial and no deterministic components, and shows that it is identical to the limiting distribution of the canonical correlation maximum likelihood estimator of Johansen, see [?]. While addition of lags of Δx_t only changes the limiting distribution of the cointegrating vector estimator, $\hat{\beta}$, in the sense that $\alpha'_\perp\beta_\perp$ has to be replaced by $\alpha'_\perp\Gamma(1)\beta_\perp$, where $\Gamma(L)\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t$ and $\Gamma(L)$ is a $(p-1)$-dimensional lag polynomial in case of a VAR(p), inclusion of deterministic components does also change the functional form of the cointegrating vector estimator, see for example [?] and [?] for the influence of the deterministic components on other kind of cointegrating vector estimators. Theorem 2 states the estimators and limiting distributions of the multiplicator and cointegrating vector estimators including deterministic components for a few commonly used specifications of the deterministic components.

Theorem 2 *When the DGP reads*

$$\Delta x_t = \alpha(\beta' x_{t-1} + \gamma' t) + \varepsilon_t, \tag{19}$$

and the number of cointegrating vectors equals r ($k-r$ unit roots), the estimators

$$\begin{aligned}\hat{\alpha} = {} & \Big(\sum_{t=1}^{T}\Delta x_t\Big(1 - \begin{pmatrix} x_{2t-1} \\ t \end{pmatrix}'\Big(\sum_{t=1}^{T}\begin{pmatrix} x_{2t-1} \\ t \end{pmatrix}\begin{pmatrix} x_{2t-1} \\ t \end{pmatrix}'\Big)^{-1} \\ & \begin{pmatrix} x_{2t-1} \\ t \end{pmatrix}\Big)x'_{1t-1}\Big)\Big(\sum_{t=1}^{T} x_{1t-1}\Big(1 - \begin{pmatrix} x_{2t-1} \\ t \end{pmatrix}'\Big(\sum_{t=1}^{T}\begin{pmatrix} x_{2t-1} \\ t \end{pmatrix}\begin{pmatrix} x_{2t-1} \\ t \end{pmatrix}'\Big)^{-1} \\ & \begin{pmatrix} x_{2t-1} \\ t \end{pmatrix}\Big)x'_{1t-1}\Big)^{-1}\end{aligned} \tag{20}$$

and

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \Big(\sum_{t=1}^{T}\begin{pmatrix} x_{t-1} \\ t \end{pmatrix}\begin{pmatrix} x_{t-1} \\ t \end{pmatrix}'\Big)^{-1}\Big(\sum_{t=1}^{T}\begin{pmatrix} x_{t-1} \\ t \end{pmatrix}\Delta x'_t\Big)\Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1} \tag{21}$$

have a limiting behavior which can be characterized by

$$\sqrt{T}(\hat{\alpha} - \alpha) \Rightarrow N(0,\ var(\beta' x - \gamma' t)^{-1} \otimes \Sigma) \tag{22}$$

$$\begin{aligned} & \begin{pmatrix} T I_k & 0 \\ 0 & T^{\frac{1}{2}} \end{pmatrix}\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} \\ \Rightarrow {} & \begin{pmatrix} 0 \\ \begin{pmatrix} (\beta'_\perp\beta_\perp)^{-1}\beta'_\perp\alpha_\perp A_1^{-1\prime} & 0 \\ 0 & 1 \end{pmatrix}\Big(\int\begin{pmatrix} W_1 \\ t \end{pmatrix}\begin{pmatrix} W_1 \\ t \end{pmatrix}'\Big)^{-1}\Big(\int\begin{pmatrix} W_1 \\ t \end{pmatrix} dW_2\Big)A'_2 \end{pmatrix} \\ \Rightarrow {} & \begin{pmatrix} 0 \\ N(0, \alpha'\Sigma^{-1}\alpha \otimes \Xi_1) \end{pmatrix}, \end{aligned} \tag{23}$$

When the DGP reads

$$\Delta x_t = \mu + \alpha\beta' x_{t-1} + \varepsilon_t, \tag{24}$$

$\mu = \alpha c' + \alpha_\perp \delta'$, and the number of cointegrating vectors equals r ($k - r$ unit roots), the estimators

$$\begin{aligned}\hat{\alpha} \;=\; & (\sum_{t=1}^{T} \Delta x_t (I - \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}' (\sum_{t=1}^{T} \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix} \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}')^{-1} \\ & \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}) x_{1t-1}')(\sum_{t=1}^{T} x_{1t-1} (I - \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}' (\sum_{t=1}^{T} \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix} \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}')^{-1} \\ & \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}) x_{1t-1}')^{-1}\end{aligned} \tag{25}$$

and

$$\begin{pmatrix} \hat{\beta} \\ \hat{c} \end{pmatrix} = (\sum_{t=1}^{T} \begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix}')^{-1} (\sum_{t=1}^{T} \begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix} \Delta x_t') \Sigma^{-1} \alpha (\alpha' \Sigma^{-1} \alpha)^{-1} \tag{26}$$

have a limiting behavior which can be characterized by

$$\sqrt{T}(\hat{\alpha} - \alpha) \Rightarrow n(0,\; var(\beta' x - c')^{-1} \otimes \Sigma)) \tag{27}$$

$$\begin{aligned} & \begin{bmatrix} \begin{pmatrix} T I_{k-r-1} & 0 \\ 0 & T^{\frac{1}{2}} \end{pmatrix} (\beta_{\perp 2}^{*\prime} \beta_{\perp 2}^{*})^{-1} \beta_{\perp 2}^{*\prime} & 0 \\ 0 & T^{\frac{1}{2}} \end{bmatrix} \begin{pmatrix} \hat{\beta}_2 - \beta_2 \\ \hat{c} - c \end{pmatrix} \\ \Rightarrow \;& \begin{pmatrix} \Delta_g & 0 \\ 0 & 1 \end{pmatrix}^{-1\prime} (\int \begin{pmatrix} W_{11} \\ \tau \\ 1 \end{pmatrix} \begin{pmatrix} W_{11} \\ \tau \\ 1 \end{pmatrix}')^{-1} \int \begin{pmatrix} W_{11} \\ \tau \\ 1 \end{pmatrix} dW_2' \Delta_2' \\ \Rightarrow \;& n(0,\; \alpha' \Sigma^{-1} \alpha \otimes \Omega_2) \end{aligned} \tag{28}$$

When the DGP reads

$$\Delta x_t = \mu + \alpha(\beta' x_{t-1} + \delta' t) + \varepsilon_t, \tag{29}$$

$\mu = \alpha c' + \alpha_\perp \delta'$, and the number of cointegrating vectors equals r ($k - r$ unit roots), the estimators

$$\begin{aligned}\hat{\alpha} \;=\; & (\sum_{t=1}^{T} \Delta x_t (I - \begin{pmatrix} x_{2t-1} \\ 1 \\ t \end{pmatrix}' (\sum_{t=1}^{T} \begin{pmatrix} x_{2t-1} \\ 1 \\ t \end{pmatrix} \begin{pmatrix} x_{2t-1} \\ 1 \\ t \end{pmatrix}')^{-1} \\ & \begin{pmatrix} x_{2t-1} \\ 1 \\ t \end{pmatrix}) x_{1t-1}')(\sum_{t=1}^{T} x_{1t-1} (I - \begin{pmatrix} x_{2t-1} \\ 1 \\ t \end{pmatrix}' (\sum_{t=1}^{T} \begin{pmatrix} x_{2t-1} \\ 1 \\ t \end{pmatrix} \begin{pmatrix} x_{2t-1} \\ 1 \\ t \end{pmatrix}')^{-1} \\ & \begin{pmatrix} x_{2t-1} \\ 1 \\ t \end{pmatrix}) x_{1t-1}')^{-1}\end{aligned} \tag{30}$$

and

$$\begin{pmatrix} \hat{\beta} \\ \hat{\mu} \\ \hat{\delta} \end{pmatrix} = \left(\sum_{t=1}^{T} \begin{pmatrix} x_{t-1} \\ 1 \\ t \end{pmatrix} \begin{pmatrix} x_{t-1} \\ 1 \\ t \end{pmatrix}'\right)^{-1} \left(\sum_{t=1}^{T} \begin{pmatrix} x_{t-1} \\ 1 \\ t \end{pmatrix} \Delta x_t'\right) \Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1} \quad (31)$$

have a limiting behavior which can be characterized by

$$\sqrt{T}(\hat{\alpha} - \alpha) \Rightarrow N(0,\ cov(\beta' x - \mu' - \delta' t)^{-1} \otimes \Sigma)) \quad (32)$$

$$\begin{pmatrix} TI_k & 0 & 0 \\ 0 & T^{\frac{1}{2}} & 0 \\ 0 & 0 & T^{\frac{3}{2}} \end{pmatrix} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\mu} - \mu \\ \hat{\delta} - \delta \end{pmatrix} \quad (33)$$

$$\Rightarrow \begin{pmatrix} 0 \\ \begin{pmatrix} (\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\alpha_\perp\Omega_1^{-1'} & 0 \\ 0 & I_2 \end{pmatrix} \left(\int \begin{pmatrix} W_1 \\ \iota \\ \tau \end{pmatrix} \begin{pmatrix} W_1 \\ \iota \\ \tau \end{pmatrix}'\right)^{-1} \left(\int \begin{pmatrix} W_1 \\ \iota \\ \tau \end{pmatrix} dW_2'\right) \Omega_2' \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 0 \\ N(0, \alpha'\Sigma^{-1}\alpha \otimes \Psi_3) \end{pmatrix},$$

where W_1, W_{11} and W_2 are $(k-r)$, $(k-r-1)$ and r dimensional stochastically independent Brownian motions, $\Omega_1 = (\alpha_\perp'\Sigma\alpha_\perp)^{\frac{1}{2}}$, $\Omega_2 = (\alpha'\Sigma^{-1}\alpha)^{\frac{1}{2}}$, $\Omega_3 = \left(\begin{pmatrix} d_\perp(\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp'\Sigma^{\frac{1}{2}} \\ dd' \end{pmatrix} \begin{pmatrix} d_\perp(\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp'\Sigma^{\frac{1}{2}} \\ dd' \end{pmatrix}'\right)^{\frac{1}{2}}$, $\tau(t) = t$, $\iota(t) = 1$, $0 \leq t \leq 1$, $\beta = \begin{pmatrix} I_r \\ -\beta_2 \end{pmatrix}$, $\beta_{2\perp} = (\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\alpha_\perp(\alpha_\perp'\alpha_\perp)^{-1}\begin{pmatrix} d_\perp \\ d \end{pmatrix}'$,

