Chapter 8

Cointegration Testing Using an Outlier Robust Wald Test

This chapter considers the same problem as Chapter 7, namely that of multivariate unit root testing or cointegration testing. There are two important differences between the present and the previous chapter. First, both chapters use a different testing principle. In Chapter 7, a robust alternative for the Gaussian pseudo likelihood ratio test of Johansen (1988) was developed, whereas an outlier robust Wald test is considered in the present chapter. A second important difference concerns the presence of deterministic drift terms in both the data generating process and the regression model. The effect of such terms was only briefly touched upon in Section 7.7. In this chapter much more emphasis is placed upon the consequences of drift terms and deterministic regressors.

As in Chapter 7, the asymptotic distribution of the cointegration test is derived. Moreover, some of the finite sample properties of the test are studied by means of a Monte-Carlo simulation experiment.

As already mentioned in Section 7.7, nonzero drift terms in the data generating process complicate the derivation of the asymptotic distributions of cointegration tests. The relevance of this statement will become apparent in the present chapter. The notation is sometimes cumbersome, while the separate treatment of all different possibilities for the data generating process and the regression models gives the material a messy appearance. This, however, seems to be unavoidable.

The chapter is set up as follows. In Section 8.1, the model and the testing procedure of Kleibergen and van Dijk (1994) are introduced (see also Kleibergen (1994)). An obvious obvious modification of the original testing procedure is discussed, which makes the test resistant to outliers. Moreover, the different sets of interesting null hypotheses are introduced, and attention is devoted to the importance of the deterministic components in the model. Section 8.2 provides the test statistics for each of the hypotheses put forward in Section 8.1. In Section 8.3, the asymptotic distribution of these test statistics is derived. In Section 8.4, the properties of the outlier robust Wald test are evaluated

by means of simulations. In Section 8.5, some brief comments can be found on the importance of the ordering of the variables for the Wald test. Some possibilities for solving the sensitivity of the Wald test with respect to this ordering are proposed. Finally, Section 8.6 concludes this chapter and presents some interesting lines for future research. The appendix contains the proofs of the theorems from the main text. The notational conventions adopted in this chapter were explained in Subsection 1.4.4.

8.1 The Model, the Testing Procedure, and the Hypotheses

This section describes the Wald cointegration test put forward by Kleibergen and van Dijk (1994). It is shown how this test can be generalized in order to yield an outlier robust cointegration test. Finally, the different interesting hypothesis are discussed. Throughout this section, it must be kept in mind that there are two classes of interesting hypotheses. First, one is interested in the number of cointegrating relationships. Second, one is interested in how the deterministic components of the model, like the constant term and the linear time trend, enter the cointegrating relationships. Considering both types of hypotheses leads to a wide variety of test statistics.

8.1.1 The Model

The model used throughout this chapter is similar to (7.10). The only difference is that a linear time trend is added to the specification. More general polynomials in t can also be dealt with, but at the cost of additional complexity and cumbersome notation. Moreover, quadratic and cubic trend functions are often deemed unrealistic from an economic perspective. Therefore, only the case of a linear time trend is considered in this chapter. Assume the data generating process is

$$\Delta y_t = \Pi y_{t-1} + \Psi_1 \Delta y_{t-1} + \ldots + \Psi_p \Delta y_{t-p} + \gamma + \delta t + \varepsilon_t, \tag{8.1}$$

with $y_t \in \mathbb{R}^k$, and

$$|(1-z)I_k - z\Pi - z(1-z)\Psi_1 - \ldots - z^p(1-z)\Psi_{p-1}| = 0 \tag{8.2}$$

implying either $|z| > 1$ or $z = 1$. Moreover, assume that the elements of y_t are at most integrated of order one. The vector y_t is observed for $t = -p, \ldots, T$ and all inference procedures are conducted conditional on the values of y_{-p} through y_0.

As was explained in Section 7.1, the rank of the matrix Π in (8.1) corresponds to the number of cointegrating relationships. Under the hypothesis that $\text{rank}(\Pi) = r$, the matrix Π can be decomposed as $\Pi = AB^\top$, with A and B two $(k \times r)$ matrices of full column rank. Note that in contrast to Chapter

7, $B^\top y_t$ now no longer needs to be stationary in the sense that the mean of $B^\top y_t$ does not depend upon t. This is a consequence of the presence of the linear trend in the data generating process (8.1).

The interpretation of the deterministic components in the model, i.e., the constant and the trend, heavily depend upon the rank of the matrix Π. Consider the two extreme cases of Π having full rank and Π being equal to zero. If Π has full rank, then y_t is stationary around a linear time trend. In contrast, if $\Pi = 0$, then (8.1) indicates that Δy_t is stationary around a linear trend, implying that y_t is the sum of a random walk process and a *quadratic* time trend. The different interpretations of the deterministic part of the model under the null hypothesis of no cointegration and under the alternative of stationarity should be kept in mind when discussing the different test statistics. If one wants the deterministic part of the model to have the same interpretation under the null and under the alternative hypothesis, additional restrictions must be imposed on the coefficients of the trend function (see the discussion of the test statistics in Subsection 8.1.3, below). These restrictions can also be tested. More on the interpretation of deterministic components in models like (8.1) can be found in Johansen (1994).

8.1.2 The Testing Procedure

In order to test for the number of cointegrating relationships, Kleibergen and van Dijk (1994) consider the LU-decomposition of the matrix Π. Under the hypothesis that $\text{rank}(\Pi) = r$, they set

$$\Pi = \alpha\beta^\top = \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} I_r & -\beta_2^\top \\ 0 & I_{k-r} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & -\alpha_{11}\beta_2^\top \\ \alpha_{21} & -\alpha_{21}\beta_2^\top + \alpha_{22} \end{pmatrix}, \quad (8.3)$$

with β_2 a $((k-r) \times r)$ matrix, α_{11} an $(r \times r)$ matrix of full rank, α_{21} a $((k-r) \times r)$ matrix, and α_{22} a $((k-r) \times (k-r))$ matrix. Note the similarity between the decomposition in (8.3) and the one used in the proof of Theorem 7.1 in Appendix 7.A. The total number of parameters in α_{11}, α_{21}, α_{22}, and β_2 equals the number of elements in the matrix Π. A nice feature of the decomposition in (8.3) is that it allows for an explicit parameterization of the hypotheses of interest (see also further below in this subsection). Using (8.3), the cointegration hypothesis can be tested using linear restrictions and ordinary regression techniques.

A decomposition of the type (8.3) is not always possible. First, the nonsingularity assumption for α_{11} may require a reordering of the elements of y_t. The nonsingularity of α_{11} plays a crucial role in the derivation of the asymptotic distribution of the Wald test for cointegration. This is undesirable, as one would, in general, like a testing procedure to be invariant under such simple transformations as reordering the variables. The case $|\alpha_{11}| = 0$ is discussed in more detail in Section 8.5.

A second problem with the decomposition in (8.3) is the presence of the identity matrix I_r in the matrix β. The normalization of the upper-left block

of β to be the unit matrix, may be invalid. As a simple example, consider the case where $k = 2$, $r = 1$, and

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} 0 & -0.5 \\ 0 & -0.5 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}.$$

An LU-decomposition of the form (8.3) is now impossible, as such a decomposition would have to satisfy $\alpha_{11} = 0$ and $\alpha_{11}\beta_2 = 0.5$. This defect can again be taken care of by reordering the elements of y_t.

Using the decomposition in (8.3) and the assumption that $|\alpha_{11}| \neq 0$, rank deficiency of the matrix Π corresponds to a rank deficiency of α_{22}. In particular, $\text{rank}(\Pi) = r$ if and only if $\alpha_{22} = 0$. For $\alpha_{22} = 0$, the cointegrating vectors are given by the columns of $(I_r, -\beta_2^\top)^\top$. Therefore, one can test for the existence of r linearly independent cointegrating relationships by testing whether $\alpha_{22} = 0$. This idea underlies the testing procedure of Kleibergen and van Dijk (1994).

Kleibergen and van Dijk propose the following two-step procedure. Define $\Gamma = (\Psi_1, \ldots, \Psi_p)$, $Z_t^\top = (\Delta X_{t-1}^\top, \ldots, \Delta X_{t-p}^\top)$, and let Γ_1, y_{1t}, Π_1, γ_1, δ_1, and ε_{1t} denote the first r rows of Γ, y_t, Π, γ, δ, and ε_t, respectively. Analoguously, Γ_2 denotes the last $(k - r)$ rows of Γ. Similar definitions hold for the other quantities. Using these definitions and the decomposition presented in (8.3), (8.1) can be split into the two parts, namely

$$\begin{aligned} \Delta y_{1t} &= \Gamma_1 Z_t + \gamma_1 + \delta_1 t + \Pi_1 y_{t-1} + \varepsilon_{1t} \\ &= \Gamma_1 Z_t + \alpha_{11}(B^\top y_{t-1} + \alpha_{11}^{-1}\gamma_1 + \alpha_{11}^{-1}\delta_1 t) + \varepsilon_{1t} \\ &= \Gamma_1 Z_t + \alpha_{11}(B^\top y_{t-1} - \tilde{\gamma}_1 - \tilde{\delta}_1 t) + \varepsilon_{1t}, \end{aligned} \tag{8.4}$$

and

$$\begin{aligned} \Delta y_{2t} &= \Gamma_2 Z_t + \gamma_2 + \delta_2 t + \Pi_2 y_{t-1} + \varepsilon_{2t} \\ &= \Gamma_2 Z_t + \alpha_{21}(B^\top y_{t-1} + \alpha_{11}^{-1}(\gamma_1 + \delta_1 t)) + \alpha_{22} y_{2,t-1} + \\ &\qquad A_\perp^\top(\gamma + \delta t) + \varepsilon_{2t} \\ &= \Gamma_2 Z_t + \alpha_{21}(B^\top y_{t-1} - \tilde{\gamma}_1 - \tilde{\delta}_1 t) + \alpha_{22} y_{2,t-1} + \\ &\qquad \tilde{\gamma}_2 + \tilde{\delta}_2 t + \varepsilon_{2t}, \end{aligned} \tag{8.5}$$

with $\tilde{\gamma}_1 = -\alpha_{11}^{-1}\gamma_1$, $\tilde{\delta}_1 = -\alpha_{11}^{-1}\delta_1$, $\tilde{\gamma}_2 = A_\perp^\top\gamma$, $\tilde{\delta}_2 = A_\perp^\top\delta$, $A^\top = (\alpha_{11}^\top, \alpha_{21}^\top)$, $A_\perp^\top = (-\alpha_{21}\alpha_{11}^{-1}, I_{k-r})$, $B^\top = (I_r, -\beta_2^\top)$, and $B_\perp^\top = (\beta_2, I_{k-r})$. Note that $B_\perp^\top B = A_\perp^\top A = 0$. Estimators of all these quantities are denoted by putting a $\hat{}$ above the parameter one is estimating, for example, $\hat{\tilde{\gamma}}_1$ for $-\hat{\alpha}_{11}^{-1}\hat{\gamma}_1$. Kleibergen and van Dijk propose to start computing the test for $\alpha_{22} = 0$ by estimating (8.4). In this way, one obtains an estimate of the cointegrating relationships, say $(\hat{B}^\top y_{t-1} - \hat{\tilde{\gamma}}_1 - \hat{\tilde{\delta}}_1)$, with $\hat{\beta}_2^\top = -\hat{\Pi}_{11}^{-1}\hat{\Pi}_{12}$. This estimate can then be used as a regressor in (8.5) instead of the unobserved quantity $(B^\top y_{t-1} - \tilde{\gamma}_1 - \tilde{\delta}_1 t)$. After estimating (8.5) with the generated regressor $(\hat{B}^\top y_{t-1} - \hat{\tilde{\gamma}}_1 - \hat{\tilde{\delta}}_1)$, the hypothesis $\alpha_{22} = 0$ can be tested using a Wald type test. This procedure has the attractive property that only simple regression techniques are needed.

There are two problems with the approach of Kleibergen and van Dijk (1994). First, the use of a generated regressor in estimating (8.5) creates an errors-in-variables problem. This complicates the derivation of the asymptotic distribution of the test statistics. The effect of this errors-in-variables problem does not vanish for $T \to \infty$. Second, Kleibergen and van Dijk use the OLS estimator for estimating (8.4) and (8.5). As mentioned in previous chapters, the OLS estimator is sensitive to outliers and other departures from the normality assumption. Therefore, it is suggested in this chapter to use the class of M estimators to obtain estimates of the parameters in (8.4) and (8.5). As in the univariate case, these estimators have better robustness properties for a variety of data generating processes. The precise formulations of the test statistics based on M estimators are presented in the next section.

8.1.3 The Hypotheses of Interest

Now turn to the different hypotheses that are of interest when testing for the existence of cointegrating relations. The type and number of deterministic components in both the data generating process and in the chosen regression model are extremely important. Therefore, the following five data generating processes are distinguished:

1. no deterministic regressors in either (8.4) or (8.5) ($\tilde{\gamma}_1 = 0$, $\tilde{\gamma}_2 = 0$, $\tilde{\delta}_1 = 0$, $\tilde{\delta}_2 = 0$);

2. an intercept in (8.4), which enters the cointegrating relationship; no additional intercept in (8.5) ($\tilde{\gamma}_2 = 0$, $\tilde{\delta}_1 = 0$, $\tilde{\delta}_2 = 0$);

3. an intercept in (8.4), which enters the cointegrating relationship; an additional intercept in (8.5) ($\tilde{\delta}_1 = 0$, $\tilde{\delta}_2 = 0$);

4. a linear time trend in (8.4), which enters the cointegrating relationship; an intercept in (8.5) ($\tilde{\delta}_2 = 0$);

5. a linear time trend in (8.4), which enters the cointegrating relationship; an additional linear time trend in (8.5).

Similarly, one can distinguish five different specifications for the regression model when estimating (8.4) and (8.5). This leads to a total of 25 combinations of regression models and data generating processes. The number of possibilities is reduced by focusing on situations where the regression model is correctly specified or over-specified. This leaves 15 combinations of data generating processes and regression models.

By testing whether $\alpha_{22} = 0$, one tests whether the number of cointegrating relationships is at most r. This immediately produces the first set of relevant hypotheses, namely $H_r^{a.b}: \ \alpha_{22} = 0$, where $H_r^{a.b}$ denotes the null hypothesis of rank$(\Pi) \leq r$ under data generating process a, $a = 1, \ldots, 5$, and regression model b, $b = a, \ldots, 5$. The alternative hypothesis is taken to be $H_k^{a.b}$. In order to determine the rank of Π exactly, a sequential testing procedure can be

devised. Starting from $r = 0$, one tests whether $H_r^{a.b}$ is rejected. If it is not rejected, one concludes that one cannot reject the hypothesis of no cointegration. Otherwise, r is increased by one and the whole procedure is repeated again. Finally, if one rejects $H_{k-1}^{a.b}$, one concludes that all the elements of y_t are stationary. This sequential testing procedure follows a different sequence of hypotheses than the procedure proposed by Kleibergen and van Dijk (1994). Kleibergen and van Dijk start with $r = k$ and reduce r by one each time the hypothesis $H_r^{a.b}$ is *not* rejected. The sequence proposed here follows the one that is usually employed for the Johansen likelihood ratio test for cointegration (see Johansen (1988, 1991, 1994)).

As mentioned at the beginning of this section, it is also interesting to know how the deterimistic components enter the model. For example, if $\tilde{\delta}_2 = 0$, the linear trend is only present in (8.4) and not in (8.5). Alternatively, if $\tilde{\delta}_2 \neq 0$, a linear trend is also present in (8.5), implying a quadratic trend in the levels of y_t, at least if $\alpha_{22} = 0$. Therefore, it is also interesting to consider joint hypotheses on the elements of α_{22}, γ_2, and δ_2. For example, if one suspects that the constant term only enters the cointegrating relationship, one would like to test $H_r^{*a.3} : \alpha_{22} = 0 \wedge \gamma_2 = 0$, for $a = 1$ of 2. Similarly, a joint hypothesis can be devised for the number of cointegrating relationships and the presence of a trend outside these relationships, $H_r^{*a.5} : \alpha_{22} = 0 \wedge \delta_2 = 0$, for $a = 1, \ldots, 4$.

In total one now has 21 hypothesis of interest for each value of r. In the next section, the appropriate test statistics are defined for each of these hypotheses. The asymptotic distributions of these statistics are derived in Section 8.3.

8.2 Construction of the Test Statistics

In this section, the test statistics are specified that correspond to the hypotheses $H_r^{a.b}$ and $H_r^{*a.b}$, presented in Subsection 8.1.3. As was mentioned previously, one first has to obtain estimates of the parameters in (8.4). In this chapter, the class of M estimators is used for obtaining the estimates. The first problem that has to be dealt with is to extend the definition of M estimators to the multivariate setting. This turns out to be reasonably straightforward in the present context. Both (8.4) and (8.5) can be viewed as sets of seemingly unrelated regressions (SUR) (see, e.g., Judge et al. (1988)). Each equation of (8.4) has the same set of regressors. The same holds for (8.5). Therefore, following Koenker and Portnoy (1990), the parameters of (8.4) and (8.5) can be estimated by computing the M estimator for each equation separately. No efficiency is lost by ignoring the cross-correlations between the error terms of the different equations.

In order to simplify the notation later on, define the matrices of parameters, Θ_1^b and Θ_2^b, and the vectors of regressors, F_t^b, Y_{1t}^b, and Y_{2t}^b, for the different regression models. The following definitions hold for the matrices containing

the coefficients:

$$\Theta_1^1 = (\Pi_1, \Gamma_1), \quad \Theta_1^2 = \Theta_1^3 = (\Theta_1^1, \gamma_1), \quad \Theta_1^4 = \Theta_1^5 = (\Theta_1^3, \delta_1),$$

and

$$\Theta_2^1 = \Theta_2^2 = (\alpha_{21}, \alpha_{22}, \Gamma_2), \quad \Theta_2^3 = \Theta_2^4 = (\Theta_2^1, \tilde{\gamma}_2), \quad \Theta_2^5 = (\Theta_2^3, \tilde{\delta}_1).$$

The regressors of the first step regression are given by

$$Y_{1t}^1 = (y_{t-1}^\top, Z_t^\top)^\top, \quad Y_{1t}^2 = Y_{1t}^3 = ((Y_{1t}^1)^\top, 1)^\top, \quad Y_{1t}^4 = Y_{1t}^5 = ((Y_{1t}^3)^\top, t)^\top.$$

In order to define the regressors of the second step regression, an estimate of the cointegrating relationship is required. Let F_t^b be the cointegrating relationship for regression model b:

$$F_t^1 = y_{1,t-1} - \beta_2^\top y_{2,t-1}, \quad F_t^2 = F_t^3 = F_t^1 - \tilde{\gamma}_1, \quad F_t^4 = F_t^5 = F_t^3 - \tilde{\delta}_1 t.$$

The second step regressors are now given by

$$Y_{2t}^1 = ((F_t^1)^\top, y_{2,t-1}^\top, Z_t^\top)^\top, \quad Y_{2t}^2 = ((F_t^2)^\top, y_{2,t-1}^\top, Z_t^\top)^\top,$$

$$Y_{2t}^3 = ((F_t^3)^\top, y_{2,t-1}^\top, Z_t^\top, 1)^\top, \quad Y_{2t}^4 = ((F_t^4)^\top, y_{2,t-1}^\top, Z_t^\top, 1)^\top,$$

and

$$Y_{2t}^5 = ((F_t^5)^\top, y_{2,t-1}^\top, Z_t^\top, 1, t)^\top.$$

If the parameters in the definition of F_t^b are replaced by their estimates, F_t^b is replaced by $\hat{F}_t^b$. Similarly, if F_t^b in the definition of Y_{2t}^b is replaced by $\hat{F}_t^b$, Y_{2t}^b is replaced by $\hat{Y}_{2t}^b$.

Let $\phi_n(\cdot)$, $n = 1, \ldots, k$, be a function that is continuously differentiable. Furthermore, let $\hat{\phi}_{nt}^b$ be equal to $\phi_n(\hat{\varepsilon}_{n1t}^b)$, with $\hat{\varepsilon}_{n1t}^b$ the nth row of $\hat{\varepsilon}_{1t}^b = \Delta y_{1t} - \hat{\Theta}_1^b Y_{1t}^b$. Then the M estimator of the parameters in (8.4) solves

$$\sum_{t=1}^{T} Y_{1t}^b \hat{\phi}_{nt}^b = 0, \qquad n = 1, \ldots, r, \tag{8.6}$$

with respect to $\hat{\Theta}_1^b$. The parameter estimates of the first set of equations can be used to construct an estimate of the cointegrating relationship, $\hat{F}_t^b$. Estimators of the parameters needed to construct $\hat{F}_t^b$, can be obtained by exploiting the relations in (8.3). The estimator of β_2 equals $-\hat{\Pi}_{11}^{-1}\hat{\Pi}_{12}$, with $\hat{\Pi}_{11}$ and $\hat{\Pi}_{12}$ the first r, respectively, the last $(k-r)$, columns of the M estimator for Π_1. The estimators for $\tilde{\gamma}_1$ and $\tilde{\delta}_1$ follow from (8.3) and (8.4), namely $\hat{\tilde{\gamma}}_1 = -\hat{\Pi}_{11}^{-1}\hat{\gamma}_1$ and $\hat{\tilde{\delta}}_1 = -\hat{\Pi}_{11}^{-1}\hat{\delta}_1$.

The M estimator for the parameters in (8.5) can be defined similarly to (8.6). Let $\hat{\eta}_{nt}^b$ denote the $(n-r)$th row of $\hat{\eta}_t^b = \Delta y_{2t} - \hat{\Theta}_2^b \hat{Y}_{2t}^b$. The M estimator of the parameters in (8.5) then solves

$$\sum_{t=1}^{T} \hat{Y}_{2t}^b \phi_n(\hat{\eta}_{nt}^b) = 0, \qquad n = r+1, \ldots, k, \tag{8.7}$$

with respect to $\hat{\Theta}_2^b$.

The estimates that are obtained by solving (8.7), can be used to test the hypotheses $H_r^{a.b}$ and $H_r^{*a.b}$. In addition to the estimates of the unknown parameters, one needs an estimate of the covariance matrix of $\hat{\Theta}_2^b$. Such an estimate can be used to construct the Wald test. The standard formula for the covariance matrix of an M estimator (see Hampel et al. (1986)) cannot be applied directly to each separate equation of (8.7), because the error terms of the different equations in (8.5) are cross-correlated. This problem is easily solved by considering (8.7) for $n = r+1, \ldots, k$ simultaneously. A second and more difficult problem concerns the presence of a generated regressor, $\hat{F}_t^b$, in (8.7). As mentioned in Section 8.1, F_t^b is replaced by $\hat{F}_t^b$ in order to obtain estimates of the parameters in (8.5). Consequently, ε_{2t} is replaced by $\eta_t = \varepsilon_{2t} - \alpha_{21}(\hat{F}_t^b - F_t^b)$. The fact that $T^{1/2}(\eta_2 - \varepsilon_{2t}) = O_p(1)$, affects the covariance matrix of the M estimator. This phenomenon has already thoroughly been studied in the context of stationary variables (see, e.g., Pagan (1984) and Oxley and McAleer (1993)). The results of Kleibergen and van Dijk (1994) and those in Section 8.3 below, indicate that similar problems arise in the context of nonstationary variables.

Based on the results of Section 8.3, the following covariance matrix estimator is proposed. Let $\hat{\Phi}_t$ be a diagonal[1] matrix with diagonal

$$(\phi_1'(\hat{\varepsilon}_{1,1t}), \ldots, \phi_r'(\hat{\varepsilon}_{r1t}), \phi_{r+1}'(\hat{\eta}_{1t}), \ldots, \phi_k'(\hat{\eta}_{k-r,t})),$$

with ϕ_n' the first order derivative of ϕ_n. Define the matrices $\hat{\Phi} = T^{-1}\sum_{t=1}^T \hat{\Phi}_t$, $\hat{A}_\perp^\top = (-\hat{\alpha}_{21}\hat{\Pi}_{11}^{-1}, I_{k-r})$, and

$$\hat{V} = T^{-1}\hat{A}_\perp^\top \hat{\Phi}^{-1} \sum_{t=1}^T \begin{pmatrix} \phi_1(\hat{\varepsilon}_{1,1t}) \\ \vdots \\ \phi_r(\hat{\varepsilon}_{r1t}) \\ \phi_{r+1}(\hat{\eta}_{1t}) \\ \vdots \\ \phi_k(\hat{\eta}_{k-r,t}) \end{pmatrix} \begin{pmatrix} \phi_1(\hat{\varepsilon}_{1,1t}) \\ \vdots \\ \phi_r(\hat{\varepsilon}_{r1t}) \\ \phi_{r+1}(\hat{\eta}_{1t}) \\ \vdots \\ \phi_k(\hat{\eta}_{k-r,t}) \end{pmatrix}^\top \hat{\Phi}^{-1}\hat{A}_\perp. \tag{8.8}$$

Finally, let vec denote the operator that stacks the columns of a matrix into a vector, and let $\otimes$ denote the Kronecker product. The appropriate scaling matrix for $\mathrm{vec}((\Theta_2^b)^\top)$ is then given by $T^{-1}\hat{V} \otimes (\hat{V}_Y^b)^{-1}$, with $\hat{V}_Y^b = T^{-1}\sum_{t=1}^T \hat{Y}_{2t}^b(\hat{Y}_{2t}^b)^\top$. The fact that there is a generated regressor problem when estimating the second set of equations, requires a slight modification of the usual covariance matrix estimator. Normally, one would expect a covariance

[1] The diagonality of the matrix $\hat{\Phi}_t$ follows from the fact that each of the equations in (8.4) and (8.5) is estimated separately. If, in contrast, an M estimator would be used based on the assumption that the ε_t in (8.1) follow some multivariate distribution, e.g., the Student t, then the matrix Φ_t would no longer be diagonal. This can be compared with the results in Chapter 7, where the matrix C_1 is nondiagonal in general and has a similar interpretation as $T^{-1}\sum_{t=1}^T \hat{\Phi}_t$.

estimate of $T^{-1}\hat{V}_{22} \otimes (\hat{V}_Y^b)^{-1}$, with $\hat{V}_{22}$ defined as $\hat{V}$ in (8.8), only with $\hat{A}_\perp^\top$ replaced by $(0, I_{k-r})$ (see Koenker and Portnoy (1990)). Because of the use of $\hat{F}_t^b$ instead of F_t^b when performing the second-step estimation, the slightly different form of the covariance matrix presented in (8.8) is appropriate.

Using the estimates of the parameters and of the scaling matrix, the different test statistics can be constructed. Let P^b be a selection matrix such that $P_b Y_{2t}^b = y_{2,t-1}$. Furthermore, let P_3^* and P_5^* be such that $P_3^* Y_{2t}^3 = (y_{2,t-1}^\top, 1)^\top$ and $P_5^* Y_{2t}^5 = (y_{2,t-1}^\top, t)^\top$, respectively. The hypothesis $H_r^{a.b}$ can then be tested using the statistic

$$t_w^b = T\text{vec}(\hat{\alpha}_{22}^\top)^\top (\hat{V} \otimes P_b (V_Y^b)^{-1} P_b^\top)^{-1} \text{vec}(\hat{\alpha}_{22}^\top). \tag{8.9}$$

The asymptotic distribution of this test statistic is derived in Section 8.3 and can be expressed as a functional of Brownian motions. The asymptotic distribution depends both on the data generating process, a, and on the regression model, b.

The hypothesis $H_r^{*a.b}$ can be tested by means of the statistic

$$t_w^{*b} = T\text{vec}(P_b^* (\hat{\Theta}_2^b)^\top)^\top (\hat{V} \otimes P_b^* (V_Y^b)^{-1} (P_b^*)^\top)^{-1} \text{vec}(P_b^* (\hat{\Theta}_2^b)^\top). \tag{8.10}$$

The asymptotic distribution of this statistic is also presented in Section 8.3 and can again be expressed as a functional of Brownian motions.

8.3 Asymptotic Distribution of the Wald Test

In this section the asymptotic distributions are derived of the test statistics t_w^b and t_w^{*b}, presented in (8.9) and (8.10), respectively. It is shown how these distributions depend upon the data generating mechanism and upon the chosen regression model. Moreover, the relationship between the cointegration test based on the OLS estimator and on an M estimator is investigated. The proofs of all theorems can be found in Appendix 8.A.

Apart from the restrictions on the data generating process that were mentioned in Section 8.1, it is assumed throughout this section that the following conditions are satisfied.

Assumption 8.1 *(i) The $\{\varepsilon_t\}$ process is independently and identically distributed (i.i.d.) with $E(\varepsilon_t) = 0$ and $0 < |E(\varepsilon_t \varepsilon_t^\top)| < \infty$; (ii) for each $n = 1, \ldots, k$, the function $\phi_n(\cdot)$ is differentiable and has a Lipschitz continuous derivative $\phi_n'(\cdot)$; $\phi_n(\cdot)$ is not identically equal to zero; (iii) $\phi_n(u)/u$ is bounded; (iv) $E(\phi_n(\varepsilon_{nt})) = 0$; (v) the decomposition (8.3) is valid with $|\alpha_{11}| \neq 0$.*

Part (i) of Assumption 8.1 is needed for the application of a functional central limit theorem. Both the 'identically' and the 'independently' part of the i.i.d. assumption can be relaxed at the cost of additional complexity. Part (ii) imposes some common smoothness conditions on the functions ϕ_n. These conditions are also found in Knight (1989, 1991) and Chapters 5 and 6. The

asymptotic theory continues to hold for non-smooth versions of ϕ_n if the density of the innovations exists and is smooth enough (compare the remarks below Assumption 7.3). Part (iii) restricts the attention to M estimators that are at most as sensitive to outlying observations as the OLS estimator. Part (iv) imposes a second centering condition on the innovations. A sufficient condition for this restriction to be satisfied is that the density of the innovations is symmetric and that the functions ϕ_n are odd. The final part of Assumption 8.1 requires that the parametric decomposition of the matrix Π is valid. The condition that $|\alpha_{11}| \neq 0$ is crucial, as was mentioned in Section 8.1. The asymptotic results are changed dramatically if this condition is violated, as is showed by Kleibergen and van Dijk (1994). Some results for the case $|\alpha_{11}| = 0$ can be found in Section 8.5.

As a first step towards establishing the asymptotic distribution of the relevant test statistics, I decompose the $\{y_t\}$ process into a deterministic component, a stationary component, and a nonstationary component. The following theorem presents the decomposition.

Theorem 8.1 *Let (8.1) be the data generating process and define* $\Psi = I + \Pi - \sum_{i=1}^{p-1} \Psi_i$. *If Assumption 8.1 is satisfied and if* $\text{rank}(\Pi) = r$, *then*

$$y_t = y_0 + C\sum_{i=1}^{t} \varepsilon_i + C\gamma t + S(1)\delta t + C\delta t(t+1)/2 + S(L)(\varepsilon_t - \varepsilon_0), \qquad (8.11)$$

with $C = B_\perp (A_\perp^\top \Psi B_\perp)^{-1} A_\perp^\top$, *and* $S(L)$ *a polynomial in the lag operator* L, $Ly_t = y_{t-1}$, *such that* $S(L)(\varepsilon_t - \varepsilon_0)$ *is a stationary process.*

Theorem 8.1 facilitates the derivation of the asymptotic distribution of the test statistics. It presents a decomposition of y_t into a linear time trend, a quadratic time trend, a stationary process $y_0 + S(L)(\varepsilon_t - \varepsilon_0)$, and a random walk process, $C\sum_{i=1}^{t} \varepsilon_i$. The decomposition in (8.11) also clearly demonstrates that a linear time trend in (8.1) can result in a quadratic drift for y_t. The presence of a quadratic drift in y_t is absent if δ is restricted to $\delta = A\delta_0$, with δ_0 some column vector of length r. Note that this is equivalent to restricting $\tilde{\delta}_2$ to be zero in (8.5). Similar results hold for the constant term γ, which can transform into a linear time trend (compare Johansen (1994) for further details).

The asymptotic distributions of both t_w^b and t_w^{*b} can be expressed as functionals of Brownian motions. The next theorem states the result for t_w^b.