$$\Psi_1 = (\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\alpha_\perp\Omega_1^{-1'}\left(\int \begin{pmatrix} W_1 \\ \iota \end{pmatrix} \begin{pmatrix} W_1 \\ \iota \end{pmatrix}'\right)^{-1}\Omega_1^{-1}\alpha_\perp'\beta_\perp(\beta_\perp'\beta_\perp)^{-1},$$

$$\Psi_2 = \begin{pmatrix} \Omega_3 & 0 \\ 0 & 1 \end{pmatrix}^{-1'}\left(\int \begin{pmatrix} W_{11} \\ \tau \\ \iota \end{pmatrix} \begin{pmatrix} W_{11} \\ \tau \\ \iota \end{pmatrix}'\right)^{-1}\begin{pmatrix} \Omega_3 & 0 \\ 0 & 1 \end{pmatrix}^{-1},$$

$$\Psi_3 = (\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\alpha_\perp\Omega_1^{-1'}\left(\int \begin{pmatrix} W_1 \\ \iota \\ \tau \end{pmatrix} \begin{pmatrix} W_1 \\ \iota \\ \tau \end{pmatrix}'\right)^{-1}\Omega_1^{-1}\alpha_\perp'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}.$$

Proof: the first and third part of the theorem are natural extensions of theorem 1. The second part of the theorem is proved in the appendix.

Theorems 1 and 2 show that the limiting distributions of elements of the cointegrating vector estimator are normal and standard (asymptotic) χ^2 tests can be performed to test hypotheses on the cointegrating vectors, see [12]. The next section discusses the use of the cointegrating vector estimator, $\hat{\beta}$, and the multiplicator, $\hat{\alpha}$, in the ECM objective function, (7), to construct a statistic to test for the number of cointegrating vectors, unit roots, in the system.

8 Cointegration testing using GMM-2SLS estimators

The GMM objective function, (7), can also be used to test for the number of cointegrating vectors, unit roots. This can be done as the optimal value of the objective function has a specific kind of limiting distribution under $H_0 : r = r^*$. In theorem 8, the functional expressions of this objective function for several specifications of the deterministic components and their limiting distributions are stated.

Theorem 8 *When (1) is the DGP and the number of cointegrating vectors equals r ((k-r) unit roots), the use of the estimators ([illegible]) and ([illegible]) for the optimal value of the GMM objective function (7) leads to a limiting behavior of this optimal value which can be characterized by*

$$X(\hat{\alpha}, \hat{\beta}) \Rightarrow tr[(\int W_1 dW_1')'(\int W_1 W_1')^{-1}(\int W_1 dW_1')]. \quad (84)$$

When ([illegible]) is the DGP and the number of cointegrating vectors equals r ((k-r) unit roots), the use of the estimators ([illegible]) and ([illegible]) for the optimal value of the GMM objective function (7) leads to a limiting behavior of this optimal value which can be characterized by

$$\begin{aligned} & X(\hat{\alpha}, \hat{\beta}, \hat{\mu}) \quad (85) \\ \Rightarrow \; & tr[(\int \begin{pmatrix} W_1 \\ 1 \end{pmatrix} dW_1')'(\int \begin{pmatrix} W_1 \\ 1 \end{pmatrix} \begin{pmatrix} W_1 \\ 1 \end{pmatrix}')^{-1}(\int \begin{pmatrix} W_1 \\ 1 \end{pmatrix} dW_1')]. \end{aligned}$$

When ([illegible]) is the DGP and the number of cointegrating vectors equals r ((k-r) unit roots), the use of the estimators ([illegible]) and ([illegible]) for the optimal value of the GMM objective function (7) leads to a limiting behavior of this optimal value which can be characterized by

$$\begin{aligned} & X(\hat{\alpha}, \hat{\beta}, \hat{\delta}) \quad (86) \\ \Rightarrow \; & tr[(\int \begin{pmatrix} W_{11} \\ t \end{pmatrix} dW_1')'(\int \begin{pmatrix} W_{11} \\ t \end{pmatrix} \begin{pmatrix} W_{11} \\ t \end{pmatrix}')^{-1}(\int \begin{pmatrix} W_{11} \\ t \end{pmatrix} dW_1')]. \end{aligned}$$

When ([illegible]) is the DGP and the number of cointegrating vectors equals r ((k-r) unit roots), the use of the estimators ([illegible]) and ([illegible]) for the optimal value of the GMM objective function (7) leads to a limiting behavior of this optimal value which can be characterized by

$$\begin{aligned} & X(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\delta}) \quad (87) \\ \Rightarrow \; & tr[(\int \begin{pmatrix} W_1 \\ t \end{pmatrix} dW_1')'(\int \begin{pmatrix} W_1 \\ t \end{pmatrix} \begin{pmatrix} W_1 \\ t \end{pmatrix}')^{-1}(\int \begin{pmatrix} W_1 \\ t \end{pmatrix} dW_1')], \end{aligned}$$

where $x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix}$, $x_{1t} : r \times 1$, $x_{2t} : (k-r) \times 1$; $\bar{x}_{t-1} = x_{t-1} - \frac{1}{T}\sum_{t=1}^{T} x_{t-1}$, $\bar{t} = t - \frac{1}{T}\sum_{t=1}^{T} t$; W_1, W_{11} *are* $(k-r)$, $(k-r-1)$ *dimensional brownian motions,* $W_1 = \begin{pmatrix} W_{11} \\ w_{12} \end{pmatrix}$, $\bar{W}_1 = W_1 - \int W_1$, $\bar{W}_{11} = W_{11} - \int W_{11}$, $\tau(t) = t$, $\iota(t) = 1$, $0 \leq t \leq 1$, $\bar{\tau} = \tau - \int \tau$, *and* Ω *is estimated by the residual sum of squares for the unrestricted model.*

Proof: for the first part a proof is given in the appendix, the other parts follow straightforwardly.

Theorems 1 to 5 show that the limiting distributions using the GMM-2SLS estimators are identical to the limiting distributions when maximum likelihood estimators are used, see [4]. As maximum likelihood estimators can be constructed in a straightforward way using canonical correlations there is not much gain when 2SLS estimators are used compared to maximum likelihood estimators from a limiting distribution perspective. Possible gains can be both in the small sample distribution of the 2SLS estimator and in model extensions as maximum likelihood estimators become analytically intractable when more complicated models are used than the one shown in (1).

In [15], it is shown that the canonical correlation cointegrating vector estimator has a small sample distribution with Cauchy type tails such that it has no finite moments. When we neglect the dynamic property of the data and assume fixed regressors, results from [10] indicate that the small sample distribution of the 2SLS cointegrating vector estimator has finite moments up to the degree $(k-r)$. This degree is determined by the $\Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}$ expression appearing in the cointegrating vector estimator $\hat{\beta}$. As $\hat{\beta}$ is specified such that it always has rank r, rank reduction of $\alpha\beta'$ implies that α has a lower rank value. In that case $\hat{\alpha}'\Sigma^{-1}\hat{\alpha}$ would not be invertible leading to the fat tails of the small sample distribution. So, cointegration tests essentially test for the rank of α and can be considered as tests for the local identification of β and are, therefore, comparable with the concentration parameter in the INSEM, see [10].

The maximum likelihood cointegrating vector estimator is appealing as it has a simple expression in the standard case. The relation between maximum likelihood cointegrating vector estimators and canonical correlations is, however, lost when extensions of the model are considered. Furthermore, model extensions often lead to analytically intractable maximum likelihood estimators. The GMM framework used in this paper offers a framework which allows for the analytical construction of cointegrating vector estimators for a general class of models. In the next sections two kind of structural break model extensions are analyzed, i.e. structural breaks in the variance (heteroscedasticity) and cointegrating vector and/or multiplicator, whose cointegrating vector maximum likelihood estimators are not of the canonical correlation type.

4 Cointegration in a Model with Heteroscedasticity

Assuming homoscedastic disturbances in (1), the maximum likelihood estimator of the cointegrating vector can be constructed using canonical correlations. This estimator has a normal limiting distribution under conditions which are more general than strict homoscedasticity, see [15], where it is shown that the weak convergence properties are retained in case of conditional heteroscedasticity with constant unconditional variances. These weak convergence properties are, however, lost when the mean of the conditional variances changes from period to period. Furthermore, also the relation between the maximum likelihood estimator and canonical correlations is lost in that case. A GMM-GLS cointegrating vector estimator can still be constructed when the functional form of the heteroscedasticity is known. We construct estimators and limiting distributions for an example of a change of the variance after a predefined period of time T_1 has evolved such that the analyzed model reads,

$$\Delta x_t = \alpha\beta' x_{t-1} + \varepsilon_t, \tag{38}$$

where

$$\begin{aligned} var(\varepsilon_t) &= \Sigma_1, & t = 1, ..., T_1 \\ &= \Sigma_2, & t = T_1 + 1, ..., T \end{aligned} \tag{39}$$

In the next subsection, the GMM cointegrating vector estimator and cointegration test and their limiting distributions are derived using a Generalized Least Squares (GLS) framework to account for the heteroscedasticity.

4.1 Generalized Least Squares Cointegration Estimators

Assuming that we know the form of heteroscedasticity, a different GMM objective function than (7) is used in the construction of the GMM estimators,

$$\begin{aligned} Q(\alpha, \beta) = \; & vec(\sum_{t=1}^{T_1} \Sigma_1^{-1} \varepsilon_t x_{t-1}' + \sum_{t=T_1+1}^{T} \Sigma_2^{-1} \varepsilon_t x_{t-1}')' \\ & (\sum_{t=1}^{T_1} (x_{t-1} x_{t-1}' \otimes \Sigma_1^{-1}) + \sum_{t=T_1+1}^{T} (x_{t-1} x_{t-1}' \otimes \Sigma_2^{-1}))^{-1} \\ & vec(\sum_{t=1}^{T_1} \Sigma_1^{-1} \varepsilon_t x_{t-1}' + \sum_{t=T_1+1}^{T} \Sigma_2^{-1} \varepsilon_t x_{t-1}'). \end{aligned} \tag{40}$$

In the next theorem the GMM estimators and their limiting distributions jointly with the limiting distribution of the optimal value of the GMM objective function are stated.