Theorem 8.2 *Given Assumption 8.1,* $\text{rank}(\Pi) = r$, *and* $\hat{\varepsilon}_{1t}^b - \varepsilon_{1t}^b = o_p(1)$ *and* $\hat{\eta}_t^b - \eta_t^b = o_p(1)$ *uniformly in* t, *then*

$$t_w^b \Rightarrow \text{tr}\left((\int B_1^{a.b} dB_2^\top)^\top (\int B_1^{a.b} (B_1^{a.b})^\top)^{-1} (\int B_1^{a.b} dB_2^\top) \right), \qquad (8.12)$$

with B_1 *and* B_2 *two standard Brownian motions such that*

$$E(B_1(s)B_2(s)^\top) = sR^{1.1},$$

with $R^{1.1}$ a diagonal matrix containing the absolute canonical correlations between $A_\perp^\top \varepsilon_t$ and $A_\perp^\top \Phi^{-1}\phi(\varepsilon_t)$, $\phi(\varepsilon_t) = (\phi_1(\varepsilon_{1,1t}), \ldots, \phi_k(\varepsilon_{k-r,2t}))^\top$,

$$\begin{aligned}
B_1^{1.1}(s) &= B_1(s) \\
B_1^{1.2}(s) &= B_1^{1.1}(s) \\
B_1^{1.3}(s) &= B_1^{1.1}(s) - \textstyle\int B_1^{1.1} \\
B_1^{1.4}(s) &= B_1^{1.3}(s) \\
B_1^{1.5}(s) &= B_1^{1.1}(s) - (4-6s)\textstyle\int B_1^{1.1} - (12s-6)\int sB_1^{1.1} \\
B_1^{2.j}(s) &= B_1^{1.j}(s) \\
B_1^{3.3}(s) &= (s-0.5, (F_1^{3.3}(s) - \textstyle\int F_1^{3.3})^\top)^\top \\
B_1^{3.4}(s) &= B_1^{3.3}(s) \\
B_1^{3.5}(s) &= B_1^{1.5}(s) \\
B_1^{4.4}(s) &= (s-0.5, (F_1^{4.4}(s) - \textstyle\int F_1^{4.4})^\top)^\top \\
B_1^{4.5}(s) &= B_1^{1.5}(s) \\
B_1^{5.5}(s) &= F_1^{5.5}(s) - (4-6s)\textstyle\int F_1^{5.5} - (12s-6)\int sF_1^{5.5},
\end{aligned}$$

for $j = 2,3,4,5$, and $F^{i.i}$ for $i = 3,4,5$ standard Brownian motions such that $E(F^{i.i}(s)B_2(s)^\top) = sR^{i.i}$, with $R^{i.i}$ a diagonal matrix[2] containing the absolute canonical correlations between $A_\perp^\top \Phi^{-1}\phi(\varepsilon_t)$ and $\zeta_{1\perp}^\top B_\perp^\top C\varepsilon_t$, $\zeta_{4\perp}^\top B_\perp^\top C\varepsilon_t$, and $\zeta_{3\perp}^\top B_\perp^\top C\varepsilon_t$, for $i = 3,4$, and 5, respectively. The matrices $\zeta_{1\perp}$, $\zeta_{3\perp}$, and $\zeta_{4\perp}$ are defined in Appendix 8.A (pages 207 and 211).

Corollary 8.1 *If $\phi(\varepsilon_t) = \varepsilon_t$, then the diagonal elements of $R^{a.b}$ are equal to unity and (8.12) reduces to the expressions found in Kleibergen and van Dijk (1994).*

Kleibergen and van Dijk (1994) derive the limiting distribution of t_w^b for the OLS estimator. Theorem 8.2 presents the generalization of their results for the class of M estimators. The main difference between (8.12) and the result for the OLS estimator is that the Brownian motion B_2 no longer coincides with the Brownian motion B_1. These two Brownian motions are, in general, imperfectly correlated. As in Chapter 7, the correlation depends upon the canonical correlations between certain linear combinations of the errors ε_t and certain linear combinations of the pseudo score $\phi(\varepsilon_t)$.

The results in Theorem 8.2 differ in one important respect from the results in Theorem 7.1 of Section 7.3. Under the null hypothesis, it appears from Theorem 8.2 that the only nuisance parameters entering the limiting distribution of the Wald test are the canonical correlations between $A_\perp^\top \Phi^{-1}\phi(\varepsilon_t)$ and $A_\perp^\top \varepsilon_t$. In contrast, the likelihood ratio test also suffers from the effects of misspecification of the likelihood function. The asymptotic distributions of the likelihood ratio test and the Wald test only coincide if the likelihood is correctly specified

[2]Note that the $R^{i.i}$'s are not square matrices for $i \neq 1$. The diagonal of these matrices has to be interpreted as the set of elements for which the row and column indices coincide.

(see Corollary 7.2) or if the OLS estimator is used (see Corollary 7.1). Similar findings were already established in the stationary setting by White (1982).

In order to perform inference with the statistic t_w^b, critical values are needed. As in Chapter 7, one can think of several strategies. First, one can ignore the imperfect correlation between B_2 and $B_1^{a.b}$ and just use the critical values of the OLS-based test. These are tabulated in, e.g., Osterwald-Lenum (1992). This strategy is only useful if ε_t and $\phi(\varepsilon_t)$ are sufficiently close, or, stated differently, if the diagonal elements of $R^{a.b}$ are sufficiently close to unity. Second, one can develop a similar strategy as in Section 7.5 and construct a corrected test that has the same mean as the OLS-based test. In order to develop such a test, one can replace B_2 by the sum of two stochastic processes, one of which is a linear transformation of the stochastic process in $B_1^{a.b}$, while the other is orthogonal to the stochastic process in $B_1^{a.b}$ (compare Theorem 7.2). Third, one can approximate the distribution of the expression in (8.12) by means of simulation. Using the estimated values of the parameters, one can construct an estimate of $R^{a.b}$. This estimate can then be used to approximate the stochastic integrals in (8.12) by drawing appropriately correlated and standardized random walks and replacing the integrals in (8.12) by averages. This process can be repeated a large number of times in order to obtain accurate approximations to the critical values of the test.

To conclude this section, the asymptotic distributions of the alternative test statistic t_w^{*b} are presented.

Theorem 8.3 *Given the conditions of Theorem 8.2,*

$$t_w^{*b} \Rightarrow \text{tr}\left((\int B_1^{*a.b} dB_2^\top)^\top (\int B_1^{*a.b} (B_1^{*a.b})^\top)^{-1} (\int B_1^{*a.b} dB_2^\top)\right),$$

with

$$\begin{aligned} B_1^{*i.3}(s) &= ((B_1^{1.1}(s))^\top, 1)^\top \\ B_1^{*j.5}(s) &= ((B_1^{1.3}(s))^\top, s - 0.5)^\top, \end{aligned}$$

for $i = 1, 2$ *and* $j = 1, \ldots, 4$.

Again, one can see that by choosing $\phi(\varepsilon_t) = \varepsilon_t$, one obtains the results of Johansen (1991, Theorem 2.2) and Kleibergen and van Dijk (1994, Theorem 4.ii). For $\phi(\varepsilon_t) \neq \varepsilon_t$, the two Brownian motions in the limiting expressions for t_w^{*b} are imperfectly correlated, giving rise to $(k - r)$ nuisance parameters in the asymptotic distribution of the test statistic.

As in Johansen (1991), it is probably straightforward to show that conditional on the number of unit roots in the system, test statistics for conducting inference on the elements of α_{ij} and β_2 have standard limiting distributions. An explicit proof of this statement is not provided, but it follows rather directly by using Lemma 8.9 from Appendix 8.A and the condition that $\alpha_{22} = 0$.

8.4 Simulation Results

In this section,the performance of the outlier robust Wald test for cointegration is discussed using a small simulation experiment. Subsection 8.4.1 discusses the results for the level of the test. Subsection 8.4.2 presents some results concerning the power of the test. Both subsections consider data generating process one ($a = 1$) and regression model five ($b = 5$). So there are no drift terms in the data generating process, while a linear trend is fitted in both (8.4) and (8.5). This departs from the simulation setup presented in Section 7.6, where no deterministic regressors were used.

8.4.1 Level

The level simulations are standard. For several values of $k - r$, a $(k - r)$-dimensional random walk process is generated as $y_t = y_{t-1} + \varepsilon_t$, with $\{\varepsilon_t\}$ and i.i.d. process and $y_0 = 0$. The number of observations is T. For each of these simulated time series, the Wald test statistic corresponding to the hypothesis $H_0^{1.5}$ is computed for several M estimators (see below). These statistics are stored over N replications. In order to estimate the α-quantile, the αNth order statistic is used. The length of the time series and the number of Monte-Carlo replications used in this subsection, are $T = 100$ and $N = 1,000$, respectively. Note that $\text{rank}(\Pi) = r = 0$ if $H_0^{1.5}$ is tested.

In order to illustrate the effects of outliers on the Wald test, I consider two distributions for the error process. The first distribution is the multivariate standard normal, while the second distribution is the truncated Cauchy, which was also used in Section 7.6.

As mentioned in Section 8.2, the parameters in (8.4) and (8.5) can be estimated equation by equation. Therefore, only a specification for the functions $\phi_n(\cdot)$, $n = 1, \ldots, k - r$, in (8.6) and (8.7) is needed. In order to illustrate the properties of the outlier robust Wald test, I use the Student t score function for $\phi_n(\cdot)$, so

$$\phi_n(e) = (\nu + 1)e/(\nu + e^2),$$

for $e \in \mathbb{R}$. Note that this does not coincide with the specification used in Chapter 7. In Chapter 7 a pseudo maximum likelihood estimator was used based on the multivariate Student t distribution. In this chapter, the pseudo likelihood is a product of univariate Student t densities. The M estimator is made scale invariant by setting e equal to the residual of the nth equation, divided by the median absolute deviation of the residuals of the nth equation. The median absolute deviation is again standardized such that it is a consistent estimator of the standard deviation of the Gaussian distribution (see (2.32) and below). The Student t score function depends upon the tuning constant or degrees of freedom parameter ν. The values considered, are $\nu = \infty, 5, 1$, with $\nu = \infty$ corresponding to the OLS-based Wald testing procedure of Kleibergen and van Dijk (1994). The results are presented in Table 8.1 and Figure 8.1.

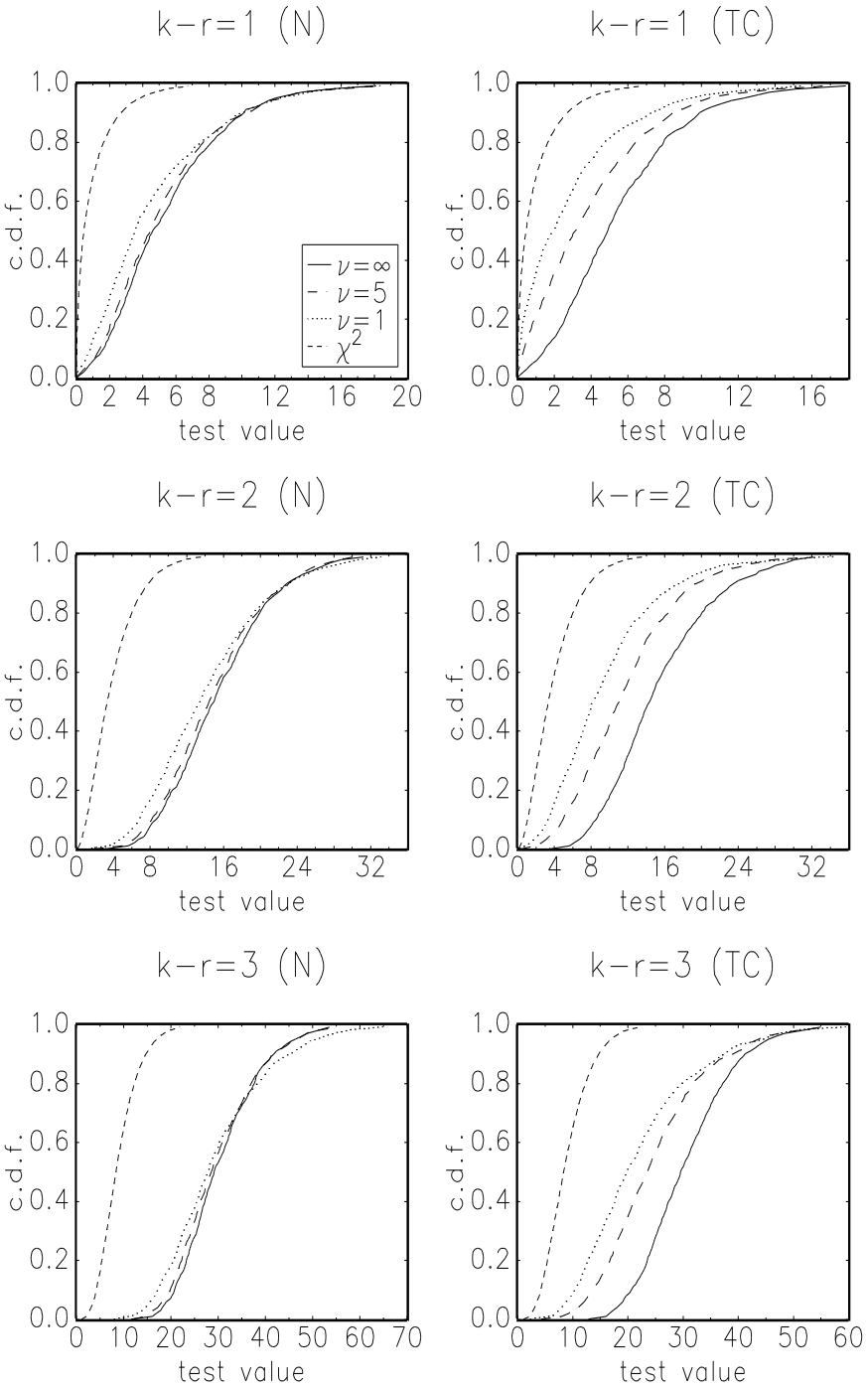

Figure 8.1.— Distributions of the cointegration Wald test for normal (N) and truncated Cauchy (TC) innovations

TABLE 8.1
Quantiles for the Cointegration Wald Test

		quantile					
$k-r$	ν	0.5	0.8	0.9	0.95	0.975	0.99
		Gaussian					
1	∞	4.593	8.120	10.125	11.984	13.987	17.809
1	5	4.490	7.661	10.287	12.446	14.032	18.066
1	1	3.511	7.505	10.131	12.461	14.576	17.443
2	∞	14.973	19.976	22.900	26.126	27.873	31.097
2	5	14.432	19.669	23.098	25.403	28.063	29.820
2	1	13.493	19.266	23.087	26.548	29.957	32.612
3	∞	29.106	37.634	41.781	46.223	49.385	53.102
3	5	28.527	36.907	41.792	46.483	49.280	53.466
3	1	27.877	38.507	45.355	51.148	55.484	63.874
		Truncated Cauchy					
1	∞	4.897	7.929	9.864	12.296	14.030	17.746
1	5	3.148	6.419	8.449	10.422	13.418	16.627
1	1	1.947	4.757	7.356	9.466	13.072	15.404
2	∞	14.322	20.050	23.546	27.240	29.477	32.102
2	5	11.023	16.451	19.325	23.177	26.129	32.155
2	1	8.188	13.701	17.752	21.164	26.104	33.746
3	∞	29.548	37.205	41.207	45.341	50.324	54.613
3	5	23.422	31.729	39.110	43.663	49.230	55.756
3	1	19.865	29.851	37.541	44.364	48.951	58.561

The table contains the αNth order statistics from $N = 1,000$ Monte-Carlo simulations of the Wald cointegration test. $k-r$ denotes the dimension of the time series. Gaussian means that the innovations were drawn from the standard multivariate normal distribution. Truncated Cauchy means that the innovations were drawn from the truncated Cauchy distribution (see Section 7.6). The parameter ν is the tuning constant of the M estimator. The robustness of the estimator decreases for increasing ν.

Both Table 8.1 and the left panels of Figure 8.1 indicate that for Gaussian innovations the c.d.f. of the test statistic for low values of ν lies to the left of the c.d.f. for $\nu = \infty$, at least for the lower quantiles. The ordering is reversed if the higher quantiles of the distribution are considered. Especially for higher values of $k-r$, the 5% critical value of the test for $\nu = 1$ lies to the right of that for $\nu = \infty$. Similar findings were noted in the previous chapters. Also note that the c.d.f.'s of all cointegration tests lie to the right of the c.d.f. of the χ^2 distribution with k^2 degrees of freedom. This phenomenon is already well known in the literature (compare Fuller (1976)).

If one considers the case of truncated Cauchy innovations, one again obtains that the c.d.f.'s of the Wald tests for cointegration lie to the right of the appropriate χ^2 distribution. The c.d.f. for $\nu = 1$, however, now clearly lies to the left of that for $\nu = 5$, which in turn lies to the left of that for $\nu = \infty$. This result was also noted in Section 7.6, where the actual size of the pseudo

likelihood ratio test appeared to be below the nominal level if the innovations were truncated Cauchy.

In order to answer the question whether the c.d.f. of the test for $\nu = 1$ should be to the left of that for $\nu = \infty$ or whether they should intersect, I conducted the following experiment. Instead of directly simulating the distribution of the Wald test, one can also try to approximate the asymptotic distribution of the Wald test by means of simulation. This can be done by replacing the stochastic processes in Theorem 8.2 by discrete realizations and by replacing the integrals by averages. Note that the asymptotic distribution in Theorem 8.2 depends upon the canonical correlations between $B_1(s)$ and $B_2(s)$. Assume for simplicity that $(k-r) = 1$ and that $R^{1.5}$ in Theorem 8.2 is equal to ρ, with $0 \leq \rho \leq 1$. Moreover, let ε_t and η_t denote two independent standard normal random variables. Let $B_{1T}(t/T) = T^{-1/2} \sum_{i=1}^{t} \varepsilon_i$, $t = 1, \ldots, T$. and $B_{3T}(t/T) = T^{-1/2} \sum_{i=1}^{t} \eta_i$, while $B_{1T}^{\tau}(t/T)$ contains the OLS residuals of a regression of B_{1T} on a constant and a trend. Then a drawing from the asymptotic distribution presented in Theorem 8.2 can be approximated by $\mathrm{tr}(C_{1T}^{\top} C_{2T}^{-1} C_{1T})$, with

$$\begin{aligned} C_{1T} &= \sum_{t=1}^{T} B_{1T}^{\tau}((t-1)/T)(B_{2T}(t/T) - B_{2T}((t-1)/T)^{\top}, \\ C_{2T} &= T^{-1} \sum_{t=1}^{T} B_{1T}^{\tau}((t-1)/T) B_{1T}^{\tau}((t-1)/T)^{\top}, \end{aligned}$$

and $B_{2T}(s) = \rho B_{1T}(s) + (1-\rho^2)^{1/2} B_{3T}(s)$. Using $N = 1{,}000$ Monte-Carlo simulations with random walks of length $T = 50$, an approximation to the asymptotic distribution function is constructed. Using larger values of T or higher values of $(k-r)$, yielded qualitatively similar results. The left panel of Figure 8.2 presents the critical values corresponding to several quantiles and several values of the correlation parameter ρ.