Theorem 4 *When the DGP in equations (38), (39) is such that the number of cointegrating vectors is r (k − r unit roots), the estimators,*

$$\begin{aligned} vec(\hat\alpha) &= ((\sum_{t=1}^{T_1} z_{t-1}z'_{t-1} \otimes \hat\Sigma_1^{-1}) + (\sum_{t=T_1+1}^{T} z_{t-1}z'_{t-1} \otimes \hat\Sigma_2^{-1}))_1^{-1} \qquad (41)\\ & vec(\sum_{t=1}^{T_1} \hat\Sigma_1^{-1}\Delta x_t z'_{t-1} + \sum_{t=T_1+1}^{T} \hat\Sigma_2^{-1}\Delta x_t z'_{t-1}), \end{aligned}$$

and

$$\begin{aligned} vec(\hat\beta') &= ((\sum_{t=1}^{T_1} z_{t-1}z'_{t-1} \otimes \hat\alpha'\hat\Sigma_1^{-1}\hat\alpha) + (\sum_{t=T_1+1}^{T} z_{t-1}z'_{t-1} \otimes \hat\alpha'\hat\Sigma_2^{-1}\hat\alpha))^{-1} \qquad (42)\\ & vec(\sum_{t=1}^{T_1} \hat\alpha'\hat\Sigma_1^{-1}\Delta x_t z'_{t-1} + \sum_{t=T_1+1}^{T} \hat\alpha'\hat\Sigma_2^{-1}\Delta x_t z'_{t-1}), \end{aligned}$$

have a limiting behavior which can be characterized by

$$\begin{aligned} \sqrt{T} vec(\hat\alpha - \alpha) &\Rightarrow n(0, (\pi(cov(\beta' z)_1 \otimes \alpha'\Sigma_1^{-1}\alpha) \qquad (43)\\ & +(1-\pi)(cov(\beta' z)_2 \otimes \alpha'\Sigma_2^{-1}\alpha))^{-1}), \end{aligned}$$

and

$$\begin{aligned} & T[vec(\beta_2 - \hat\beta_2)] \qquad (44)\\ \Rightarrow\; & ((\beta'_\perp\beta_\perp)^{-1}\beta'_\perp \otimes I_r)((C_1\int_0^\pi W_1W_1'C_1' \otimes \alpha'\Sigma_1^{-1}\alpha) +\\ & (\int_\pi^1 (C_2W_1(t) + C_1W_1(\pi))(C_2W_1(t) + C_1W_1(\pi))'dt \otimes \alpha'\Sigma_2^{-1}\alpha))^{-1}\\ & vec[\Omega_1(\int_0^\pi dW_2W_1')C_1' + \Omega_2(\int_\pi^1 dW_2(t)(C_2W_1(t) + C_1W_1(\pi))'dt)], \end{aligned}$$

The limiting behavior of the optimal value of the GMM objective function, can be characterized by

$$\begin{aligned} & Q(\hat\alpha, \hat\beta) \qquad (45)\\ \Rightarrow\; & vec[C_1(\int_0^\pi dW_1W_1')C_1' + C_2(\int_\pi^1 dW_1(t)(C_2W_1(t) + C_1W_1(\pi))'dt)]' \end{aligned}$$

$$((\Delta_1 \int_0^{\pi} W_1 W_1' \Delta_1' \otimes a_\perp' \Sigma_1^{-1} a_\perp) + (\int_\pi^1 (\Delta_2 W_1(t) - \Delta_1 W_1(\pi))$$
$$(\Delta_2 W_1(t) - \Delta_1 W_1(\pi))' dt \otimes a_\perp' \Sigma_2^{-1} a_\perp))^{-1}$$
$$vec[\Delta_1 (\int_0^{\pi} dW_2 W_1') \Delta_1' + \Delta_2 (\int_\pi^1 dW_2 W_1(t) (\Delta_2 W_1(t) - \Delta_1 W_1(\pi))' dt)],$$

where $\pi = \frac{T_1}{T}$, W_1 *and* W_2, *are stochastically independent* r, $(k-r)$ *dimensional Brownian motions with identity covariance matrices,* $\Delta_1 = (a_\perp' \Sigma_1^{-1} a_\perp)^{\frac{1}{2}}$, $\Delta_2 = (a_\perp' \Sigma_2^{-1} a_\perp)^{\frac{1}{2}}$, $\Omega_1 = (a' \Sigma_1 a)^{\frac{1}{2}}$, $\Omega_2 = (a' \Sigma_2 a)^{\frac{1}{2}}$, $\hat{\Sigma}_1 = \frac{1}{T_1 - k} \sum_{t=1}^{T_1} \Delta y_t (I - y_{t-1}' (\sum_{t=1}^{T_1} y_{t-1} y_{t-1}')^{-1} y_{t-1}) \Delta y_t'$, $\hat{\Sigma}_2 = \frac{1}{T - T_1 - k} \sum_{t=T_1+1}^{T} \Delta y_t (I - y_{t-1}' (\sum_{t=T_1+1}^{T} y_{t-1} y_{t-1}')^{-1} y_{t-1}) \Delta y_t'$, $cov(\beta' y)_1 = \beta' \sum_{i=0}^{\infty} C_i^* \Sigma_1 C_i^{*\prime} \beta$, $cov(\beta' y)_2 = \beta' \sum_{i=0}^{\infty} C_i^* \Sigma_2 C_i^{*\prime} \beta$, $(\ )_1^{-1}$ *are the first* k *rows of* $(\ )^{-1}$.

Proof: the asymptotic results for subsamples using the fraction π stem from [8]. Using these asymptotics for subsamples, the other results follow straightforwardly from the proofs of theorems 1-3.

The cointegrating vector estimator in theorem 4 is a 2SLS estimator as it is constructed in two sequential steps. In the first step, we estimate Π in (3) using least squares and use its first r columns to construct a. Furthermore, we construct $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ as the sum of squared residuals of the two subsamples. In the second step, we construct the estimator $\hat{\beta}$ (42).

Theorem 4 shows that the cointegrating vector estimator $\hat{\beta}$ has a normal limiting distribution. When we use a cointegrating vector estimator which neglects the heteroscedasticity of the disturbances, we cannot find accurate expressions of the covariance matrix such that it is hard to test hypotheses on the cointegrating vector in that case. Although the cointegrating relationships are not weakly stationary in this case, as they have a different variance in each of the two variance regimes, they still show mean reversion. The estimators and limiting distributions from theorem 4 can be extended to more variance shifts and other moment conditions (relationships) for the variances can be incorporated. The limiting distribution of the optimal value of the GMM objective function depends on the relative change of the covariance matrix and the relative time period during which the variance differs, π. As it is not known what the true values of these parameters are, they are typically replaced by sample estimates. The resulting distribution is in that case no longer the true limiting distribution but only an approximation of it. It is interesting to investigate whether nonparametric covariance estimators, like the White covariance matrix estimator, see [17], can be used to overcome these difficulties. These covariance matrix estimators can directly be used in the GMM objective function but expressions of the resulting limiting distributions are still unknown.

Cointegration with structural breaks

In this section, we investigate the influence of a change in the value of the multiplicator, α, and cointegrating vector, β, at T_1. The model, therefore, is

$$\begin{aligned} \Delta y_t &= \alpha\beta' y_{t-1} + \varepsilon_t & t = 1, ..., T_1, \\ \Delta y_t &= \delta\gamma' y_{t-1} + \varepsilon_t & t = T_1 + 1, ..., T, \end{aligned} \tag{46}$$

where ε_t, $t = 1, ..., T$, are Gaussian white noise disturbances with covariance matrix Ω. The GMM objective function corresponding with this model reads,

$$\begin{aligned} Q(\alpha, \beta, \gamma, \delta) = {} & vec(\sum_{t=1}^{T_1} \varepsilon_t y_{t-1}', \sum_{t=T_1+1}^{T} \varepsilon_t y_{t-1}')' \\ & \begin{pmatrix} ((\sum_{t=1}^{T_1} y_{t-1}y_{t-1}')^{-1} \otimes \Omega^{-1}) & 0 \\ 0 & ((\sum_{t=T_1+1}^{T} y_{t-1}y_{t-1}')^{-1} \otimes \Omega^{-1}) \end{pmatrix} \\ & vec(\sum_{t=1}^{T_1} \varepsilon_t y_{t-1}', \sum_{t=T_1+1}^{T} \varepsilon_t y_{t-1}'), \end{aligned} \tag{47}$$

where $vec(A, B) = (vec(A)' \; vec(B)')'$. In theorem 6, the GMM estimators of the cointegrating vector, multiplicator and their limiting distributions are stated jointly with the limiting distribution of the GMM objective function. As the cointegrating vector estimators and multiplicators all have normal limiting distribution, standard χ^2 tests can be performed to test for the equality of the parameters in each of the two periods. Theorem 6 also states the estimators and their limiting distributions, which can be used when either the cointegrating vectors or multiplicators in each of the two periods are equal to one another.