The left panel of Figure 8.2 indicates that the critical values corresponding to the α-quantile are increasing functions of ρ for $\alpha \geq 0.8$. In order to get some feeling for this result, consider the two extreme cases $\rho = 1$ and $\rho = 0$. For $\rho = 1$, the asymptotic distribution of the Wald test is just a multivariate generalization of the Dickey-Fuller distribution, yielding critical values that are above the standard χ^2 critical values (compare, e.g., Fuller (1976)). For $\rho = 0$, in contrast, the asymptotic distribution of the Wald test is just a χ^2 distribution with $(k-r)^2$ degrees of freedom. This follows from the independence of B_1 and B_2 for $\rho = 0$.

In order to study the relation between the critical values of the test and the value of the tuning constant ν, a relation between ν and the correlation parameter ρ is needed. The relation between ν and ρ can be easily visualized if the half-line $\nu > \infty$ is mapped onto the unit interval. Therefore, define the function

$$f(\nu) = \begin{cases} \nu/2 & \text{for } 0 < \nu \leq 1 \\ 1 - (2\nu)^{-1} & \text{for } 1 < \nu \end{cases} . \tag{8.13}$$

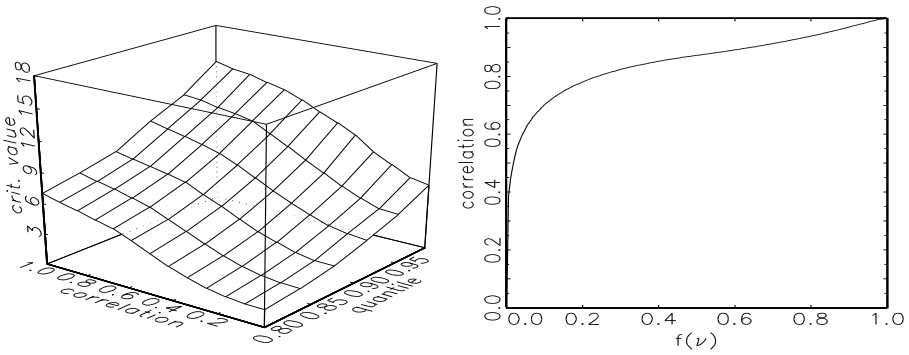

Figure 8.2.— Critical Values for Different Correlations between the M Estimator and the OLS Score (left panel) and the Correlation as a Function of $f(\nu)$, defined in (8.13) (right panel).

Note that $0 < f(\nu) \leq 1$ for $0 < \nu \leq \infty$ and that $f(\cdot) : \mathbb{R}^+ \to (0, 1]$ is increasing and bijective. For each value of ν, the correlation between B_1 and B_2 can be computed. The right panel of Figure 8.2 displays the relation between the correlation ρ and $f(\nu)$. It appears that the correlation is an increasing function of $f(\nu)$ and, therefore, an increasing function of the degrees of freedom parameter ν. Therefore, one would expect the critical values of the Wald test for $\nu = 1$ to lie to the left of those for $\nu = \infty$. This contrasts with the results presented in Table 8.1 and the left panels of Figure 8.1. My conjecture is that this discrepancy is due to the finite sample properties of the test. The Wald test based on M estimators needs larger samples to reach its asymptotic distribution that its OLS-based counterpart. This is clearly illustrated by the simulations in Chapter 6, in particular Table 6.1. For $T = 100$, the 1% critical value of the Wald test (M_2^i for $i = 0, 1, 2$) based on OLS lies closer to zero than the critical value for $\nu = 3$. For $T = 5,000$ the ordering is reversed, as is expected from the asymptotic theory.

Concluding, one can say that the critical values of the Wald cointegration test based on M estimators have as yet to be obtained by means of extensive simulations. The asymptotic distribution theory appears to provide a poor approximation to the finite sample distribution of the test. The accuracy of the approximation to the finite sample distribution by means of the asymptotic distribution seems to worsen if the dimension of the system is increased. Therefore, it seems a useful topic for future research to generalize the results of Abadir and Larson (1994) to non-OLS estimators. If this can be done, valid finite sample critical values can perhaps be calculated without resorting to

time consuming simulations.

8.4.2 Power

As in Chapter 7, the focus is not so much on the absolute power of the Wald cointegration test, but rather on the relative power of the Wald test based on M estimators versus the power of the OLS-based Wald test. Apart from the model presented in this subsection, other experiments were performed using alternative regression models and other dimensions of the system $(k-r)$. These experiments resulted in qualitatively similar conclusions. Therefore, they are not reported.

The data generating mechanism used in this subsection is

$$\Delta y_t = \begin{pmatrix} -c/T & 0 & 0 \\ 0 & -c/(2T) & 0 \\ 0 & 0 & -c/(3T) \end{pmatrix} y_{t-1} + \varepsilon_t, \tag{8.14}$$

with y_t a 3-dimensional vector, $T = 100$, and $c = 0, 1, 5, 10, 20$. As in Subsection 8.4.1, ε_t again follows either a standard Gaussian distribution or a truncated Cauchy distribution (see Section 7.6). The fitted regression model contains an unrestricted linear trend. Therefore, the critical values can be found in Table 8.1. Model (8.14) generates a time series with nearly-nonstationary behavior (compare Section 7.2). If $c = 0$, then (8.14) presents a system of independent random walks. For $0 < c < T$, however, there are three cointegrating relations with loading factors that decrease to zero as the sample size increases. This also appears by looking at the roots of the system, which are $(1 - c/T)^{-1}$, $(1 - c/(2T))^{-1}$, and $(1 - c/(3T))^{-1}$, respectively. So the cointegrating relation corresponding to the smaller roots should be more difficult to detect than the relations corresponding to the larger roots. The innovation process was again either standard Gaussian or truncated Cauchy.

Using (8.14), one can calculate the Wald cointegration test for the values of ν used in Subsection 8.4.1. These can be compared with the critical values from Table 8.1. Table 8.2 presents the number of rejections of the null hypothesis $H_r^{1.5}$ for $r = 0, 1, 2$ at a nominal significance level of 5%.

First, consider the results for Gaussian innovations. The power of the test appears to be very low. Even if the roots of the system are as small as 0.8, the rejection frequencies are at most 28% for $\nu = \infty$ and $\nu = 5$ and at most 19% for $\nu = 1$. This corresponds to the univariate results presented in Section 6.4. The tests seem unable to detect the third cointegrating relationship: the rejection frequency for $H_2^{1.5}$ barely exceeds the nominal level of 5%. Finally, the cointegration tests for $\nu = \infty$ and $\nu = 5$ perform equally well. In contrast, the test for $\nu = 1$ appears to have a much lower power than the test for $\nu = \infty$. This can be expected from the optimality result in Section 7.4: for Gaussian innovations, the OLS estimator usually performs best from a minimum expected mean squared error perspective.

TABLE 8.2
Rejection Frequencies for the Cointegration Wald Test

		Gaussian			Truncated Cauchy		
ν	c	$r=0$	$r=1$	$r=2$	$r=0$	$r=1$	$r=2$
	0	0.028	0.015	0.007	0.058	0.015	0.025
	1	0.048	0.028	0.013	0.045	0.033	0.025
∞	5	0.048	0.018	0.022	0.072	0.035	0.007
	10	0.075	0.060	0.022	0.077	0.050	0.033
	20	0.280	0.110	0.058	0.307	0.120	0.075
	0	0.037	0.013	0.007	0.022	0.007	0.003
	1	0.037	0.033	0.013	0.030	0.010	0.010
5	5	0.058	0.028	0.010	0.090	0.070	0.010
	10	0.090	0.072	0.018	0.338	0.165	0.063
	20	0.285	0.150	0.052	0.925	0.623	0.253
	0	0.040	0.018	0.010	0.003	0.007	0.005
	1	0.028	0.033	0.013	0.010	0.013	0.005
1	5	0.050	0.037	0.013	0.077	0.058	0.020
	10	0.085	0.072	0.015	0.340	0.168	0.070
	20	0.185	0.120	0.070	0.920	0.667	0.313

The table contains the rejection frequencies of the hypothesis $H_r^{1.5}$ for different values of r and different estimators. The nominal size of the test is 5% and the entries are based on 400 Monte-Carlo simulations. The left three columns present the results for model (8.14) and standard Gaussian innovations. The right three columns present the result for truncated Cauchy innovations (see Section 7.6). The M estimators are explained in Subsection 8.4.1.

Now consider the results for the truncated Cauchy innovations. The rejection frequencies of the test for $\nu = \infty$ seem to have increased marginally compared to the case with Gaussian innovations. The overall picture, however, remains the same: the power is low and the third cointegrating relationship is barely detected. For the non-OLS estimators, the rejection frequencies are below the nominal level (compare Section 7.6). For $c \geq 5$ the power of the tests for $\nu = 5$ and $\nu = 1$ exceeds the power of the OLS-based test. The test for $\nu = 1$ appears to have a somewhat higher power against distant alternatives than the test for $\nu = 5$, but the difference is very small. The power of both these tests is quite high. For $c = 20$, the tests based on $\nu = 5$ and $\nu = 1$ find at least one cointegrating relationship in more that 90% of the simulations. Even the third cointegrating relation is detected with a frequency exceeding 25%, a number that almost matches the 28% for $c = 20$, $r = 0$, and Gaussian innovations.

Based on the power simulations of the Wald cointegration test, it seems best to use an M estimator that is reasonably efficient for Gaussian innovations and that, at the same time, possesses some robustness properties. In the simulations, the best candidate meeting these criteria appeared to be the

estimator based on $\nu = 5$. For Gaussian innovations, the performance of the Wald test based on this estimator was approximately the same as that of the OLS-based Wald test. For truncated Cauchy innovations, the test for $\nu = 5$ performed much better than the test for $\nu = \infty$. Moreover, the test for $\nu = 5$ did not perform much worse than the more robust test based on $\nu = 1$.

8.5 Ordering of the Variables

In Sections 8.3 and 8.4, the asymptotic and finite sample properties of the Wald cointegration test were considered for the case $|\alpha_{11}| \neq 0$, where α_{11} was introduced in (8.3). In the present section, the properties of the Wald test are studied for the case $|\alpha_{11}| = 0$. First, the effect of $|\alpha_{11}| = 0$ on the asymptotic distribution of the Wald test is discussed. This is done in Subsection 8.5.1. Then, several possibilities for solving the problems caused by the singularity of α_{11} are reviewed. This is done in Subsections 8.5.2 and 8.5.3. Finally, Subsection 8.5.4 presents the results of a small simulation experiment similar to the one presented in Section 8.4, only with $\alpha_{11} = 0$ instead of $|\alpha_{11}| \neq 0$.

8.5.1 Asymptotic Distribution Theory for $|\alpha_{11}| = 0$

If α_{11} is singular, the asymptotic properties of the estimator for β_2 change quite dramatically. As stated below (8.5), $\beta_2^\top$ is estimated by $-\hat{\Pi}_{11}^{-1}\hat{\Pi}_{12}$. From (8.3) and Lemma 8.4 it follows that $\hat{\Pi}_{11}$ converges to α_{11}. Therefore, if α_{11} is singular, one can expect a change in the convergence behavior of $\hat{\beta}_2$. It turns out that $\hat{\beta}_2$ is still a consistent estimator for β_2, only the rate at which the estimator converges to its limiting value is lower than for the case $|\alpha_{11}| \neq 0$, namely $T^{1/2}$ instead of T. This result is presented in Lemma 8.11 in Appendix 8.A.

The slower convergence rate of $\hat{\beta}_2$ also affects the convergence behavior of the estimators for the parameters of the second-step regression. In general, the rate of convergence of these estimators drops by 1/2 if $|\alpha_{11}| = 0$. This is most easily seen in the last formula in the proof of Lemma 8.9, where $T^{1/2}\hat{\Pi}_{11}$ is now $O_p(1)$ rather than $O_p(T^{1/2})$. Also the limiting distributions of the second-step estimators change dramatically. Consequently, the distribution of the cointegration Wald test is also changed and is no longer given by the result in Theorem 8.2. For the OLS estimator, the limiting distribution of the test for regression model 3, data generating process 3, and $\alpha_{11} = 0$, is given in Theorem 5 of Kleibergen and van Dijk (1995). In this chapter, I refrain from deriving the asymptotic distribution of the Wald test for the case $|\alpha_{11}| = 0$, because the expressions would be very complicated without providing valuable new insights into the properties of the test. It is more important to note that the assumption $|\alpha_{11}| \neq 0$ is crucial for the validity of asymptotic distributions presented in Theorem 8.2. Therefore, one has to make sure that the condition is met when the test is applied to empirical data. Some problems and solutions

associated with checking whether $|\alpha_{11}| \neq 0$ or not, are discussed in the next subsection.

8.5.2 A Pretest for $|\alpha_{11}| = 0$

As was mentioned in Section 8.1, the singularity of α_{11} can often be undone by reordering the variables in the system. In this subsection, a pretesting procedure is discussed for determining whether α_{11} is singular or not. It turns out that this pretest cannot be applied if one does not know the number of unit roots (or the number of cointegrating relations) in the system. As the number of unit roots is exactly the unknown parameter one is trying to determine, it is obvious that the pretest does not provide a solution to the problems caused by the singularity of α_{11}. In the next subsection I discuss some alternative cointegration testing procedures that do not depend on the true value of α_{11}.

Before concluding that pretesting is infeasible in the present context, one must first construct a sensible pretesting procedure. Given this procedure, one can then investigate its properties and show why the test fails if the number of unit roots in unknown.

From Lemma 8.4 in Appendix 8.A it follows that under the hypothesis that the number of cointegrating relationships is r, $T^{1/2}\hat{\Pi}_{11}$ is asymptotically normally distributed. This fact can be exploited for constructing a pretesting procedure. Conditional on the result of the pretesting procedure, one proceeds by either reordering the variables or by testing whether $\alpha_{22} = 0$. The most simple pretesting procedure tests whether $|\alpha_{11}|$ is significantly different from zero. If one cannot reject the hypothesis that $|\alpha_{11}| = 0$, the variables must be reordered. Otherwise, one can proceed with the cointegration testing procedure described in Sections 8.1 and 8.2. If there is no ordering of the variables for which one can reject the hypothesis $|\alpha_{11}| = 0$, then one cannot proceed with the cointegration test. Note that some care has to be taken if one wants to conclude that the number of cointegrating relations is smaller than the postulated r, given that $|\alpha_{11}| = 0$ cannot be rejected for all possible orderings of the variables. This is due to the fact that if the rank of the matrix Π is smaller than r, a test for $|\alpha_{11}| = 0$ has a nonstandard limiting distribution (see the discussion below).[3]

If one uses the pretesting procedure described above, it becomes important what sequence of tests is used for determining the number of cointegrating relations. This chapter suggests to start with the hypothesis $r = 0$ and then to increase the number of cointegrating relations by one at each rejection of the null hypothesis. Kleibergen and van Dijk (1994), instead, suggest to start with the hypothesis $r = k - 1$ and then to decrease the number of cointegrating

[3] Also note the following. Consider the testing sequence proposed in this chapter, namely the one that starts from the hypothesis $r = 0$ and proceeds by increasing r at each rejection of the null hypothesis. Then concluding that r must be decreased if one cannot reject $|\alpha_{11}| = 0$ for any ordering of the variables, implies that one has to go back to a null hypothesis that one has already rejected in an earlier stage of the testing sequence.

relations by one at each faillure to reject the null. It is argued here that the first testing sequence is more suitable when combined with the pretesting procedure for α_{11}. Assume that the number of cointegrating relations is equal to $k-1 > r_0 > 0$. Using the sequence of tests proposed in this chapter, one then starts with $H_0 : \ r = 0$. If small sample effects are ignored, this hypothesis is rejected. Next, one considers a test for the hypothesis $H_1 : \ r \leq 1$. Given that $k - 1 > r_0 > 0$, there always exist an ordering for which $|\alpha_{11}| \neq 0$. This holds for all hypotheses H_r with $r \leq r_0$. Therefore, all hypotheses H_r for $r < r_0$ are rejected in large enough samples, while H_{r_0} is not rejected with a probability of, say, 95%. In contrast, the testing sequence suggested by Kleibergen and van Dijk (1994) starts with $H_{k-1} : \ r = k - 1$. Given the true number of cointegrating relations $r_0 < k - 1$, there exists no ordering of the variables for which $|\alpha_{11}| = 0$. Moreover, a test based on $|\alpha_{11}| = 0$ has a nonstandard limiting distribution, which makes it difficult (if not impossible) to construct a feasible inference procedure.