Theorem 5 *When the DGP in [illegible] is such that the number of cointegrating vectors is r (k-r unit roots), the estimators,*

$$\begin{aligned} \hat{\alpha} = {} & (\sum_{t=1}^{T_1} \Delta y_t (I - y_{2t-1}'(\sum_{t=1}^{T_1} y_{2t-1}y_{2t-1}')^{-1} y_{2t-1}) y_{1t-1}') \\ & (\sum_{t=1}^{T_1} y_{1t-1} (I - y_{2t-1}'(\sum_{t=1}^{T_1} y_{2t-1}y_{2t-1}')^{-1} y_{2t-1}) y_{1t-1}')^{-1}, \end{aligned} \tag{48}$$

$$\hat{\delta} = (\sum_{t=T_1+1}^{T} \Delta y_t (I - y_{2t-1}'(\sum_{t=T_1+1}^{T} y_{2t-1}y_{2t-1}')^{-1} y_{2t-1}) y_{1t-1}') \tag{49}$$

$$(\sum_{t=T_1+1}^{T} x_{1t-1}(1 - x'_{2t-1}(\sum_{t=T_1+1}^{T} x_{2t-1}x'_{2t-1})^{-1}x_{2t-1})x'_{1t-1})^{-1},$$

and

$$\hat{\beta} = (\sum_{t=1}^{T_1} x_{t-1}x'_{t-1})^{-1}(\sum_{t=1}^{T_1} x_{t-1}\Delta y'_t)\Sigma_1^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma_1^{-1}\hat{\alpha})^{-1}, \quad (50)$$

$$\hat{\gamma} = (\sum_{t=T_1+1}^{T} x_{t-1}x'_{t-1})^{-1}(\sum_{t=T_1+1}^{T} x_{t-1}\Delta y'_t)\Sigma_2^{-1}\hat{\theta}(\hat{\theta}'\Sigma_2^{-1}\hat{\theta})^{-1}, \quad (51)$$

have a limiting behavior which can be characterized by,

$$\sqrt{T}(\hat{\alpha} - \alpha) \Rightarrow n(0,\ \text{cov}(\beta' x)^{-1} \otimes \kappa\Sigma), \quad (52)$$

$$\sqrt{T}(\hat{\theta} - \theta) \Rightarrow n(0,\ \text{cov}(\gamma' x)^{-1} \otimes (1-\kappa)\Sigma), \quad (53)$$

and

$$T(\hat{\beta} - \beta) \Rightarrow \begin{pmatrix} 0 \\ (\beta'_\perp\beta_\perp)^{-1}\beta'_\perp\alpha\Lambda_1^{-1\prime}(\int_0^\kappa W_1W_1')^{-1}(\int_0^\kappa W_1 dW_2')Q'_1 \end{pmatrix}, \quad (54)$$

$$\begin{aligned} T(\gamma_2 - \hat{\gamma}_2) \Rightarrow & [(\int_\kappa^1(\gamma'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_1 W_1(\kappa) + \gamma'_\perp\gamma_\perp(\theta'_\perp\gamma_\perp)^{-1}\Lambda_2 W_1(t)) \quad (55) \\ & (\gamma'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_1 W_1(\kappa) + \gamma'_\perp\gamma_\perp(\theta'_\perp\gamma_\perp)^{-1}\Lambda_2 W_1(t))'dt)]^{-1} \\ & \int_\kappa^1[\gamma'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_1 W_1(\kappa) + \gamma'_\perp\gamma_\perp(\theta'_\perp\gamma_\perp)^{-1}\Lambda_2 W_1(t)) \\ & dW_2(t)'dt]Q'_2. \end{aligned}$$

The limiting behavior of the optimal value of the objective function can be characterized by,

$$\begin{aligned} & LR(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\theta}) \quad (56) \\ \Rightarrow\ & \text{vec}((\int_0^\kappa W_1 dW_1')')'((\int_0^\kappa W_1W_1')^{-1} \otimes I_{k-r})\text{vec}((\int_0^\kappa W_1 dW_1')') \\ & + \text{vec}(\int_\kappa^1[\Lambda_2^{-1}\theta'_\perp\gamma_\perp(\gamma'_\perp\gamma_\perp)^{-1}\gamma'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_1 W_1(\kappa) \\ & + W_1(t)]dW_1(t)')'([\int_\kappa^1(\Lambda_2^{-1}\theta'_\perp\gamma_\perp(\gamma'_\perp\gamma_\perp)^{-1}\gamma'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_1 W_1(\kappa) \\ & + W_1(t))(\Lambda_2^{-1}\theta'_\perp\gamma_\perp(\gamma'_\perp\gamma_\perp)^{-1}\gamma'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_1 W_1(\kappa) \\ & + W_1(t))'dt]^{-1} \otimes I_{k-r})\text{vec}(\int_\kappa^1[\Lambda_2^{-1}\theta'_\perp\gamma_\perp(\gamma'_\perp\gamma_\perp)^{-1}\gamma'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\Lambda_1 W_1(\kappa) \\ & + W_1(t)]dW_1(t)'). \end{aligned}$$

When the model in (48) is such that the cointegrating vectors are equal in the two subsamples, $\beta = \gamma$, which can be tested for using an asymptotic χ^2 test, the GMM estimator for β reads (estimators for α and δ result from (48) and (49)),

$$\begin{aligned}
vec(\hat{\beta}') &= [(\sum_{t=1}^{T_1} z_{t-1}z_{t-1}' \otimes \hat{\alpha}'\Sigma_1^{-1}\hat{\alpha}) + \\
&\quad (\sum_{t=T_1+1}^{T} z_{t-1}z_{t-1}' \otimes \hat{\delta}'\Sigma_2^{-1}\hat{\delta})]^{-1}[(I_k \otimes \hat{\alpha}'\Sigma_1^{-1}) \\
&\quad (vec(\sum_{t=1}^{T_1} \Delta y_t z_{t-1}')) + (I_k \otimes \hat{\delta}'\Sigma_2^{-1})(vec(\sum_{t=T_1+1}^{T} \Delta y_t z_{t-1}'))] \\
&= vec(\begin{pmatrix} I_m \\ -\hat{\beta}_2 \end{pmatrix}') + [(\sum_{t=1}^{T_1} z_{t-1}z_{t-1}' \otimes \hat{\alpha}'\Sigma_1^{-1}\hat{\alpha}) \\
&\quad + (\sum_{t=T_1+1}^{T} z_{t-1}z_{t-1}' \otimes \hat{\delta}'\Sigma_2^{-1}\hat{\delta})]^{-1}[vec(\sum_{t=1}^{T_1} \hat{\alpha}'\Sigma_1^{-1}\varepsilon_t z_{t-1}') \\
&\quad + vec(\sum_{t=T_1+1}^{T} \hat{\delta}'\Sigma_2^{-1}\varepsilon_t z_{t-1}')]
\end{aligned} \tag{57}$$

and the limiting behavior of this estimator can be characterized by,

$$\begin{aligned}
&T vec(\hat{\beta}_2 - \beta_2) \\
\Rightarrow\ & ((\beta_\perp'\beta_\perp)^{-1} \otimes I_m)[((\alpha_\perp'\beta_\perp)^{-1}C_1(\int_0^u W_1W_1')C_1'(\beta_\perp'\alpha_\perp)^{-1} \otimes \alpha'\Sigma^{-1}\alpha) \\
&+ ([\int_u^1((\alpha_\perp'\beta_\perp)^{-1}C_1W_1(u) + (\delta_\perp'\beta_\perp)^{-1}C_2W_1(t))((\alpha_\perp'\beta_\perp)^{-1}C_1W_1(u) \\
&+ (\delta_\perp'\beta_\perp)^{-1}C_2W_1(t))'dt] \otimes \delta'\Sigma^{-1}\delta)]^{-1} vec(\Omega_1(\int_0^u dW_2W_1')C_1'(\beta_\perp'\alpha_\perp)^{-1} \\
&+ \Omega_2\int_u^1[dW_2(t)(W_1'(u)C_1'(\beta_\perp'\alpha_\perp)^{-1} + W_1'(t)C_2'(\beta_\perp'\delta_\perp)^{-1})dt]).
\end{aligned} \tag{58}$$

When the model in (48) is such that the multiplicators of cointegrating vectors are equal in the two subsamples, $\alpha = \delta$, which can be tested for using a χ^2 test, the GMM estimator for α reads (estimators for β and γ result from (50) and (51))

$$\hat{\alpha} = (\sum_{t=1}^{T_1} \Delta y_t z_{t-1}'\hat{\beta} + \sum_{t=T_1+1}^{T} \Delta y_t z_{t-1}'\hat{\gamma})(\hat{\beta}'\sum_{t=1}^{T_1} z_{t-1}z_{t-1}'\hat{\beta} + \hat{\gamma}'\sum_{t=T_1+1}^{T} z_{t-1}z_{t-1}'\hat{\gamma})^{-1} \tag{59}$$

and its limiting behavior can be characterized by

$$\sqrt{T}(\hat{\alpha} - \alpha) \Rightarrow n(0, u\,cov(\beta'z)_1 + (1-u)cov(\gamma'z)_2) \tag{60}$$

where $\pi = \frac{T_1}{T}$, W_1 and W_2, are stochastically independent r, $(k-r)$ dimensional Brownian motions with identity covariance matrices, $\Phi_1 = (\alpha_\perp'\Sigma^{-1}\alpha_\perp)^{\frac{1}{2}}$, $\Phi_2 = (\gamma_\perp'\Sigma^{-1}\gamma_\perp)^{\frac{1}{2}}$, $\Omega_1 = (\alpha'\Sigma\alpha)^{\frac{1}{2}}$, $\Omega_2 = (\gamma'\Sigma\gamma)^{\frac{1}{2}}$, $\hat{\Sigma}_1 = \frac{1}{T_1-k}\sum_{t=1}^{T_1}\Delta x_t(I - x_{t-1}'(\sum_{t=1}^{T_1} x_{t-1}x_{t-1}')^{-1}x_{t-1})\Delta x_t'$, $\hat{\Sigma}_2 = \frac{1}{T-T_1-k}\sum_{t=T_1+1}^{T}\Delta x_t(I - x_{t-1}'(\sum_{t=T_1+1}^{T} x_{t-1}x_{t-1}')^{-1}x_{t-1})\Delta x_t'$, $var(\beta'x)_1 = \beta'\sum_{i=0}^{\infty} C_{1i}\Sigma_1 C_{1i}'\beta$, $var(\gamma'x)_2 = \delta'\sum_{i=0}^{\infty} C_{2i}\Sigma_2 C_{2i}'\delta$, and $C_1(L)$, $C_2(L)$ are the Vector Moving Average representations of the first and second subsets.

Proof: again uses asymptotics for subsamples, see [8], and results from proofs of theorems 1-5.