Although one can prove that $\hat{\alpha}_{11}$ is asymptotically normally distributed, this is not sufficient for constructing a test statistic to test whether $|\alpha_{11}| = 0$. In addition, a consistent estimator for the asymptotic variance of $T^{1/2}|\alpha_{11}|$ is needed. A suitable test statistic is presented in the next theorem. This test statistic has a standard limiting distribution.

Theorem 8.4 *Let the conditions of Theorem 8.2 be satisfied and let* $\text{rank}(\Pi) = r$. *Further assume that one is performing the pretesting procedure described above for the hypothesis* $H_r^{a.b}$ *or* $H_r^{*a.b}$. *If* $\alpha_{11} = 0$, *then*

$$T\left(\text{tr}((\hat{\Pi}_{11}^{\top})^{-1}\hat{V}_{\xi}^{-1}\hat{\Pi}_{11}^{-1}\hat{V}_{\phi 1})\right)^{-1} \xrightarrow{d} \chi^2(1),$$

where $\xrightarrow{d}$ *denotes convergence in distribution,*

$$\begin{aligned}
\hat{V}_{\xi} &= T^{-1}\sum_{t=1}^{T}(y_{1,t-1} - \tilde{\beta}_2^{\top}y_{2,t-1})(y_{1,t-1} - \tilde{\beta}_2^{\top}y_{2,t-1})^{\top}, \\
\hat{V}_{\phi 1} &= T^{-1}\sum_{t=1}^{T}\hat{\Phi}_{11}^{-1}\begin{pmatrix}\phi_1(\hat{\varepsilon}_{1,1t})\\ \vdots \\ \phi_r(\hat{\varepsilon}_{r1t})\end{pmatrix}\begin{pmatrix}\phi_1(\hat{\varepsilon}_{1,1t})\\ \vdots \\ \phi_r(\hat{\varepsilon}_{r1t})\end{pmatrix}^{\top}\hat{\Phi}_{11}^{-1},
\end{aligned}$$

$\hat{\Phi}_{11}$ *the upper-left* $(r \times r)$ *block of* $\hat{\Phi}$, *defined above (8.8), and* $\tilde{\beta}_2$ *the estimator for the parameters in the regression model*

$$y_{1t} = \beta_2^{\top}y_{2t} + u_{1t},$$

where this regression model is augmented with a constant for the cases $a.2$ *and* $a.3$ *and with a trend for the cases* $a.4$ *and* $a.5$ $(a = 1, \ldots, 5)$.

Note that the parameters from the cointegrating relations (β_2) are estimated somewhat differently in Theorem 8.4 than in Section 8.1. As explained

in Subsection 8.5.1, the estimator for β_2 from Section 8.1 is $T^{-1/2}$-consistent rather than T^{-1}-consistent if $|\alpha_{11}| = 0$. As a result, the asymptotic variance of $|\hat{\alpha}_{11}|$ cannot be consistently estimated with this estimator. In contrast, the estimator $\tilde{\beta}_2$ is still T^{-1}-consistent even if $|\alpha_{11}| = 0$. The only important condition for $\tilde{\beta}_2$ to be T^{-1}-consistent is the validity of the restriction of the leading $(r \times r)$ submatrix of B to be the unit matrix, where B was defined below (8.5).

Theorem 8.4 reveals that a pretest statistic can be constructed that has a standard limiting distribution if the postulated number of unit roots is correct. Determining the number of unit roots, however, is a sequential process. It is therefore interesting to now what the properties of the test are if rank(Π) is in fact greater than r. Disappointingly, the test statistic from Theorem 8.4 no longer has a standard limiting distribution if rank(Π) $> r$, thus leading to either too few or too many rejections relative to the nominal significance level of the test. To illustrate this point, consider the case where the true number of cointegrating relations is $r+1$, so one more than the number that is tested. In that case, $\tilde{\beta}_2$ will, in general, only be $T^{1/2}$-consistent instead of T-consistent. As a result, the matrix $\hat{V}_\xi$ will converge to a random variate with a nondegenerate distribution.[4] This contrasts with the situation in Theorem 8.4, where $\hat{V}_\xi$ converges to a matrix of constants. Moreover, even if $\hat{V}_\xi$ converges to a matrix of constants, for example if the true number of cointegrating relations is k, then the test statistic in Theorem 8.4 is based on the wrong estimate of the variance of $\sqrt{T}|\alpha_{11}|$.

Concluding this subsection, it seems as yet infeasible to design a pretest for the hypothesis that $|\alpha_{11}| = 0$. The procedure of Kleibergen and van Dijk (1994) starts at the wrong hypothesis, namely $H_{k-1}^{a.b}$ instead of $H_0^{a.b}$. As a result, if the true number of cointegrating relations is small, the Kleibergen and van Dijk procedure almost automatically violates one of the main assumptions used for deriving the limiting distribution of the Wald test, namely $|\alpha_{11}| \neq 0$. The sequence of tests proposed in this subsection does not suffer from this defect. However, no useful pretesting procedure could be constructed for testing the assumption $|\alpha_{11}| \neq 0$. This was due to the fact that no proper estimate of the variance of $|\alpha_{11}|$ was available if the true number of cointegrating relations was allowed to exceed the postulated number.

8.5.3 Testing Procedures that are Independent of α_{11}

Instead of testing whether α_{11} is singular, one can follow a different route and try to remove the dependence of the cointegration test on the true value of

[4] This is due to the fact that one of the components of $\hat{V}_\xi$, namely

$$T^{-1}(\tilde{\beta}_2 - \beta_2)^\top \sum_{t=1}^{T} y_{2,t-1} y_{2-t-1}^\top (\tilde{\beta}_2 - \beta_2),$$

generally has a nondegenerate limiting distribution if $\tilde{\beta}_2 - \beta_2$ is $O_p(T^{-1/2})$ rather than $O_p(T^{-1})$.

α_{11}. As noted in Subsection 8.5.1, the singularity of α_{11} mainly affects the limiting distribution of the Wald test through its effect on $\hat{\beta}_2$. Therefore, an obvious way to remove the dependence of the Wald test on α_{11} is to remove the dependence of $\hat{\beta}_2$ on α_{11}. This can be achieved by using an alternative estimator for β_2. Kleibergen (1994) proposes to estimate β_2 by running the regressions

$$y_{1t} = \beta_2^\top y_{2t} + u_{1t}, \tag{8.15}$$

(compare Theorem 8.4). The problem with this approach is that the limiting distribution of $\hat{\beta}_2$ will be different from the one suggested in Section 8.1 due to the fact that the u_{1t} process will be non-i.i.d. in general. A result of the temporal dependence of the u_{1t} process is that additional nuisance parameters enter the limiting distribution. These nuisance parameters can be estimated using nonparametric methods as in, e.g., Phillips (1987). Eliminating the effect of these nuisance parameters on the limiting distribution of the Wald cointegration test, however, might prove a non-trivial task.

Instead of fixing the defects of the Wald test, one can use different testing principles, like tests based on the likelihood ratio (LR) or on the Lagrange multiplier principle. The LR test was thoroughly discussed in Chapter 7. This testing procedure did not suffer from a dependence on the true value of α_{11}. The LR test, however, depended in a complicated way on nuisance parameters that could not easily be eliminated. Alternatively, one could try to construct an LM test. This can be done as follows. Consider the simple case of a VAR model of order one, data generating process $a = 1$, and regression model $b = 1$. LM tests for the more complicated models can be derived analogously. Under the hypothesis $H_r^{1.1} : \alpha_{22} = 0$, (8.4) and (8.5) can be jointly written as $y_t = AB^\top y_{t-1} + \varepsilon_t$. Let $\tilde{A}$ and $\tilde{B}^\top = (I_r, -\tilde{\beta}_2^\top)$ denote the parameter estimates under the null hypothesis and let $\tilde{\varepsilon}_t$ denote the corresponding residuals. Next, define

$$\begin{aligned}
\ell(A,B) &= A_\perp^\top (L_1(A,B))^{-1} (T^{-1} \sum_{t=1}^{T} y_{2,t-1} \otimes \phi(\varepsilon_t)), \\
L(A,B) &= A_\perp^\top (L_1(A,B))^{-1} L_2(A,B) (L_1(A,B)^\top)^{-1} A_\perp, \\
L_1(A,B) &= T^{-2} \sum_{t=1}^{T} y_{2,t-1} y_{2,t-1}^\top \otimes \Phi_t, \\
L_2(A,B) &= T^{-2} \sum_{t=1}^{T} y_{2,t-1} y_{2,t-1}^\top \otimes \phi(\varepsilon_t)\phi(\varepsilon_t)^\top,
\end{aligned}$$

with $\varepsilon_t = y_t - AB^\top y_{t-1}$ and $A_\perp^\top$ such that $A_\perp^\top A = 0$. Note that $A_\perp^\top$ need not be equal to $(\alpha_{21}\alpha_{11}^{-1}, I_{k-r})$ due to the possible singularity of α_{11}. The following theorem now follows directly along the lines of the proof of Theorem 7.1.

Theorem 8.5 *Let the assumptions of Theorem 8.2 be satisfied, save the requirement that* $|\alpha_{11}| \neq 0$. *Then*

$$LM_r = \ell(\tilde{A},\tilde{B})^\top (L(\tilde{A},\tilde{B}))^{-1} \ell(\tilde{A},\tilde{B})$$

$$\Rightarrow \quad \text{tr}\left((\int B_1 dB_2^\top)^\top (\int B_1 B_1^\top)^{-1} (\int B_1 dB_2^\top)\right),$$

with B_1 and B_2 as defined in Theorem 8.2.

Theorem 8.5 reveals that the LM test has the same limiting distribution as the one presented in Theorem 8.2 and, moreover, does not depend on whether $|\alpha_{11}| = 0$ or not. An LM test based on a robust estimator, therefore, seems the most promising for application purposes. The only complication of the LM test as opposed to the Wald test is that estimation becomes more difficult. This is due to the cross-equations restrictions between (8.4) and (8.5) under the null hypothesis of no-cointegration.

8.5.4 Simulation Results for $|\alpha_{11}| = 0$

In this subsection, the results of three simulation experiments are presented, illustrating the size and power behavior of the cointegration Wald test for the case $|\alpha_{11}| = 0$. The setup of the simulations is the same as in Subsection 8.4.2, only the data generating mechanisms differ.

In the first experiment, the data generating process is given by

$$\Delta y_t = (0, 0, c/T)(1, -1, -1)^\top y_{t-1} + \varepsilon_t, \tag{8.16}$$

with $y_t \in \mathbb{R}^3$, $T = 100$, and c taking the values mentioned in Subsection 8.4.2. The roots corresponding to this system are 1, 1, and $(1 - c/T)^{-1}$. For $c \neq 0$, (8.16) contains one cointegrating relationship. Moreover, α_{11} is singular when testing $H_1^{1.5}$ and $H_2^{1.5}$. The rejection frequencies of the Wald test for the different M estimators discussed in Subsection 8.4.1 are presented in Table 8.3. These frequencies are based on 400 Monte-Carlo replications using standard Gaussian ε_t's.

It appears from the left three columns of Table 8.3 that the Wald test is able to detect at least one cointegrating relationship if sufficiently distant alternatives are considered. Moreover, there appears to be no evidence for the presence of two or more cointegrating relationships. For the test of the hypothesis $r \leq 1$, the rejection frequencies apear to increase somewhat with c, but it seems that this effect is only due to sampling variability. Unreported results show that for $c = 100$ the rejection frequencies over 400 Monte-Carlo simulations are again about 6%.

In the second experiment, the joint failure of $|\alpha_{11}| \neq 0$ and $B^\top = (I, -\beta_2^\top)$ was investigated. The two-dimensional data generating process considered, is given by

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -c/T \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \tag{8.17}$$

while the relevant hypothesis to be tested is $H_1^{1.5}$. Note that $A^\top = (0, -c/T)$ and $B^\top = (0, 1)$. So there are now two defects. First, α_{11} is singular, which results in a discrepancy between the asymptotic distribution of the test and the asymptotic distribution presented in Theorem 8.2. Second, the leading

TABLE 8.3
Power Simulations for $|\alpha_{11}| = 0$

		Experiment 1			Experiment 2			
					Gaussian		T. Cauchy	
ν	c	$r=0$	$r=1$	$r=2$	$r=0$	$r=1$	$r=0$	$r=1$
	0	0.052	0.033	0.018	0.052	0.013	0.055	0.040
	1	0.025	0.010	0.007	0.085	0.037	0.060	0.030
∞	5	0.080	0.025	0.003	0.085	0.050	0.052	0.035
	10	0.240	0.063	0.010	0.120	0.122	0.128	0.110
	20	0.785	0.113	0.005	0.318	0.380	0.343	0.398
	0	0.055	0.033	0.013	0.065	0.013	0.050	0.015
	1	0.037	0.010	0.003	0.083	0.033	0.072	0.010
5	5	0.065	0.030	0.000	0.095	0.028	0.085	0.085
	10	0.198	0.070	0.007	0.117	0.105	0.383	0.365
	20	0.725	0.120	0.003	0.330	0.295	0.922	0.715
	0	0.050	0.030	0.007	0.043	0.020	0.028	0.010
	1	0.060	0.020	0.010	0.058	0.028	0.043	0.010
1	5	0.055	0.030	0.007	0.070	0.030	0.090	0.080
	10	0.138	0.055	0.005	0.098	0.080	0.380	0.335
	20	0.475	0.092	0.005	0.225	0.195	0.905	0.642

The columns under the heading Experiment 1 contain the rejection frequencies of the Wald cointegration test for the M estimators described in Section 8.4 and data generating process (8.16). The columns under the heading Experiment 2 contain the results for the same estimators and data generating process (8.17). Gaussian means that the disturbances in the model were generated from the multivariate standard normal, while T. Cauchy means that they were generated from a truncated Cauchy distribution (see Section 7.6).

number in $B^\top$ cannot be normalized to unity. The rejection frequencies of the Wald test for Gaussian and truncated Cauchy ε_t are presented in the right four columns of Table 8.3.

The most important feature that emerges from the right columns of Table 8.3 is that the rejection frequencies of the hypothesis $H_1^{1.5}: \ r \leq 1$ strongly increase with the parameter c. This means that the Wald test detects too much cointegration. In fact, there is only one cointegrating relation in (8.17), while the Wald test rejects the hypothesis $H_1^{1.5}$ for a considerable number of simulations if $c = 20$. Therefore, it appears to be very important to check whether the assumption $B^\top = (I_r, -\beta_2^\top)$ is satisfied when applying the test to empirical data. Alternatively, one could try to circumvent the problem of a singular leading submatrix in $B^\top$ by choosing a different parameterization (or normalization). This can be done by imposing the following restrictions. First, the $(r \times r)$ matrix $\tilde{\beta}_1$ and the $((k-r) \times r)$ matrix $\tilde{\beta}_2$ must be chosen such that $\tilde{B}^\top \tilde{B} = I_r$, with $\tilde{B} = (\tilde{\beta}_1^\top, \tilde{\beta}_2^\top)$. Second, $\tilde{\alpha}_{11}$ must be chosen as a symmetric matrix. Note that for an arbitrary matrix $\Pi_1 = (\Pi_{11}, \Pi_{12})$, one can always find matrices satisfying the above criteria and $\Pi_1 = \tilde{\alpha}_{11}\tilde{B}^\top$. If $U\Sigma V^\top$

denotes the singular value decomposition of Π_1, then one can set $\tilde{\alpha}_{11} = U\Sigma U^\top$ and $\tilde{B} = VU^\top$. Therefore, the parameterization described above provides an alternative for the parameterization in (8.3). Using this parameterization of (8.17), additional simulations revealed that the Wald test still rejected the hypothesis $H_1^{1.5}$ in a large number of cases.