Theorem 6 shows that the GMM estimators of the cointegrating vector and multiplicator have normal limiting distributions in case of breaks in the cointegrating vector and/or multiplicator. Similar to the limiting distribution of the optimal value of the GMM objective function in case of heteroscedasticity, the limiting distribution of the optimal value of the GMM objective function again depends on model parameters and the relative length of the subsamples. An approximation of this limiting distribution can again be constructed using the estimated values of the parameters, α, β, γ, δ and T_1. As this leads to a rather complicated testing procedure, it may be preferable to fix the number of cointegrating vectors a priori and just perform tests on the estimated cointegrated vectors and multiplicators, which are straightforward to construct. This reasoning also holds for the cointegration tests discussed in the previous section.

5 Conclusions

A GMM framework for cointegration analysis is developed allowing for extensions of the models which are analyzable using the maximum likelihood procedure documented in the literature. As examples, model extensions incorporating heteroscedasticity and structural breaks are discussed and the resulting cointegration estimators are shown to have normal limiting distributions while the optimal value of the GMM objective function has a limiting distribution which is a Brownian motion functional with additional parameters resulting from the change of properties of the involved Brownian motions. These additional parameters are essentially the parameters in the model with vary over time resulting in heteroscedasticity or structural breaks. In future work, we will apply the developed framework for e.g. cointegration analysis in financial series, for example term structure of interest rates. As heteroscedasticity is a stylized fact of these series, the standard cointegration procedures cannot be applied here as they lead to incorrect (asymptotic) variances of the estimators.

Appendix

Proof of theorem 1.

In [4], it is proved that the stochastic process x_t, from (1), can be represented by:

$\Delta x_t = C(L)\Sigma^{\frac{1}{2}}\varepsilon_t$,

where $\beta = \begin{pmatrix} I_r & -\beta_2' \end{pmatrix}'$, and ε_t is a $k-$variate Gaussian white noise process with zero mean and identity covariance matrix. Consequently,

$x_t = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\Sigma^{\frac{1}{2}}\sum_{i=1}^{t}\varepsilon_i + C^*(L)\Sigma^{\frac{1}{2}}\varepsilon_t$,

$x_{1t} = \beta_2'(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\Sigma^{\frac{1}{2}}\sum_{i=1}^{t}\varepsilon_i + \begin{pmatrix} I_r & 0 \end{pmatrix} C^*(L)\Sigma^{\frac{1}{2}}\varepsilon_t$,

$x_{2t} = (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\Sigma^{\frac{1}{2}}\sum_{i=1}^{t}\varepsilon_i + \begin{pmatrix} 0 & I_{k-r} \end{pmatrix} C^*(L)\Sigma^{\frac{1}{2}}\varepsilon_t$,

where $C(L) = C(1) + (1-L)C^*(L)$, $C^*(L) = \sum_{i=0}^{\infty} C_i^* L^i$.

The least squares estimator of α, $\hat{\alpha}$, can also be expressed as

$$\begin{aligned}
\hat{\alpha} - \alpha &= (\sum_{t=1}^{T} u_t(I - x_{2t-1}'(\sum_{t=1}^{T} x_{2t-1}x_{2t-1}')^{-1}x_{2t-1})x_{1t-1}') \\
&\quad (\sum_{t=1}^{T} x_{1t-1}(I - x_{2t-1}'(\sum_{t=1}^{T} x_{2t-1}x_{2t-1}')^{-1}x_{2t-1})x_{1t-1}')^{-1} \\
&= (\sum_{t=1}^{T} u_t(x_{1t-1} - \hat{\beta}_2' x_{2t-1})')(\sum_{t=1}^{T}(x_{1t-1} - \hat{\beta}_2' x_{2t-1})(x_{1t-1} - \hat{\beta}_2' x_{2t-1})')^{-1}
\end{aligned}$$

where $\hat{\beta}_2 = (\sum_{t=1}^{T} x_{2t-1}x_{2t-1}')^{-1}x_{2t-1}x_{1t-1}'$. $\hat{\beta}_2$ is a superconsistent estimator of β_2 and can therefore be treated as equal to β_2 in the derivation of the limiting distribution of $\hat{\alpha}$. Since

$\frac{1}{T}\sum_{t=1}^{T}(x_{1t-1} - \hat{\beta}_2' x_{2t-1})(x_{1t-1} - \hat{\beta}_2' x_{2t-1})' \Rightarrow var(\beta' x) = \beta' \sum_{i=0}^{\infty} C_i^* \Sigma C_i^{*\prime} \beta$,

and, $T^{-\frac{1}{2}}(\sum_{t=1}^{T} u_t(x_{1t-1} - \hat{\beta}_2' x_{2t-1})') \Rightarrow N(0,\ var(\beta' x) \otimes \Sigma)$, the limiting distribution of $\hat{\alpha}$ becomes

$$\sqrt{T}(\hat{\alpha} - \alpha) \Rightarrow N(0,\ var(\beta' x)^{-1} \otimes \Sigma).$$

With respect to the cointegrating vector,

$$\begin{aligned}
\hat{\beta} &= (\sum_{t=1}^{T} x_{t-1}x_{t-1}')^{-1}(\sum_{t=1}^{T} x_{t-1}\Delta x_t')\Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1} \\
&= \begin{pmatrix} I_r \\ (\sum_{t=1}^{T} x_{t-1}x_{t-1}')_2^{-1}(\sum_{t=1}^{T} x_{t-1}(x_{t-1}'\beta\alpha' + u_t'))\Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1} \end{pmatrix}
\end{aligned}$$

$$= \begin{pmatrix} I_r \\ (\sum_{t=1}^{T} z_{t-1}z_{t-1}')_2^{-1}(\sum_{t=1}^{T} z_{t-1}(z_{t-1}'\beta a' + \varepsilon_t'))\Omega^{-1}a(a'\Omega^{-1}a)^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} I_r \\ -\beta_2 \end{pmatrix} + \begin{pmatrix} 0 \\ (\sum_{t=1}^{T} z_{t-1}z_{t-1}')_2^{-1}(\sum_{t=1}^{T} z_{t-1}\varepsilon_t')\Omega^{-1}a(a'\Omega^{-1}a)^{-1} \end{pmatrix}$$

where $(\sum_{t=1}^{T} z_{t-1}z_{t-1}')_2^{-1}$ indicates the last $(k-r)$ rows of $(\sum_{t=1}^{T} z_{t-1}z_{t-1}')^{-1}$ and $\hat{a}$ is a consistent estimator of a such that the difference between $\hat{a}$ and a will only affects orders of convergence exceeding T. Furthermore $\hat{a} = (\sum_{t=1}^{T} z_{t-1}z_{t-1}')_1^{-1}(\sum_{t=1}^{T} z_{t-1}\Delta y_t')$, where $(\sum_{t=1}^{T} z_{t-1}z_{t-1}')_1^{-1}$ indicates the first r rows of $(\sum_{t=1}^{T} z_{t-1}z_{t-1}')^{-1}$.

To analyse the limiting behavior of $\hat{\beta}$, we have to determine the limiting expressions of both $(\sum_{t=1}^{T} z_{t-1}z_{t-1}')_2^{-1}$ and $(\sum_{t=1}^{T} z_{t-1}\varepsilon_t')\Omega^{-1}a(a'\Omega^{-1}a)^{-1}$. Starting with the latter expression, its limiting behavior can be analyzed using the stochastic trend specification of z_{t-1}.

$$\begin{aligned}(\sum_{t=1}^{T} z_{t-1}\varepsilon_t')\Omega^{-1}a(a'\Omega^{-1}a)^{-1} &= (\sum_{t=1}^{T}\beta_{\perp}(a_{\perp}'\beta_{\perp})^{-1}a_{\perp}'\Omega^{\frac{1}{2}}(\sum_{i=1}^{t-1}\eta_i)\eta_t'\Omega^{-\frac{1}{2}}a \\ &+ \sum_{t=1}^{T} C^*(L)\Omega^{\frac{1}{2}}\eta_{t-1}\eta_t'\Omega^{-\frac{1}{2}}a)(a'\Omega^{-1}a)^{-1}\end{aligned}$$

Since $\Omega^{\frac{1}{2}}a_{\perp}$ is orthogonal to $\Omega^{-\frac{1}{2}}a$, i.e. $(\Omega^{\frac{1}{2}}a_{\perp})'\Omega^{-\frac{1}{2}}a = a_{\perp}'a = 0$, the brownian motions appearing in the limiting expression are independent,

$$\frac{1}{T}\sum_{t=1}^{T} a_{\perp}'\Omega^{\frac{1}{2}}(\sum_{i=1}^{t-1}\eta_i)\eta_t'\Omega^{-\frac{1}{2}}a \Rightarrow \sigma_1\int W_1 dW_2'\sigma_2,$$

since $\frac{1}{\sqrt{T}}a_{\perp}'\Omega^{\frac{1}{2}}(\sum_{i=1}^{T-1}\eta_i) \Rightarrow \sigma_1 W_1$, W_1 is a $(k-r)$ dimensional brownian motion with covariance matrix I_{k-r} and $\sigma_1 = (a_{\perp}'\Omega a_{\perp})^{\frac{1}{2}}$, W_2 is a r dimensional brownian motion with covariance matrix I_r and W_2 is stochastically independent of W_1, $\sigma_2 = (a'\Omega^{-1}a)^{\frac{1}{2}}$.