In the third experiment, the following model was used:

$$\Delta y_{1t} = \varepsilon_{1t}; \quad y_{2t} = 10^c y_{1t} + \varepsilon_{2t}. \tag{8.18}$$

Note that there is one cointegrating relationship, namely $10^c y_{1t} - y_{2t}$. The number of rejections of the (true) hypothesis $H_1^{1.5}$ over 1,000 simulations is plotted in Figure 8.3 for different values of c. The number of rejections should equal approximately 5%. This number is approximately obtained for $c \leq 0$ if the variables are ordered as $(y_{2t}, y_{1t})^\top$ (see the dashed line). For the same ordering of the variables and $c > 0$, the test appears to be undersized. Now consider the reverse ordering of the variables, $(y_{1t}, y_{2t})^\top$, such that $|\alpha_{11}| = 0$. Then the number of rejections is a decreasing function of c. For $c > 0$, the test is clearly undersized. For negative values of c, however, the number of rejections increases dramatically. This result explains the two previous findings. If the stochastic trend is strongly present in both components of the time series, i.e., c is large and positive, then the Wald test is undersized (compare the first experiment). If one of the components is almost stationary and the variables are incorrectly ordered, i.e., c is large and negative, then the Wald test is heavily oversized (compare the second experiment).

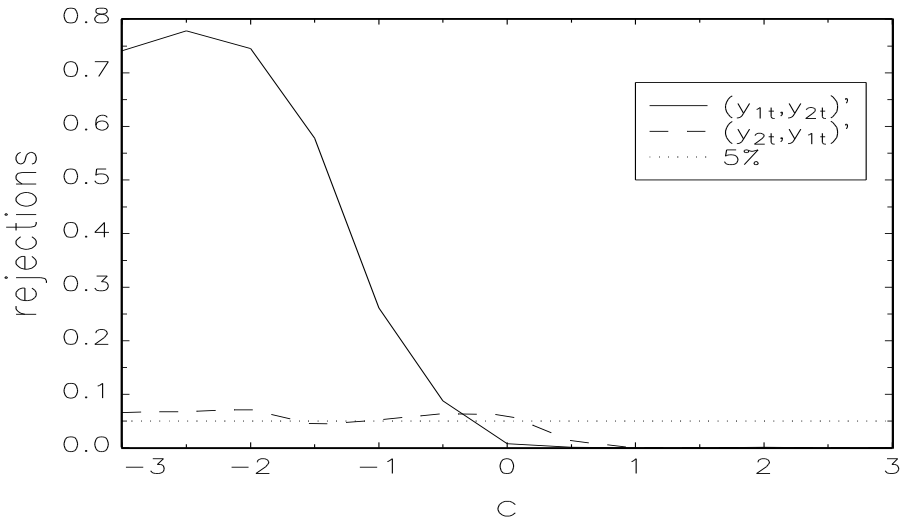

Figure 8.3.— Rejection Frequencies of the Wald Test for Different Orderings of the Variables for the Model $\Delta y_{1t} = \varepsilon_{1t}$ and $y_{2t} = 10^c y_{1t} + \varepsilon_{2t}$

8.6 Concluding Remarks

This chapter discussed the properties of an outlier robust cointegration test based on the Wald principle. The idea of the Wald test was taken from Kleibergen and van Dijk (1994). It turned out that the Wald test can be computed by performing robust single equation regressions. This is an advantage compared to the pseudo likelihood ratio test discussed in Chapter 7. Another advantage of the Wald test compared to the pseudo likelihood ratio test, is that less nuisance parameters are present in the limiting distribution of the test statistic. It was also discussed how the asymptotic distribution of the test is affected by the presence of deterministic components in either the data generating process or the fitted regression model.

The main disadvantage of the Wald test is that the ordering of the variables can be extremely important for the asymptotic distribution of the test. If the variables are not properly ordered, the Wald test can either underreject or overreject. No pretest procedure could be developed for testing whether the chosen ordering is correct. As an alternative, a robust Lagrange Multiplier (LM) cointegration test was discussed. The LM test has two advantages. First, its asymptotic distribution does not depend upon the ordering of the variables. Second, the limiting distribution of the LM test depends on the same number of nuisance parameters as the limiting distribution of the Wald test. This number of nuisance parameters is smaller than or equal to the number of nuisance parameters that enter the limiting distribution of the (pseudo) likelihood ratio test discussed in Chapter 7. Therefore, the LM cointegration test seems to be the best test at the moment for constructing outlier robust cointegration testing procedures that depend upon a small set of nuisance parameters.

The properties of the Wald cointegration test were illustrated by means of a small simulation experiment. As in Chapter 7, for Gaussian innovations the OLS-based test had the highest power. For fat-tailed innovations, however, tests based a Student t M estimator distribution performed best. Especially the estimator based upon the Student t likelihood with five degrees of freedom performed well, both for Gaussian innovations and truncated Cauchy innovations.

Another conclusion that emerged from the simulations, is that the asymptotic distribution provides a poor approximation to the finite sample distribution of the robust cointegration tests. It would, therefore, be a valuable contribution to the econometric literature if one could extend the finite sample results of Abadir and Larson (1994) for the OLS estimator to cointegration tests based on M estimators.

8.A Proofs

In this appendix, the theorems of Sections 8.3 and 8.5 are proved. First, the representation Theorem 8.1 is proved. Next, an elaborate proof of Theorem 8.2 follows. This latter proof directly follows the two-step procedure of Kleibergen and van Dijk

(1994). Therefore, it has to take account of the errors-in-variables problem in the second-step regression. An alternative proof using a systems approach rather than the two-step procedure can also be constructed based on the methods used in Appendix 7.A.

Proof of Theorem 8.1. The proof mimics that of Johansen (1991, Theorem 4). Using his techniques, one can easily derive that

$$(\tilde{Z}_t^\top, \tilde{Y}_t^\top)^\top = \tilde{A}(L)^{-1}(A, A_\perp)^\top(\varepsilon_t + \gamma + \delta t),$$

with

$$\tilde{A}(L) = \begin{pmatrix} -A^\top AB^\top B + A^\top\tilde{\Psi}(L)B(1-L) & A^\top\tilde{\Psi}(L)B_\perp \\ A_\perp^\top\tilde{\Psi}(L)B(1-L) & A_\perp^\top\tilde{\Psi}(L)B_\perp \end{pmatrix},$$

$\tilde{\Psi}(L) = (I + \Pi - \sum_{i=1}^{p-1}\Psi_i L^i)$, $\tilde{Z}_t = (B^\top B)^{-1}B^\top y_t$, and $\tilde{Y}_t = (B_\perp^\top B_\perp)^{-1}B_\perp^\top \Delta y_t$. Define

$$C(L) = (\Delta B, B_\perp)\tilde{A}(L)^{-1}(A, A_\perp)^\top,$$

and $S(L) = (C(L) - C(1))/(1-L)$. Note that $C(1) = C$ and $\Delta y_t = B_\perp\tilde{Y}_t + B\Delta\tilde{Z}_t$. Therefore,

$$\begin{aligned} \Delta y_t &= C(L)(\varepsilon_t + \gamma + \delta t) \\ &= (C + S(L)\Delta)(\varepsilon_t + \gamma + \delta t) \\ &= C\varepsilon_t + C\gamma + C\delta t + S(L)(\Delta\varepsilon_t + \delta) \\ &= C\varepsilon_t + C\gamma + S(1)\delta + C\delta t + S(L)\Delta\varepsilon_t. \end{aligned}$$

(8.11) now follows easily from the fact that $y_t = y_0 + \Delta y_1 + \ldots + \Delta y_t$. □

Corollary 8.2

$$E(B^\top y_t) = \tilde{C}(1)\gamma + \tilde{S}(1)\delta + \tilde{C}(1)\delta t,$$

with $\tilde{C}(1) = -(A^\top A)^{-1}A^\top(I - \Psi C)$, *and*

$$\tilde{S}(1) = -\tilde{C}(1)(\Psi B, -\Psi'(1)B_\perp)\tilde{A}(1)^{-1}(A, A_\perp)^\top.$$

Moreover, $\tilde{C}(1) = B^\top S(1)$.

Proof. Define $\tilde{C}(L) = (B^\top B, 0)\tilde{A}(L)^{-1}(A, A_\perp)^\top$ and $\tilde{S}(L) = (\tilde{C}(L) - \tilde{C}(1))/(1-L)$. Using the notation of the proof of Theorem 8.1, one obtains

$$\begin{aligned} E(B^\top y_t) &= E(B^\top B\tilde{Z}_t) \\ &= E(\tilde{C}(L)(\varepsilon_t + \gamma + \delta t)) \\ &= E((\tilde{C}(1) + \tilde{S}(L)\Delta)(\gamma + \delta t)) \\ &= \tilde{C}(1)\gamma + \tilde{S}(1)\delta + \tilde{C}(1)\delta t. \end{aligned}$$

The final part of the corollary follows by noting that $B^\top S(1) = -B^\top\partial C(z)/\partial z|_{z=1}$ and

$$-B^\top \left.\frac{\partial C(z)}{\partial z}\right|_{z=1} = (B^\top B, 0)\tilde{A}(1)^{-1}(A, A_\perp)^\top = \tilde{C}(1).$$

□

In the remainder of this appendix, the order of the VAR polynomial in (8.1), $p+1$, is set equal to 1. The asymptotic distributions of the test statistics are not

changed if $p > 0$. Moreover, it is continually assumed that the conditions of Theorem 8.2 are satisfied.

In order to prove Theorem 8.2, some additional definitions are needed. Let

$$\phi(\varepsilon_t) = (\phi_{1t}, \ldots, \phi_{kt})^\top$$

with ϕ_{nt} denoting $\phi_n(\cdot)$ evaluated at the nth row of ε_t. Let $\phi_{1\cdot}(\varepsilon_{1t})$ and $\phi_{2\cdot}(\varepsilon_{2t})$ denote the first r and last $k-r$ rows of $\phi(\varepsilon_t)$, respectively. Define Φ_t analogously to $\hat{\Phi}_t$, with $\hat{\varepsilon}^b_{1t}$ and $\hat{\eta}_t$ in the definition of $\hat{\Phi}_t$ replaced by ε_{1t} and ε_{2t}, respectively. Moreover, let Φ be a $(k \times k)$ diagonal matrix with the nth diagonal element equal to $E(\phi_n'(e_n^\top \varepsilon_t))$, where e_n is the nth column of I_k. Φ_{11} denotes the upper-left $(r \times r)$-block of Φ. Furthermore, Φ_{22}, $\Phi_{11,t}$ and $\Phi_{22,t}$ are defined in the obvious way.

Let $\mu_c^{a\cdot} = E(B^\top y_t)$ for $a = 1, 2, 3$, and $\mu_c^{a\cdot} + \tau_c^{a\cdot} t = E(B^\top y_t)$ for $a = 4, 5$. For completeness, $\tau_c^{a\cdot} = 0$ for $a = 1, 2, 3$. The quantities $\mu_c^{a\cdot}$ and $\tau_c^{a\cdot}$ are used to correct for the mean or trend in the cointegrating relationships. From the second part of Corollary 8.2 it follows that $\tau_c^{4\cdot} = \tau_c^{5\cdot} = B^\top S(1)\delta$.

Let $\mathrm{diag}(a_1, a_2)$ denote a block-diagonal matrix with blocks a_1 and a_2. The following matrices $K_T^{a.b}$ are needed in order to standardize the y_t process for the various data generating processes a and regression models b.

$$\begin{aligned}
K_T^{1.1} &= (B, T^{-1/2} B_\perp)^\top \\
K_T^{1.2} &= \begin{pmatrix} B^\top & -\mu_c^{1\cdot} \\ T^{-1/2} B_\perp^\top & 0 \\ 0 & 1 \end{pmatrix} \\
K_T^{1.3} &= K_T^{1.2} \\
K_T^{1.4} &= \mathrm{diag}(K_T^{1.2}, 1/T) \\
K_T^{1.5} &= K_T^{1.4} \\
K_T^{3.3} &= \begin{pmatrix} B^\top & -\mu_c^{3\cdot} \\ T^{-1} \zeta_1^\top B_\perp^\top & 0 \\ T^{-1/2} \zeta_{1\perp}^\top B_\perp^\top & 0 \\ 0 & 1 \end{pmatrix} \\
K_T^{3.4} &= \begin{pmatrix} B^\top & -\mu_c^{3\cdot} & 0 \\ T^{-1/2} B_\perp^\top & 0 & -T^{-1/2} B_\perp^\top C\gamma \\ 0 & 1 & 0 \\ 0 & 0 & T^{-1} \end{pmatrix} \\
K_T^{3.5} &= K_T^{3.4} \\
K_T^{4.4} &= \begin{pmatrix} B^\top & -\mu_c^{4\cdot} & -B^\top S(1)\delta \\ T^{-1/2} B_\perp^\top & 0 & -T^{-1/2} B_\perp^\top \zeta_2 \\ 0 & 1 & 0 \\ 0 & 0 & T^{-1} \end{pmatrix} \\
K_T^{4.5} &= K_T^{4.4} \\
K_T^{5.5} &= \begin{pmatrix} B^\top & -\mu_c^{5\cdot} & -B^\top S(1)\delta \\ T^{-2} \zeta_3^\top B_\perp^\top & 0 & 0 \\ T^{-1/2} \zeta_{3\perp}^\top B_\perp^\top & 0 & -T^{-1/2} \zeta_{3\perp}^\top B_\perp^\top \zeta_2 \\ 0 & 1 & 0 \\ 0 & 0 & T^{-1} \end{pmatrix},
\end{aligned}$$

with

$$\begin{aligned}\zeta_1 &= B_\perp^\top C\gamma/(\gamma^\top C^\top B_\perp B_\perp^\top C\gamma)\\ \zeta_2 &= (C\gamma + S(1)\delta)\\ \zeta_3 &= 2B_\perp^\top C\delta/(\delta^\top C^\top B_\perp B_\perp^\top C\delta).\end{aligned}$$

$K_T^{2.2}$ to $K_T^{2.5}$ are equal to $K_T^{1.2}$ to $K_T^{1.5}$, with $\mu_c^{1\cdot}$ replaced by $\mu_c^{2\cdot}$.

The following lemma follows directly from Phillips and Durlauf (1986).

Lemma 8.1

$$T^{-1/2}\sum_{t=1}^{\lfloor sT\rfloor}\begin{pmatrix}\varepsilon_t\\ \phi(\varepsilon_t)\end{pmatrix} \Rightarrow W(s) = \begin{pmatrix}W_1(s)\\ W_2(s)\end{pmatrix} = \begin{pmatrix}W_1\\ W_2\end{pmatrix},$$

with $W(s)$ a vector Brownian motion with covariance matrix

$$\Omega = \begin{pmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{pmatrix},$$

and $\lfloor x\rfloor$ denoting the integer part of x. Ω and Ω_{11} are $(k^2\times k^2)$ and $(k\times k)$-matrices, respectively.

The following results are now easily established.

Lemma 8.2 *$diag(0, I_{k-r+i})K_T^{a.b}Y_{1,\lfloor sT\rfloor}^b \Rightarrow g^{a.b}$, with $i = 0$ for $b = 1$, $i = 1$ for $b = 2, 3$, $i = 2$ for $b = 4, 5$, and*

$$\begin{aligned}g^{i.1} &= (W_1(s)^\top C^\top B_\perp)^\top,\\ g^{i.2} &= (W_1(s)^\top C^\top B_\perp, 1)^\top,\\ g^{i.3} &= g^{i.2},\\ g^{j.4} &= (W_1(s)^\top C^\top B_\perp, 1, s)^\top,\\ g^{j.5} &= g^{j.4},\\ g^{3.3} &= (s, W_1(s)^\top C^\top B_\perp\zeta_{1\perp}, 1)^\top,\\ g^{5.5} &= (s^2, W_1(s)^\top C^\top B_\perp\zeta_{3\perp}, 1, s)^\top,\end{aligned}$$

for $i = 1, 2$ and $j = 1, 2, 3, 4$.