Also the limiting behavior of $(\sum_{t=1}^{T} z_{t-1}z_{t-1}')^{-1}$ is determined by the stochastic trend specification.

$$(\sum_{t=1}^{T} z_{t-1}z_{t-1}')^{-1} = \begin{pmatrix} \beta & \beta_{\perp} \end{pmatrix}[\begin{pmatrix} \beta & \beta_{\perp} \end{pmatrix}'(\sum_{t=1}^{T} z_{t-1}z_{t-1}')\begin{pmatrix} \beta & \beta_{\perp} \end{pmatrix}]^{-1}\begin{pmatrix} \beta & \beta_{\perp} \end{pmatrix}'$$

So, the limiting behavior of $\begin{pmatrix} \beta & \beta_{\perp} \end{pmatrix}'(\sum_{t=1}^{T} z_{t-1}z_{t-1}')\begin{pmatrix} \beta & \beta_{\perp} \end{pmatrix}$ is,

$$\begin{aligned}&\begin{pmatrix} T^{-\frac{1}{2}}\beta & T^{-1}\beta_{\perp} \end{pmatrix}'(\sum_{t=1}^{T} z_{t-1}z_{t-1}')\begin{pmatrix} T^{-\frac{1}{2}}\beta & T^{-1}\beta_{\perp} \end{pmatrix} \\ \Rightarrow &\begin{pmatrix} var(\beta' z) & 0 \\ 0 & \beta_{\perp}'\beta_{\perp}(a_{\perp}'\beta_{\perp})^{-1}\sigma_1(\int W_1 W_1')\sigma_1'(a_{\perp}'\beta_{\perp})^{-1\prime}\beta_{\perp}'\beta_{\perp} \end{pmatrix}\end{aligned}$$

as

$$T^{-1}\sum_{t=1}^{T}\beta' C^*(L)\Sigma^{\frac{1}{2}}\xi_{t-1}\xi'_{t-1}\Sigma^{\frac{1}{2}\prime}C^{*\prime}(L)\beta \Rightarrow var(\beta' x),$$

$$\begin{aligned}
& T^{-2}\beta'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\sum_{t=1}^{T}(\alpha'_\perp\Sigma^{\frac{1}{2}}\sum_{i=1}^{t-1}\xi_i)(\sum_{i=1}^{t-1}\xi'_i\Sigma^{\frac{1}{2}\prime}\alpha_\perp)(\alpha'_\perp\beta_\perp)^{-1\prime}\beta'_\perp\beta_\perp \\
\Rightarrow\; & \beta'_\perp\beta_\perp(\alpha'_\perp\beta_\perp)^{-1}A_1(\int W_1W'_1)A'_1(\alpha'_\perp\beta_\perp)^{-1\prime}\beta'_\perp\beta_\perp,
\end{aligned}$$

$$T^{-1\frac{1}{2}}\sum_{t=1}^{T}\beta' C^*(L)\Sigma^{\frac{1}{2}}\xi_{t-1}(\sum_{i=1}^{t-1}\xi'_i\Sigma^{\frac{1}{2}\prime}\alpha_\perp)(\alpha'_\perp\beta_\perp)^{-1\prime}\beta'_\perp\beta_\perp \Rightarrow 0.$$

Consequently,

$$\begin{aligned}
& \left[\begin{pmatrix} T^{-\frac{1}{2}}\beta & T^{-1}\beta_\perp \end{pmatrix}'(\sum_{t=1}^{T}x_{t-1}x'_{t-1})\begin{pmatrix} T^{-\frac{1}{2}}\beta & T^{-1}\beta_\perp \end{pmatrix}\right]^{-1} \\
\Rightarrow\; & \begin{pmatrix} var(\beta' x)^{-1} & 0 \\ 0 & (\beta'_\perp\beta_\perp)^{-1}\beta'_\perp\alpha_\perp A_1^{-1\prime}(\int W_1W'_1)^{-1}A_1^{-1}\alpha'_\perp\beta_\perp(\beta'_\perp\beta_\perp)^{-1} \end{pmatrix}
\end{aligned}$$

and

$$\begin{aligned}
& (\sum_{t=1}^{T}x_{t-1}x'_{t-1})^{-1} \\
\Rightarrow\; & O(T)\beta\, var(\beta' x)^{-1}\beta' + \\
& O(T^2)\beta_\perp(\beta'_\perp\beta_\perp)^{-1}\beta'_\perp\alpha_\perp A_1^{-1\prime}(\int W_1W'_1)^{-1}A_1^{-1}\alpha'_\perp\beta_\perp(\beta'_\perp\beta_\perp)^{-1}\beta'_\perp.
\end{aligned}$$

where $O(T^k)$ indicates that the limiting behavior of this part is proportional to T^k. The latter part governs the limiting behavior of $(\sum_{t=1}^{T}x_{t-1}x'_{t-1})_2^{-1}$, which can be characterized by:

$$T^2(\sum_{t=1}^{T}x_{t-1}x'_{t-1})_2^{-1} \Rightarrow (\beta'_\perp\beta_\perp)^{-1}\beta'_\perp\alpha_\perp A_1^{-1\prime}(\int W_1W'_1)^{-1}A_1^{-1}\alpha'_\perp\beta_\perp(\beta'_\perp\beta_\perp)^{-1}\beta'_\perp$$

as $\beta_\perp = \begin{pmatrix} \beta'_2 \\ -I_{k-r} \end{pmatrix}$. So, the limiting expression for the cointegrating vector estimator becomes,

$$\begin{aligned}
T(\hat\beta-\beta) &\Rightarrow \begin{pmatrix} 0 \\ (\beta'_\perp\beta_\perp)^{-1}\beta'_\perp\alpha_\perp A_1^{-1\prime}(\int W_1W'_1)^{-1}(\int W_1 dW'_2)A'_2 \end{pmatrix} \\
&\Rightarrow \begin{pmatrix} 0 \\ N(0, \alpha'\Sigma^{-1}\alpha \otimes \text{[illegible]}) \end{pmatrix},
\end{aligned}$$

where $\Psi = (\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\alpha_\perp b_1^{-1\prime}(\ldots)^{-1}b_1^{-1}\alpha_\perp'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}$ and can be approximated by $(\frac{1}{T^2}\sum_{t=1}^{T} x_{2t-1}(I - x_{1t-1}'(\sum_{t=1}^{T} x_{1t-1}x_{1t-1}')^{-1}x_{1t-1})x_{2t-1}')^{-1}$ $(\Rightarrow \Psi)$.

Proof of theorem 2 (only the second part of theorem 2 is proved).

When the DGP of x_t reads,

$$\Delta x_t = \alpha_\perp\delta' + \alpha(\beta' x_{t-1} + \mu') + \varepsilon_t,$$

where $\mu = \alpha_\perp\delta' + \alpha\mu'$, it has the stochastic trend representation, see [4],

$\Delta x_t = C(L)(\mu + \Sigma^{\frac{1}{2}}\xi_t)$, $\beta = \begin{pmatrix} I_r & -\beta_2' \end{pmatrix}'$,

where ξ_t is a $k-$variate Gaussian white noise process with zero mean and identity covariance matrix. Consequently,

$x_t = \beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'(t\alpha_\perp\delta' + \Sigma^{\frac{1}{2}}\sum_{i=1}^{t}\xi_i) + C^*(1)\alpha\mu' + C^*(L)\Sigma^{\frac{1}{2}}\xi_t$,

$x_{1t} = \beta_2'(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'(t\alpha_\perp\delta' + \Sigma^{\frac{1}{2}}\sum_{i=1}^{t}\xi_i) + \begin{pmatrix} I_r \\ 0 \end{pmatrix}'(C^*(1)\alpha\mu' + C^*(L)\Sigma^{\frac{1}{2}}\xi_t)$,

$x_{2t} = (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'(t\alpha_\perp\delta' + \Sigma^{\frac{1}{2}}\sum_{i=1}^{t}\xi_i) + \begin{pmatrix} 0 \\ I_{k-r} \end{pmatrix}'(C^*(1)\alpha\mu' + C^*(L)\Sigma^{\frac{1}{2}}\xi_t)$,

where $C(L) = C(1) + (1-L)C^*(L)$, $C^*(L) = \sum_{i=0}^{\infty} C_i^* L^i$, $\beta' C^*(1)\alpha = I_r$. The least squares estimator of α, $\hat\alpha$, can also be expressed as

$$\begin{aligned}
&\hat\alpha - \alpha = \\
&(\sum_{t=1}^{T}\varepsilon_t(I - \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}'(\sum_{t=1}^{T}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}')^{-1}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix})x_{1t-1}') \\
&(\sum_{t=1}^{T}x_{1t-1}(I - \begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}'(\sum_{t=1}^{T}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}')^{-1}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix})x_{1t-1}')^{-1} \\
= \;&(\sum_{t=1}^{T}\varepsilon_t(x_{1t-1} - \begin{pmatrix} \hat\beta_2 \\ \hat a \end{pmatrix}'\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix})')(\sum_{t=1}^{T}(x_{1t-1} - \begin{pmatrix} \hat\beta_2 \\ \hat a \end{pmatrix}' \\
&\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix})(x_{1t-1} - \begin{pmatrix} \hat\beta_2 \\ \hat a \end{pmatrix}'\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix})')^{-1}
\end{aligned}$$

with $\begin{pmatrix} \hat\beta_2 \\ \hat a \end{pmatrix} = (\sum_{t=1}^{T}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}')^{-1}\sum_{t=1}^{T}\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix}x_{1t-1}'$. $\hat\beta_2$ is a superconsistent estimator of β_2. Since

$\frac{1}{T}\sum_{t=1}^{T}(x_{1t-1} - \begin{pmatrix} \hat\beta_2 \\ \hat a \end{pmatrix}'\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix})(x_{1t-1} - \begin{pmatrix} \hat\beta_2 \\ \hat a \end{pmatrix}'\begin{pmatrix} x_{2t-1} \\ 1 \end{pmatrix})'$

$\Rightarrow var(\beta' x - \mu') = \beta'\sum_{i=0}^{\infty} C_i^*\Sigma C_i^{*\prime}\beta$, and,

$T^{-\frac{1}{2}}(\sum_{t=1}^{T}\alpha_t(x_{1t-1}-\begin{pmatrix}\hat{\beta}_2\\ \hat{\mu}\end{pmatrix}'\begin{pmatrix}x_{2t-1}\\ 1\end{pmatrix})') \Rightarrow n(0,\ var(\beta' x-\mu')\otimes\Sigma)$, the limiting distribution of $\hat{\alpha}$ becomes

$$\sqrt{T}(\hat{\alpha}-\alpha) \Rightarrow n(0,\ var(\beta' x-\mu')^{-1}\otimes\Sigma).$$

With respect to the cointegrating vector,

$$\begin{aligned}\begin{pmatrix}\hat{\beta}\\ \hat{\mu}\end{pmatrix} &= (\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}')^{-1}(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\Delta x_t')\Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}\\ &= \begin{pmatrix}I_k\\ -\hat{\beta}_2\\ \hat{\mu}\end{pmatrix}\end{aligned}$$

$$\begin{aligned}\begin{pmatrix}-\hat{\beta}_2\\ \hat{\mu}\end{pmatrix} &= ((\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}')^{-1}_2(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}((\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}'\begin{pmatrix}\beta\\ \mu\end{pmatrix}\alpha'\\ &\quad +\delta\alpha_\perp'+\varepsilon_t'))\Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}\\ &= (\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}')^{-1}_2(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\\ &\quad (\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}'\begin{pmatrix}\beta\\ \mu\end{pmatrix}\alpha'+\varepsilon_t'))\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\\ &= \begin{pmatrix}-\beta_2\\ \mu\end{pmatrix}+(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}')^{-1}_2\\ &\quad (\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\varepsilon_t')\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\end{aligned}$$

as $\alpha_\perp'\Sigma^{-1}\alpha = 0$ since $\Sigma^{-1} = P\Lambda P'$, $\Lambda = diag(\lambda_i) = \sum_{i=1}^{k}\lambda_i\iota_i\iota_i'$, $PP' = I_k$, $\alpha_\perp' PP'\alpha = h_\perp' h = 0$, $P'\alpha = h = (h_1' ... h_k')'$, $\alpha_\perp'\Sigma^{-1}\alpha = h_\perp'\Lambda h = \sum_{i=1}^{k}\lambda_i h_{\perp,i}' h_i = 0$ as $h_{\perp,i}' h_i = 0\ \forall i$. $(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}')^{-1}_2$ indicates the last $(k-r+1)$ rows of $(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}')^{-1}$ and $\hat{\alpha}$ is a consistent estimator of α such that the difference between $\hat{\alpha}$ and α will only affect orders of convergence exceeding T. To analyze the limiting behavior of $\hat{\beta}$, we have to determine the limiting expressions of both $(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}')^{-1}_2$ and $(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\varepsilon_t')\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}$. Starting with the latter expression, its limiting behavior can be analyzed using the stochastic trend specification.

$$(\sum_{t=1}^{T}\begin{pmatrix}x_{t-1}\\ 1\end{pmatrix}\varepsilon_t')\Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1}$$

$$= (\sum_{t=1}^{T} \begin{pmatrix} Y_{\perp}(a_{\perp}' Y_{\perp})^{-1} a_{\perp}' (t\mu\lambda' - \Sigma^{\frac{1}{2}}(\sum_{i=1}^{t-1}\zeta_i)) \\ 1 \end{pmatrix} \zeta_t' \Sigma^{-\frac{1}{2}} a$$

$$= \sum_{t=1}^{T} \begin{pmatrix} C^*(1)\mu\lambda' - C^*(L)\Sigma^{\frac{1}{2}}\zeta_{t-1} \\ 0 \end{pmatrix} \zeta_t' \Sigma^{-\frac{1}{2}} a)(a'\Sigma^{-1}a)^{-1}$$

Since $\Sigma^{\frac{1}{2}} a_{\perp}$ is orthogonal to $\Sigma^{-\frac{1}{2}} a$, i.e. $(\Sigma^{\frac{1}{2}} a_{\perp})'\Sigma^{-\frac{1}{2}} a = a_{\perp}' a = 0$, the brownian motions appearing in the limiting expression are independent,

$$\begin{pmatrix} T^{-1}\lambda_{\perp} \\ T^{-\frac{3}{2}}\lambda \end{pmatrix} (a_{\perp}' a_{\perp})^{-1} a_{\perp}' \sum_{t=1}^{T} (t\mu\lambda' - \Sigma^{\frac{1}{2}}(\sum_{i=1}^{t-1}\zeta_i))\zeta_t' \Sigma^{-\frac{1}{2}} a)$$

$$\Rightarrow \Delta_S \int \begin{pmatrix} W_{11} \\ \tau \end{pmatrix} dW_2' \Delta_2'$$

and

$$\begin{pmatrix} T^{-1} I_{k-r-1} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix} Y_{\perp}^{*\prime} \quad (\sum_{t=1}^{T} Y_{\perp}(a_{\perp}' Y_{\perp})^{-1} a_{\perp}' (t\mu\lambda' - \Sigma^{\frac{1}{2}}(\sum_{i=1}^{t-1}\zeta_i))\zeta_t' \Sigma^{-\frac{1}{2}} a)$$

$$\Rightarrow \qquad \Delta_S \int \begin{pmatrix} W_{11} \\ \tau \end{pmatrix} dW_2' \Delta_2'$$

So,

$$\begin{pmatrix} Y_{\perp}^{*\prime} \begin{pmatrix} T^{-1} I_{k-r-1} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix} & 0 \\ 0 & T^{-\frac{1}{2}} \end{pmatrix}' (\sum_{t=1}^{T} \begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix} \varepsilon_t') \Sigma^{-1} a (a'\Sigma^{-1}a)^{-1}$$

$$\Rightarrow \begin{pmatrix} \Delta_S & 0 \\ 0 & 1 \end{pmatrix} \int \begin{pmatrix} W_{11} \\ \tau \\ \iota \end{pmatrix} dW_2' \Delta_2'$$

where $Y_{\perp}^{*} = Y_{\perp}(Y_{\perp}' Y_{\perp})^{-1} Y_{\perp}' a_{\perp} (a_{\perp}' a_{\perp})^{-1} \begin{pmatrix} \lambda_{\perp} \\ \lambda \end{pmatrix}'$, $\lambda_{\perp}\lambda' = 0$, W_{11} is a $(k - r - 1)$ dimensional brownian motion with covariance matrix I_{k-r-1} and $\Delta_S = (\begin{pmatrix} \lambda_{\perp}(a_{\perp}' a_{\perp})^{-1} a_{\perp}' \Sigma^{\frac{1}{2}} \\ \lambda\lambda' \end{pmatrix} \begin{pmatrix} \lambda_{\perp}(a_{\perp}' a_{\perp})^{-1} a_{\perp}' \Sigma^{\frac{1}{2}} \\ \lambda\lambda' \end{pmatrix}')^{\frac{1}{2}}$, W_2 is a r dimensional brownian motion with covariance matrix I_r and W_2 is stochastically independent of W_{11}, $\Delta_2 = (a'\Sigma^{-1}a)^{\frac{1}{2}}$, $\tau(t) = t$, $\iota(t) = 1$, $0 \le t \le 1$.

Also the limiting behavior of $(\sum_{t=1}^{T} \begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix}')^{-1}$ is determined by the stochastic trend specification.

$$(\sum_{t=1}^{T} \begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix} \begin{pmatrix} x_{t-1} \\ 1 \end{pmatrix}')^{-1}$$

$$= \begin{pmatrix} G_1 & G_* & 0 \\ 0 & 0 & 1 \end{pmatrix} \left[\begin{pmatrix} G_1 & G_* & 0 \\ 0 & 0 & 1 \end{pmatrix}' \left(\sum_{t=1}^{T} \begin{pmatrix} z_{t-1} \\ 1 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ 1 \end{pmatrix}'\right) \begin{pmatrix} G_1 & G_* & 0 \\ 0 & 0 & 1 \end{pmatrix} \right]^{-1} \begin{pmatrix} G_1 & G_* & 0 \\ 0 & 0 & 1 \end{pmatrix}'$$

$$\begin{pmatrix} T^{-\frac{1}{2}} G_1 & G_* \begin{pmatrix} T^{-1} I_{k-m-1} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix} & 0 \\ 0 & 0 & T^{-\frac{1}{2}} \end{pmatrix}' \left(\sum_{t=1}^{T} \begin{pmatrix} z_{t-1} \\ 1 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ 1 \end{pmatrix}'\right)$$

$$\begin{pmatrix} T^{-\frac{1}{2}} G_1 & G_* \begin{pmatrix} T^{-1} I_{k-m-1} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix} & 0 \\ 0 & 0 & T^{-\frac{1}{2}} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} var(G_1' z - a') + a'a & 0 & a' \\ 0 & \Delta_B\left(\int \begin{pmatrix} W_{11} \\ r \end{pmatrix} \begin{pmatrix} W_{11} \\ r \end{pmatrix}'\right)\Delta_B' & 0 \\ a & 0 & 1 \end{pmatrix}$$

Consequently,

$$\left(\begin{pmatrix} T^{-\frac{1}{2}} G_1 & \begin{pmatrix} T^{-1} I_{k-m-1} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix} G_* & 0 \\ 0 & 0 & T^{-\frac{1}{2}} \end{pmatrix}' \left(\sum_{t=1}^{T} \begin{pmatrix} z_{t-1} \\ 1 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ 1 \end{pmatrix}'\right)\right.$$

$$\left.\begin{pmatrix} T^{-\frac{1}{2}} G_1 & \begin{pmatrix} T^{-1} I_{k-m-1} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix} G_* & 0 \\ 0 & 0 & T^{-\frac{1}{2}} \end{pmatrix}\right)^{-1}$$

$$\Rightarrow \begin{pmatrix} var(G_1' z - a')^{-1} & 0 & -var(G_1' z - a')^{-1} a' \\ 0 & \Delta_B^{-1\prime}\left(\int \begin{pmatrix} W_{11} \\ r \end{pmatrix} \begin{pmatrix} W_{11} \\ r \end{pmatrix}'\right)^{-1} \Delta_B^{-1} & 0 \\ -a\, var(G_1' z - a')^{-1} & 0 & 1 + a\, var(G_1' z - a')^{-1} a' \end{pmatrix}$$