Proof. All of these results follow rather easily from the decomposition derived in Theorem 8.1, the weak convergence of the partial sum process in Lemma 8.1, and the continuous mapping theorem (see Billingsley (1968)). □

Lemma 8.3 $T^{-1}\sum_{t=1}^{\lfloor sT\rfloor} K_T^{a.b}Y_{1t}^b(K_T^{a.b}Y_{1t}^b)^\top \Rightarrow G^{a.b}$, *with*

$$\begin{aligned}G^{1.1} &= \begin{pmatrix}\Xi_1 & 0\\ 0 & \int\tilde{W}_1\tilde{W}_1^\top\end{pmatrix},\\ G^{i.2} &= \begin{pmatrix}\Xi_1 & 0 & 0\\ 0 & \int\tilde{W}_1\tilde{W}_1^\top & \int\tilde{W}_1\\ 0 & \int\tilde{W}_1^\top & 1\end{pmatrix},\end{aligned}$$

$$
\begin{aligned}
G^{i.3} &= G^{i.2}, \\
G^{j.4} &= \begin{pmatrix} \Xi_1 & 0 & 0 & 0 \\ 0 & \int \tilde{W}_1 \tilde{W}_1^\top & \int \tilde{W}_1 & \int s\tilde{W}_1 \\ 0 & \int \tilde{W}_1^\top & 1 & \frac{1}{2} \\ 0 & \int s\tilde{W}_1^\top & \frac{1}{2} & \frac{1}{3} \end{pmatrix}, \\
G^{j.5} &= G^{j.4}, \\
G^{3.3} &= \begin{pmatrix} \Xi_1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \int s\tilde{W}_1^\top \zeta_{1\perp} & \frac{1}{2} \\ 0 & \zeta_{1\perp}^\top \int s\tilde{W}_1 & \zeta_{1\perp}^\top \int \tilde{W}_1 \tilde{W}_1^\top \zeta_{1\perp} & \zeta_{1\perp}^\top \int \tilde{W}_1 \\ 0 & \frac{1}{2} & \int \tilde{W}_1^\top \zeta_{1\perp} & 1 \end{pmatrix}, \\
G^{5.5} &= \begin{pmatrix} \Xi_1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \int s^2 \tilde{W}_1^\top \zeta_{3\perp} & \frac{1}{3} & \frac{1}{4} \\ 0 & \zeta_{3\perp}^\top \int s^2 \tilde{W}_1 & \zeta_{3\perp}^\top \int \tilde{W}_1 \tilde{W}_1^\top \zeta_{3\perp} & \zeta_{3\perp}^\top \int \tilde{W}_1 & \zeta_{3\perp}^\top \int s\tilde{W}_1 \\ 0 & \frac{1}{3} & \int \tilde{W}_1^\top \zeta_{3\perp} & 1 & \frac{1}{2} \\ 0 & \frac{1}{4} & \int s\tilde{W}_1^\top \zeta_{3\perp} & \frac{1}{2} & \frac{1}{3} \end{pmatrix},
\end{aligned}
$$

for $i = 1, 2$ *and* $j = 1, 2, 3, 4$, *with* $\tilde{W}_1 = B_\perp^\top C W_1$. Ξ_1 *equals the variance of the cointegrating relationships,*[5]

$$\Xi_1 = E\{(B^\top y_t - \mu_c^{a\cdot} - \tau_c^{a\cdot} t)(B^\top y_t - \mu_c^{a\cdot} - \tau_c^{a\cdot} t)^\top\}.$$

Proof. All of these results follow directly from Theorem 8.1, Lemma 8.1, and the continuous mapping theorem (see Billingsley (1968)). □

The results of Lemmas 8.2 and 8.3 can be used to derive the asymtotic distribution of $(\hat{\Theta}_1^b - \Theta_1^b)$. This is done in the following lemma.

Lemma 8.4 *Let* $\hat{\varepsilon}_{1t}^b - \varepsilon_{1t} = o_p(1)$ *uniformly for* $t = 1, \ldots, T$. *Then, if* $\operatorname{rank}(\Pi) = r$,

$$T^{1/2}(I_r \otimes (K_T^{a.b})^\top)^{-1} \operatorname{vec}((\hat{\Theta}_1^{a.b} - \Theta_1^b)^\top) \Rightarrow (\Phi_{11} \otimes G^{a.b})^{-1} \left(\int dW_{21} \otimes \begin{pmatrix} d\xi_1 \\ g^{a.b} \end{pmatrix} \right),$$

with $W_{21}(s)$ *denoting the first* r *rows of* $W_2(s)$ *and* $\int (dW_{21}) \otimes (d\xi_1)$ *denoting a normally distributed random variate.*

Proof. The M estimator is given by the set of parameter values that solves

$$\sum_{t=1}^{T} Y_{1t}^b \hat{\phi}_{nt}^b = 0 \qquad (n = 1, \ldots, r). \tag{8.19}$$

The r different equations in (8.19) can be stacked into the system of equations $\sum_{t=1}^{T} \phi_{1\cdot}(\hat{\varepsilon}_t^b) \otimes Y_{1t}^b = 0$. Taking a first order Taylor expansion of this system around the true parameter values, one obtains

$$0 = T^{-1} \sum_{t=1}^{T} \phi_{1\cdot}(\varepsilon_{1t}) \otimes K_T^{a.b} Y_{1t}^b + (T^{-1} \sum_{t=1}^{T} \Phi_{11,t} \otimes K_T^{a.b} Y_{1t}^b (K_T^{a.b} Y_{1t}^b)^\top) \cdot$$

[5]Note that Ξ_1 does not depend on the data generating process nor on the regression model.

$$(I_r \otimes (K_T^{a.b})^\top)^{-1}\mathrm{vec}((\hat{\Theta}_1^b - \Theta_1^b)^\top) + R_T^{a.b}, \tag{8.20}$$

with $R_T^{a.b}$ denoting the remainder term. The Lipschitz continuity of $\phi_n'(\cdot)$ and the fact that $(\hat{\varepsilon}_{1t}^b - \varepsilon_{1t}) = o_p(1)$ uniformly for $t = 1, \ldots, T$, together imply that

$$R_T^{a.b} = o_p(1) \cdot (I_r \otimes (K_T^{a.b})^\top)^{-1}\mathrm{vec}((\hat{\Theta}_1^b - \Theta_1^b)^\top)$$

(compare Appendix 6.A). Therefore,

$$T^{1/2}(I_r \otimes (K_T^{a.b})^\top)^{-1}\mathrm{vec}((\hat{\Theta}_1^b - \Theta_1^b)^\top) =$$

$$(T^{-1}\sum_{t=1}^{T}\Phi_{11,t} \otimes K_T^{a.b}Y_{1t}^b(K_T^{a.b}Y_{1t}^b)^\top + o_p(1))^{-1} \cdot (T^{-1/2}\sum_{t=1}^{T}\phi_{1.}(\varepsilon_{1t}) \otimes K_T^{a.b}Y_{1t}^b). \tag{8.21}$$

Using Lemma 8.2 and Theorem 3.2 of Hansen (1992), one obtains

$$T^{-1}\sum_{t=1}^{T}\Phi_{11,t} \otimes K_T^{a.b}Y_{1t}^b(K_T^{a.b}Y_{1t}^b)^\top \Rightarrow \Phi_{11} \otimes G^{a.b}.$$

From Theorem 2.1 of Hansen (1992), it follows that

$$T^{-1/2}\sum_{t=1}^{T}\phi_{1.}(\varepsilon_{1t}) \otimes K_T^{a.b}Y_{1t}^b \Rightarrow \int dW_{21} \otimes \begin{pmatrix} d\xi_1 \\ g^{a.b} \end{pmatrix},$$

with $\int(dW_{21} \otimes (d\xi_1))$ denoting $\lim_{T\to\infty} T^{-1/2}\sum_{t=1}^{T}\phi_{1.}(\varepsilon_{1t}) \cdot (B^\top y_{t-1} - \mu_c^{a\cdot} - \tau_c^{a\cdot}t)$. This proves the theorem. □

Lemma 8.4 reveals that certain linear combinations of the parameters from the first step regression are $T^{1/2}$-consistent, while other combinations are T-consistent. Therefore, the elements of $\hat{\Pi}_1$ are only $T^{1/2}$-consistent in general. Despite this fact, the following lemma shows that $\hat{\beta}_2 = -\hat{\Pi}_{12}\hat{\Pi}_{11}^{-1} = \beta_2 + O_p(T^{-1})$.

Lemma 8.5 *Given the assumptions of Lemma 8.4,*

$$T\mathrm{vec}(\hat{\beta}_2 - \beta_2) = O_p(1).$$

Proof. Define Λ^b as the set of parameters in the cointegrating relationship of regression model b, e.g., $\Lambda^4 = (\beta_2^\top, \tilde{\gamma}_1, \tilde{\delta}_1)$. Moreover, let $\lambda^b = -\Pi_{11}\Lambda^b$. It is easily checked that

$$\begin{aligned}
(\hat{\Lambda}^b - \Lambda^b) &= -(\hat{\Pi}_{11}^{-1}\hat{\lambda}^b - \Pi_{11}^{-1}\lambda^b)^\top \\
&= -\hat{\Pi}_{11}^{-1}(\lambda^b + (\hat{\lambda}^b - \lambda^b) - [\Pi_{11} + (\hat{\Pi}_{11} - \Pi_{11})]\Pi_{11}^{-1}\lambda^b) \\
&= -\hat{\Pi}_{11}^{-1}(\hat{\lambda}^b - \lambda^b - (\hat{\Pi}_{11} - \Pi_{11})\Pi_{11}^{-1}\lambda^b) \\
&= -\hat{\Pi}_{11}^{-1}((\hat{\Pi}_{11} - \Pi_{11})\Lambda^b + (\hat{\lambda}^b - \lambda^b)I) \\
&= -\hat{\Pi}_{11}^{-1}(\hat{\Theta}_1^b - \Theta_1^b)((\Lambda^b)^\top, I)^\top.
\end{aligned}$$

Note that the upper-left $(k \times (k-r))$ block of $((\Lambda^b)^\top, I)^\top$ equals $B_\perp$. Therefore,

$$\mathrm{vec}((\hat{\Lambda}^b - \Lambda^b)^\top) = (-\hat{\Pi}_{11}^{-1} \otimes ((\Lambda^b)^\top, I))\mathrm{vec}((\hat{\Theta}_1^b - \Theta_1^b)^\top).$$

The lemma now follows immediately from Lemma 8.4 by noting that the first $(k-r)$ rows of $T^{1/2}((\Lambda^b)^\top, I)(K_T^{a.b})^\top$ are $O(1)$ for all considered combinations of a and b. □

Remark 8.1 For some combinations $(a.b)$, there exist linear combinations of $T(\hat{\beta}_2 - \beta_2)$ that are $o_p(1)$. For example, for combination 3.3 a vector v exists, such that $v^\top(\hat{\beta}_2 - \beta_2) = O_p(T^{-3/2})$.

Lemma 8.4 presents the asymptotic distribution of the estimators that are used in the first step regression, while Lemma 8.5 presents the results for the estimators of the parameters that enter the cointegrating relationship. In order to derive the asymptotic distribution of the estimators that are used in the second step regression, define the matrices $\hat{L}_b$, such that $\hat{Y}^b_{2t} = \hat{L}_b Y^b_{1t}$. For example,

$$\hat{L}_2 = \begin{pmatrix} I_r & -\hat{\beta}_2^\top & -\hat{\tilde{\gamma}}_1 \\ 0 & I_{k-r} & 0 \end{pmatrix}.$$

Note that the matrices $\hat{L}_b$ need not be square. The matrices L_b are defined similarly, $Y^b_{2t} = L_b Y^b_{1t}$. The relation between $\hat{L}_b$ and L_b is presented in the following lemma.

Lemma 8.6

$$\hat{L}_b - L_b = -\begin{pmatrix} I_r \\ 0 \end{pmatrix} \hat{\Pi}_{11}^{-1}(\hat{\Theta}^b_1 - \Theta^b_1) \begin{pmatrix} \Lambda^b \\ I \end{pmatrix} (0, I),$$

where the column dimension of zero block in the final matrix $(0, I)$ *is equal to* r.

Proof. It is easily checked that either

$$\hat{L}_b = \begin{pmatrix} I & -\hat{\Lambda}^b \\ 0 & I \end{pmatrix}, \qquad \text{or} \qquad \hat{L}_b = \begin{pmatrix} I & -\hat{\Lambda}^b \\ 0 & (I, 0) \end{pmatrix}.$$

Therefore, $\hat{L}_b - L_b = -(I, 0)^\top(\hat{\Lambda}^b - \Lambda^b)(0, I)$. The result now follows by using Lemma 8.5. □

The following matrices are needed to normalize the vector process $\{Y^b_{2t}\}$.

$$\tilde{K}_T^{3.3} = \begin{pmatrix} I_r & 0 & \tilde{\gamma}_1 - \mu_c^{3.} \\ 0 & T^{-1}\zeta_1^\top(B_\perp^\top B_\perp) & 0 \\ 0 & T^{-1/2}\zeta_{1\perp}^\top(B_\perp^\top B_\perp) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\tilde{K}_T^{3.4} = \tilde{K}_T^{3.3},$$

$$\tilde{K}_T^{3.5} = \begin{pmatrix} I_r & 0 & \tilde{\gamma}_1 - \mu_c^{3.} & 0 \\ 0 & T^{-1/2}(B_\perp^\top B_\perp) & 0 & -T^{-1/2}B_\perp^\top C\gamma \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & T^{-1} \end{pmatrix},$$

$$\tilde{K}_T^{4.4} = \begin{pmatrix} I_r & 0 & \tilde{\gamma}_1 - \mu_c^{4.} \\ 0 & T^{-1}\zeta_4^\top & 0 \\ 0 & T^{-1/2}\zeta_{4\perp}^\top(B_\perp^\top B_\perp) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\tilde{K}_T^{4.5} = \begin{pmatrix} I_r & 0 & \tilde{\gamma}_1 - \mu_c^{4\cdot} & 0 \\ 0 & T^{-1/2}(B_\perp^\top B_\perp) & 0 & -T^{-1/2}(B_\perp^\top B_\perp)(0, I_{k-r})\zeta_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & T^{-1} \end{pmatrix},$$

$$\tilde{K}_T^{5.5} = \begin{pmatrix} I_r & -\hat{B}^\top B_\perp & \tilde{\gamma}_1 - \mu_c^{5\cdot} & \hat{\tilde{\delta}}_1 - \tau_c^{5\cdot} \\ 0 & T^{-2}\zeta_3^\top (B_\perp^\top B_\perp) & 0 & 0 \\ 0 & T^{-1/2}\zeta_{3\perp}^\top (B_\perp^\top B_\perp) & 0 & -T^{-1/2}\zeta_{3\perp}^\top (B_\perp^\top B_\perp)(0, I_{k-r})\zeta_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & T^{-1} \end{pmatrix},$$

with

$$\zeta_4 = (B_\perp^\top B_\perp)(0, I_{k-r})\zeta_2 / \|B_\perp (0, I_{k-r})\zeta_2\|^2.$$

For the remaining combinations, define the matrices

$$\begin{aligned} \tilde{K}_T^{i.1} &= \mathrm{diag}(I_r, T^{-1/2}(B_\perp^\top B_\perp)), \\ \tilde{K}_T^{i.2} &= \tilde{K}_T^{i.1}, \\ \tilde{K}_T^{i.3} &= \mathrm{diag}(I_r, T^{-1/2}(B_\perp^\top B_\perp), 1), \\ \tilde{K}_T^{i.4} &= \tilde{K}_T^{i.3}, \end{aligned}$$

and

$$\tilde{K}_T^{i.5} = \mathrm{diag}(I_r, T^{-1/2}(B_\perp^\top B_\perp), 1, T^{-1}).$$

Using Lemma 8.5, the following two lemmas can be established by tedious, but straightforward algebraic manipulations.

Lemma 8.7 *Under the conditions of Lemma 8.4,*

$$\lim_{T\to\infty} K_T^{a.b}((\Lambda^b)^\top, I)^\top \tilde{M}_1^b (K_T^{a.b})^{-1} - \tilde{M}_{2T}^{a.b} = 0,$$

with

$$\begin{aligned} \tilde{M}_1^1 &= (0_{(k-r)\times r}, I_{k-r}), \\ \tilde{M}_1^2 &= (0_{(k-r+1)\times r}, I_{k-r+1}), \\ \tilde{M}_1^3 &= \tilde{M}_1^2, \\ \tilde{M}_1^4 &= (0_{(k-r+2)\times r}, I_{k-r+2}), \\ \tilde{M}_1^5 &= \tilde{M}_1^4, \end{aligned}$$

and

$$\tilde{M}_{2T}^{a.1} = \begin{pmatrix} 0_{r\times r} & 0 \\ 0 & I_{k-r} \end{pmatrix},$$

$$\tilde{M}_{2T}^{a.2} = \begin{pmatrix} 0_{r\times r} & 0 & \tilde{\gamma}_1 - \mu_c^{a\cdot} \\ 0 & I_{k-r} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\begin{aligned}
\tilde{M}_{2T}^{a.3} &= \tilde{M}_{2T}^{a.2}, \\
\tilde{M}_{2T}^{a.4} &= \begin{pmatrix} 0_{r\times r} & 0 & \tilde{\gamma}_1 - \mu_c^{a\cdot} & T(\tilde{\delta}_1 - \tau_c^{a\cdot}) \\ 0 & I_{k-r} & 0 & m_{2T}^{a.} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \\
\tilde{M}_{2T}^{a.5} &= \tilde{M}_{2T}^{a.4}.
\end{aligned}$$

$m_{2T}^{a.} = 0$ *for* $a < 5$, *and*

$$m_{2T}^{5.} = T^{1/2}\zeta_{3\perp}^{\top}(\beta_2\tilde{\delta}_1 - (B_\perp^{\top}B_\perp)\beta_2(B^{\top}B)^{-1}B^{\top}\zeta_2).$$

Proof. As mentioned, the proofs are straightforward. Only note that for case 4.4 it follows from Corollary 8.2 that $\tilde{\delta}_1 = B^{\top}\zeta_2$. This fact is needed to prove that $m_{2T}^{4.} = 0$. □

Lemma 8.8 *Under the conditions of Lemma 8.4,*

$$\underset{T\to\infty}{plim}\ \tilde{K}_T^{a.b}\hat{L}_b(K_T^{a.b})^{-1} = N^{a.b},$$

with

$$N^{1.1} = I_k, \quad N^{i.2} = (I_k, 0_{k,1}), \quad N^{j.3} = I_{k+1},$$
$$N^{i.4} = (I_{k+1}, 0_{k+1,1}), \quad N^{k.5} = I_{k+2},$$

$$N^{3.4} = \begin{pmatrix} I_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \zeta_{1\perp}^{\top} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$N^{4.4} = \begin{pmatrix} I_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \zeta_{4\perp}^{\top} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

for $i = 1, 2$, $j = 1, 2, 3$, *and* $k = 1, \ldots, 5$.