The limiting behavior of $(\sum_{t=1}^{T} z_{t-1} z_{t-1}')_2^{-1}$ now becomes

$$\begin{pmatrix} G_{2*} \begin{pmatrix} T I_{k-m-1} & 0 \\ 0 & T^{\frac{3}{2}} \end{pmatrix} & 0 \\ 0 & T^{\frac{1}{2}} \end{pmatrix} \left(\sum_{t=1}^{T} \begin{pmatrix} z_{t-1} \\ 1 \end{pmatrix} \begin{pmatrix} z_{t-1} \\ 1 \end{pmatrix}'\right)_2^{-1}$$

$$\begin{pmatrix} G_* \begin{pmatrix} T I_{k-m-1} & 0 \\ 0 & T^{\frac{3}{2}} \end{pmatrix} & 0 \\ 0 & T^{\frac{1}{2}} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \Delta_B & 0 \\ 0 & 1 \end{pmatrix}^{-1\prime} \left(\int \begin{pmatrix} W_{11} \\ r \\ 1 \end{pmatrix} \begin{pmatrix} W_{11} \\ r \\ 1 \end{pmatrix}'\right)^{-1} \begin{pmatrix} \Delta_B & 0 \\ 0 & 1 \end{pmatrix}^{-1}$$

where $\beta_2^* = (Y_\perp' Y_\perp)^{-1} Y_\perp' \alpha_\perp (\alpha_\perp' \alpha_\perp)^{-1} \begin{pmatrix} \lambda_\perp \\ \lambda \end{pmatrix}'$. So, the limiting expression for the cointegrating vector estimator becomes,

$$
\begin{aligned}
& \left(\begin{pmatrix} T I_{k-r-1} & 0 \\ 0 & T^{\frac{3}{2}} \end{pmatrix} \begin{pmatrix} (\beta_2^{*\prime} \beta_2^{*})^{-1} \beta_2^{*\prime} & 0 \\ 0 & T^{\frac{1}{2}} \end{pmatrix} \right) \begin{pmatrix} \beta_2 - \hat{\beta}_2 \\ \hat{\gamma} - \gamma \end{pmatrix} \\
\Rightarrow & \begin{pmatrix} \delta_2 & 0 \\ 0 & 1 \end{pmatrix}^{-1\prime} \left(\int \begin{pmatrix} W_{11} \\ \tau \\ 1 \end{pmatrix} \begin{pmatrix} W_{11} \\ \tau \\ 1 \end{pmatrix}' \right)^{-1} \int \begin{pmatrix} W_{11} \\ \tau \\ 1 \end{pmatrix} dW_2' \delta_2' \\
\Rightarrow & N(0, \alpha' \Sigma^{-1} \alpha \otimes \Omega_2)
\end{aligned}
$$

where $\Omega_2 = \begin{pmatrix} \delta_2 & 0 \\ 0 & 1 \end{pmatrix}^{-1\prime} \left(\int \begin{pmatrix} W_{11} \\ \tau \\ 1 \end{pmatrix} \begin{pmatrix} W_{11} \\ \tau \\ 1 \end{pmatrix}' \right)^{-1} \begin{pmatrix} \delta_2 & 0 \\ 0 & 1 \end{pmatrix}^{-1}$.

Proof of theorem 5 (only the first part is proved, the other proofs are similar).

The optimal value of the SSR objective function reads

$$
\begin{aligned}
S(\hat{\alpha}, \hat{\beta}) = & \; vec(\sum_{t=1}^{T} \varepsilon_t x_{t-1}')' ((\sum_{t=1}^{T} x_{t-1} x_{t-1}')^{-1} \otimes \Sigma^{-1}) vec(\sum_{t=1}^{T} \varepsilon_t x_{t-1}') \\
= & \; vec(\sum_{t=1}^{T} (\Delta y_t - \hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}\hat{\alpha}'\Sigma^{-1}(\sum_{t=1}^{T} \Delta y_t x_{t-1}')(\sum_{t=1}^{T} x_{t-1} x_{t-1}')^{-1} \\
& \sum_{t=1}^{T} x_{t-1} x_{t-1}')' ((\sum_{t=1}^{T} x_{t-1} x_{t-1}')^{-1} \otimes \Sigma^{-1}) \\
& vec(\sum_{t=1}^{T} (\Delta y_t - \hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}\hat{\alpha}'\Sigma^{-1}(\sum_{t=1}^{T} \Delta y_t x_{t-1}')(\sum_{t=1}^{T} x_{t-1} x_{t-1}')^{-1} \\
& \sum_{t=1}^{T} x_{t-1} x_{t-1}') \\
= & \; vec((\Sigma^{-1} - \Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}\hat{\alpha}'\Sigma^{-1})(\sum_{t=1}^{T} \Delta y_t x_{t-1}'))' \\
& ((\sum_{t=1}^{T} x_{t-1} x_{t-1}')^{-1} \otimes \Sigma) vec((\Sigma^{-1} - \Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}\hat{\alpha}'\Sigma^{-1}) \\
& (\sum_{t=1}^{T} \Delta y_t x_{t-1}'))
\end{aligned}
$$

This functional consists of two parts, $((\sum_{t=1}^{T} x_{t-1} x_{t-1}')^{-1} \otimes \Sigma)$ and $vec((\Sigma^{-1} -$

$\Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}\hat{\alpha}'\Sigma^{-1})(\sum_{t=1}^{T}\Delta x_t x_{t-1}'))$, each of which limiting behavior is analyzed separately. Starting with the latter expression,

$$
\begin{aligned}
& \frac{1}{T} vec((\Sigma^{-1} - \Sigma^{-1}\hat{\alpha}(\hat{\alpha}'\Sigma^{-1}\hat{\alpha})^{-1}\hat{\alpha}'\Sigma^{-1})(\sum_{t=1}^{T}\Delta x_t x_{t-1}')) \\
= & \frac{1}{T} vec((\hat{\alpha}_\perp(\hat{\alpha}_\perp'\Sigma\hat{\alpha}_\perp)^{-1}\hat{\alpha}_\perp'(\sum_{t=1}^{T}\Delta x_t x_{t-1}')) \\
\Rightarrow & \frac{1}{T} vec(\alpha_\perp(\alpha_\perp'\Sigma\alpha_\perp)^{-1}\sum_{t=1}^{T}(\alpha_\perp'\Sigma^{\frac{1}{2}}\varepsilon_t)(\sum_{j=1}^{t-1}\varepsilon_j'\Sigma^{\frac{1}{2}}\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp')) \\
\Rightarrow & (\beta_\perp(\alpha_\perp'\beta_\perp)^{-1} \otimes \alpha_\perp(\alpha_\perp'\Sigma\alpha_\perp)^{-1}) vec(C_1(\int W_1 dW_1')'C_1')
\end{aligned}
$$

While

$$
\begin{aligned}
& T^2((\sum_{t=1}^{T}x_{t-1}x_{t-1}')^{-1} \otimes \Sigma) \\
\Rightarrow & (\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\alpha_\perp C_1^{-1\prime}(\int W_1 W_1')^{-1}C_1^{-1}\alpha_\perp'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' \otimes \Sigma).
\end{aligned}
$$

So,

$$
\begin{aligned}
& W(\hat{\alpha}, \hat{\beta}) \\
\Rightarrow & vec(C_1(\int W_1 dW_1')'C_1')'(\beta_\perp(\alpha_\perp'\beta_\perp)^{-1} \otimes \alpha_\perp(\alpha_\perp'\Sigma\alpha_\perp)^{-1})' \\
& (\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\alpha_\perp C_1^{-1\prime}(\int W_1 W_1')^{-1}C_1^{-1}\alpha_\perp'\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' \otimes \Sigma) \\
& (\beta_\perp(\alpha_\perp'\beta_\perp)^{-1} \otimes \alpha_\perp(\alpha_\perp'\Sigma\alpha_\perp)^{-1}) vec(C_1(\int W_1 dW_1')'C_1') \\
\Rightarrow & vec(C_1(\int W_1 dW_1')'C_1')'(C_1^{-1\prime}(\int W_1 W_1')^{-1}C_1^{-1} \otimes (\alpha_\perp'\Sigma\alpha_\perp)^{-1}) \\
& vec(C_1(\int W_1 dW_1')'C_1') \\
\Rightarrow & vec((\int W_1 dW_1')')'((\int W_1 W_1')^{-1} \otimes I_{k-r}) vec((\int W_1 dW_1')') \\
\Rightarrow & tr[(\int W_1 dW_1')'(\int W_1 W_1')^{-1}(\int W_1 dW_1')]
\end{aligned}
$$

References

[1] Engle, R.F., and C.W.J. Granger, 1987, Co-integration and Error Correction : Representation, Estimation and Testing, *Econometrica*, 55, 251-276

[2] Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey

[3] Hansen, L.P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029-1054

[4] Johansen, S., 1991, Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 59, 1551-1580

[5] Kleibergen, F., and H.K. van Dijk, 1994, Direct Cointegration Testing in Error Correction Models, *Journal of Econometrics*, 40, 63-86

[6] Kleibergen, F. and H.K. van Dijk, 1994, On the Shape of the Likelihood/Posterior in Cointegration Models, *Econometric Theory*, 10, 514-551

[7] Kleibergen, F., and P.H. Franses, *Direct Cointegration Testing in Periodic Vector Autoregressive Models*, Tinbergen Institute Research papers, 95-178

[8] Newey, W.K., and K.D. West, 1987, A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 703-708

[9] Perron, P., 1989, The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica*, 57, 1361-1401

[10] Phillips, P.C.B., 1983, Exact Small Sample Theory in the Simultaneous Equations Model, in: Z. Griliches and M.D. Intrilligator, eds., *Handbook of Econometrics*, Vol. I, North-Holland Publishing Co., Amsterdam

[11] Phillips, P.C.B., 1987, Time Series Regression with a Unit Root, *Econometrica*, 55, 277-301

[12] Phillips, P.C.B., 1991, Optimal Inference in Cointegrated Systems, *Econometrica*, 59, 283-307

[13] Phillips, P.C.B., 1994, Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating Coefficients in Error Correction Models, *Econometrica*, 62, 73-93

[14] Phillips, P.C.B., and P. Perron, 1988, Testing for a Unit Root in Time Series Regression, *Biometrika*, 75, 335-346

[15] Phillips, P.C.B., and V. Solo, 1992, Asymptotics for Linear Processes, *Annals of Statistics*, 20, 971-1001

[16] Quintos, C., 1994, *Analysis of Cointegrating Vectors using the GMM approach*, Unpublished Manuscript, Washington University

[17] White, H., 1980, A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica*, 48, 817-838