Proof. As mentioned before, the statements follow from straightforward algebraic manipulations. However, the computations are simplified considerably by noting that for $(a.b) \neq (5.5)$, it suffices to check the statements for $\tilde{K}_T^{a.b}L_b(K_T^{a.b})^{-1}$, rather than for $\tilde{K}_T^{a.b}\hat{L}_b(K_T^{a.b})^{-1}$. This can be seen as follows. By Lemma 8.6 and the definition of $\tilde{K}_T^{a.b}$,

$$\tilde{K}_T^{a.b}(\hat{L}_b - L_b)(K_T^{a.b})^{-1} = -\begin{pmatrix} I_r \\ 0 \end{pmatrix}\hat{\Pi}_{11}^{-1}(\hat{\Theta}_1^b - \Theta_1^b)\begin{pmatrix} \Lambda^b \\ I \end{pmatrix}(0, I)(K_T^{a.b})^{-1}.$$

From Lemma 8.4, it follows that

$$\hat{\Pi}_{11}^{-1}(\hat{\Theta}_1^b - \Theta_1^b)(K_T^{a.b})^{-1} = o_p(1),$$

while from Lemma 8.7, it follows that

$$K_T^{a.b}((\Lambda^b)^\top, I)^\top (0, I)(K_T^{a.b})^{-1} = O(1),$$

for $(a.b) \neq 5.5$. This proves the statement. □

The result from Lemma 8.8 is that $\tilde{K}^{a.b}\hat{Y}_{2t}^{a.b}$ and $N^{a.b}K^{a.b}Y_{1t}$ weakly converge to the same limiting process. Because $N^{a.b}$ is often of a simple form, this simplifies the derivations of the asymptotic distribution of $\hat{\Theta}_2^b$ below.

Lemma 8.9 *Under the assumptions of Lemma 8.4,*

$$T^{1/2}(I \otimes (\tilde{K}_T^{a.b})^\top)^{-1}\mathrm{vec}((\hat{\Theta}_2^b - \Theta_2^b)^\top) -$$
$$(\Phi_{22} \otimes N^{a.b}G^{a.b}(N^{a.b})^\top)^{-1}(\Phi_{22} \otimes N^{a.b}G^{a.b})(-\alpha_{21}\alpha_{11}^{-1} \otimes (\tilde{M}_{2T}^{a.b})^\top) \cdot$$
$$(\Phi_{11} \otimes G^{a.b})^{-1} \cdot (\int dW_{21} \otimes (d\xi_1^\top, (g^{a.b})^\top)^\top) \Rightarrow$$
$$(\Phi_{22} \otimes N^{a.b}G^{a.b}(N^{a.b})^\top)^{-1}(\int dW_{22} \otimes N^{a.b}(d\xi_1^\top, (g^{a.b})^\top)^\top). \tag{8.22}$$

Remark 8.2 The notation in (8.22) is nonstandard. Usually, the second term on the left-hand side of the weak convergence symbol would appear on the right-hand side. This can be done for all combinations $(a.b) \neq (5.5)$. For this last case, however, the matrix $\tilde{M}_{2T}^{5.5}$ diverges, as follows from Lemma 8.7. Therefore, the notation as in (8.22) is used.

Proof of Lemma 8.9. The M estimator for the second step regression solves

$$\sum_{t=1}^{T} \phi_{2.}(\hat{\eta}_{2t}^b) \otimes \hat{Y}_{2t}^b = 0. \tag{8.23}$$

The proof mimics that of Lemma 8.4. The remainder terms of the Taylor expansion have the same properties as the remainder term in the proof of Lemma 8.4. Therefore, they can be neglected. Taking a first order Taylor expansion of (8.23) around the disturbance ε_{2t} and using Lemmas 8.7 and 8.8, one obtains

$$\begin{aligned}
0 &= \left(T^{-1}\sum_{t=1}^{T} \phi_{2.}(\varepsilon_{2t}) \otimes \hat{Y}_{2t}^b\right) + \left(T^{-1}\sum_{t=1}^{T} \Phi_{22,t} \otimes \hat{Y}_{2t}^b\right)(\hat{\eta}_{2t}^b - \varepsilon_{2t}) + \tilde{R}_T \Leftrightarrow \\
0 &= \left(T^{-1}\sum_{t=1}^{T} \phi_{2.}(\varepsilon_{2t}) \otimes \tilde{K}_T^{a.b}\hat{L}_b Y_{1t}^b\right) + \left(T^{-1}\sum_{t=1}^{T} \Phi_{22,t} \otimes \tilde{K}_T^{a.b}\hat{L}_b Y_{1t}^b\right) \cdot \\
&\qquad \mathrm{vec}((\hat{\eta}_{2t}^b - \eta_{2t}^b + \eta_{2t}^b - \varepsilon_{2t})^\top) + R_T \\
&= \left(T^{-1}\sum_{t=1}^{T} \phi_{2.}(\varepsilon_{2t}) \otimes \tilde{K}_T^{a.b}\hat{L}_b(K_T^{a.b})^{-1}K_T^{a.b}Y_{1t}^b\right) +
\end{aligned}$$

$$\left(T^{-1}\sum_{t=1}^{T}\Phi_{22,t}\otimes\tilde{K}_T^{a.b}\hat{L}_b(K_T^{a.b})^{-1}K_T^{a.b}Y_{1t}^b\right)\cdot$$
$$\mathrm{vec}(-(\hat{Y}_{2t}^b)^\top(\hat{\Theta}_2^b-\Theta_2^b)^\top+(Y_{1t}^b)^\top(\tilde{M}_1^b)^\top(\hat{\Lambda}^b-\Lambda^b)^\top\alpha_{21}^\top)+R_T$$

$$= \left(T^{-1}\sum_{t=1}^{T}\phi_{2.}(\varepsilon_{2t})\otimes N^{a.b}K_T^{a.b}Y_{1t}^b\right)-$$
$$\left(T^{-1}\sum_{t=1}^{T}\Phi_{22,t}\otimes N^{a.b}K_T^{a.b}Y_{1t}^b(N^{a.b}K_T^{a.b}Y_{1t}^b)^\top\right)\cdot$$
$$(I\otimes(\tilde{K}_T^{a.b})^\top)^{-1}\mathrm{vec}((\hat{\Theta}_2^b-\Theta_2^b)^\top)+$$
$$\left(T^{-1}\sum_{t=1}^{T}\Phi_{22,t}\otimes N^{a.b}K_T^{a.b}Y_{1t}^b(K_T^{a.b}Y_{1t}^b)^\top\right)\cdot$$
$$(\alpha_{21}\otimes((K_T^{a.b})^\top)^{-1}(\tilde{M}_1^b)^\top)\mathrm{vec}((\hat{\Lambda}^b-\Lambda^b)^\top)+o_p(1)$$

$$= \left(T^{-1}\sum_{t=1}^{T}\phi_{2.}(\varepsilon_{2t})\otimes N^{a.b}K_T^{a.b}Y_{1t}^b\right)-$$
$$\left(T^{-1}\sum_{t=1}^{T}\Phi_{22,t}\otimes N^{a.b}K_T^{a.b}Y_{1t}^b(N^{a.b}K_T^{a.b}Y_{1t}^b)^\top\right)\cdot$$
$$(I\otimes(\tilde{K}_T^{a.b})^\top)^{-1}\mathrm{vec}((\hat{\Theta}_2^b-\Theta_2^b)^\top)+$$
$$\left(T^{-1}\sum_{t=1}^{T}\Phi_{22,t}\otimes N^{a.b}K_T^{a.b}Y_{1t}^b(K_T^{a.b}Y_{1t}^b)^\top\right)\cdot$$
$$(-\alpha_{21}\hat{\Pi}_{11}^{-1}\otimes(M_{2T}^{a.b})^\top)(I\otimes(K_T^{a.b})^\top)^{-1}\mathrm{vec}((\hat{\Theta}_1^b-\Theta_1^b)^\top)+o_p(1)$$

with $\tilde{R}_T$ and R_T remainder terms. By applying Lemmas 8.2, 8.3, 8.4, and Theorems 2.1 and 3.3 of Hansen (1992), the lemma is established. □

Under the present assumptions, it is easily checked that

$$\hat{V}\xrightarrow{p}A_\perp^\top\Phi^{-1}E(\phi(\varepsilon_t)\phi(\varepsilon_t)^\top)\Phi^{-1}A_\perp,$$

and

$$\tilde{K}_T^{a.b}\hat{V}_Y^b(\tilde{K}_T^{a.b})^\top\Rightarrow N^{a.b}G^{a.b}N^{a.b},$$

with $\hat{V}$ and $\hat{V}_Y^b$ as defined in Section 8.2. This results in the following lemma.

Lemma 8.10

$$\tilde{K}_T^{a.b}P_b^\top(P_b(\hat{V}_Y^b)^{-1}P_b^\top)^{-1}P_b(\tilde{K}_T^{a.b})^\top\Rightarrow P_b^\top(P_b(G^{a.b})^{-1}P_b^\top)^{-1}P_b.$$

Proof. Only case (5.5) is proved. The other cases follow rather easily. First note that for data generating process 5, it follows from Lemma 8.5 that $(\hat{\beta}_2-\beta_2)=O_p(T^{-1})$ and

$$T^2\zeta_3^\top(B_\perp^\top B_\perp)^{-1}(\hat{\beta}_2-\beta_2)=O_p(T^{-1/2}).$$

The result now follows by noting that

$$\begin{pmatrix}T^2(\zeta_3^\top\zeta_3)^{-1}\zeta_3^\top\\ T^{1/2}(\zeta_{3\perp}^\top\zeta_{3\perp})^{-1}\zeta_{3\perp}^\top\end{pmatrix}(B_\perp^\top B_\perp)^{-1}P_5(\tilde{K}_T^{5.5})^\top=$$

$$\begin{pmatrix} T^2(\zeta_3^\top\zeta_3)^{-1}\zeta_3^\top \\ T^{1/2}(\zeta_{3\perp}^\top\zeta_{3\perp})^{-1}\zeta_{3\perp}^\top \end{pmatrix} \begin{pmatrix} (\hat\beta_2-\beta_2)^\top(B_\perp^\top B_\perp)^{-1} \\ \zeta_3^\top/T^2 \\ \zeta_{3\perp}^\top/T^{1/2} \\ 0 \\ 0 \end{pmatrix}^\top = P_5 + o_p(1).$$

□

Using all the intermediate results derived above, one can now prove Theorem 8.2.

Proof of Theorem 8.2. First, the result is proved for regression model 1 and data generating process 1. From Lemma 8.9, it follows that

$$T(I\otimes(B_\perp^\top B_\perp)^{-1})\mathrm{vec}(\hat\alpha_{22}^\top) = T^{1/2}(I\otimes P_1((\tilde K_T^{1.1})^\top)^{-1})\mathrm{vec}(\hat\Theta_2-\Theta_2)^\top) \Rightarrow$$

$$\begin{aligned}(-\alpha_{21}\alpha_{11}^{-1}\Phi_{11}^{-1}\otimes(P_1G^{1.1}P_1^\top)^{-1})(\textstyle\int dW_{21}\otimes g^{1.1}) \quad + \\ (\Phi_{22}^{-1}\otimes(P_1G^{1.1}P_1^\top)^{-1})(\textstyle\int dW_{22}\otimes g^{1.1}) = \\ (A_\perp^\top\Phi^{-1}\otimes(\textstyle\int\tilde W_1\tilde W_1^\top)^{-1})(\textstyle\int dW_2\otimes\tilde W_1).\end{aligned}$$

Define the normalization matrices

$$\begin{aligned}\Upsilon_1^{1.1} &= (B_\perp^\top C\Omega_{11}C^\top B_\perp)^{-1/2}B_\perp^\top C, \\ \Upsilon_2 &= (A_\perp^\top\Phi^{-1}\Omega_{22}\Phi^{-1}A_\perp)^{-1/2}A_\perp^\top\Phi^{-1}.\end{aligned}$$

Then it follows from lemma 8.9 and 8.10, that

$$\begin{aligned} t_w^1 &= T\mathrm{vec}(\hat\alpha_{22}^\top)^\top(\hat V\otimes P_b(V_Y^b)^{-1}P_b^\top)^{-1}\mathrm{vec}(\hat\alpha_{22}^\top) \\ &\Rightarrow (\textstyle\int dA_\perp^\top\Phi^{-1}W_2\otimes\tilde W_1)^\top(I_{k-r}\otimes(\int\tilde W_1\tilde W_1^\top)^{-1})\cdot \\ &\qquad ((A_\perp^\top\Phi^{-1}\Omega_{22}\Phi^{-1}A_\perp)\otimes(\textstyle\int\tilde W_1\tilde W_1^\top))\cdot \\ &\qquad (I_{k-r}\otimes(\textstyle\int\tilde W_1\tilde W_1^\top)^{-1})(\int dA_\perp^\top\Phi^{-1}W_2\otimes\tilde W_1) \\ &= (\textstyle\int d\Upsilon_2W_2\otimes\Upsilon_1^{1.1}W_1)^\top(I_{k-r}\otimes\int\Upsilon_1^{1.1}W_1(\Upsilon_1^{1.1}W_1)^\top)^{-1}(\int d\Upsilon_2W_2\otimes\Upsilon_1^{1.1}W_1) \\ &= \mathrm{tr}\left((\textstyle\int(\Upsilon_1^{1.1}W_1)d(\Upsilon_2W_2)^\top)^\top(\int\Upsilon_1^{1.1}W_1(\Upsilon_1^{1.1}W_1)^\top)^{-1}(\int(\Upsilon_1^{1.1}W_1)d(\Upsilon_2W_2)^\top)\right). \end{aligned} \tag{8.24}$$

Note that $\Upsilon_1^{1.1}W_1$ and Υ_2W_2 are two correlated, standard Brownian motions. Now let $\Upsilon_3\Upsilon_4\Upsilon_5^\top$ denote the singular value decomposition of $E(\Upsilon_1^{1.1}W_1(1)W_2(1)\Upsilon_2^\top)$. Then the limiting distribution of t_w^b is left unaltered if $\Upsilon_1^{1.1}W_1$ is replaced by $B_1^{1.1} = \Upsilon_3^\top\Upsilon_1^{1.1}W_1$ and Υ_2W_2 is replaced by $B_2 = \Upsilon_5^\top\Upsilon_2W_2$. This proves case 1.1. The other cases are proved similarly, although the computations are more involved. Therefore, I also present the proof of case 3.4, which turns out to be tedious, but straightforward.

Define $\check W_1 = \zeta_{1\perp}^\top\tilde W_1$. Now note that $N^{3.4}G^{3.4}(N^{3.4})^\top = G^{3.3}$, and

$$P_4(N^{3.4}G^{3.4}(N^{3.4})^\top)^{-1}N^{3.4}G^{3.4}(\tilde M_{2T}^{3.4})^\top(G^{3.4})^{-1}(d\xi_1^\top,(g^{3.4})^\top)^\top =$$

$$\begin{pmatrix} \int(s-0.5)^2 & \int(s-0.5)(\check W_1^\mu)^\top \\ \int(s-0.5)(\check W_1^\mu) & \int(\check W_1^\mu)(\check W_1^\mu)^\top \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 & 0 & -\int s \\ 0 & 0 & I_{k-r-1} & -\int\check W_1 \end{pmatrix}.$$

$$\begin{pmatrix} I_r & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \zeta_{1\perp}^\top & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Xi_1 & 0 & 0 & 0 \\ 0 & \int \tilde{W}_1 \tilde{W}_1^\top & \int s \tilde{W}_1 & \int \tilde{W}_1 \\ 0 & \int s \tilde{W}_1^\top & \int s^2 & \int s \\ 0 & \int \tilde{W}_1^\top & \int s & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 \\ I_{k-r+1} \\ 0 \end{pmatrix} \begin{pmatrix} \int \tilde{W}_1^\mu (\tilde{W}_1^\mu)^\top & \int (s-0.5) \tilde{W}_1^\mu \\ \int (s-0.5)(\tilde{W}_1^\mu)^\top & \int (s-0.5)^2 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{W}_1^\mu \\ s-0.5 \end{pmatrix} =$$

$$\begin{pmatrix} \int (s-0.5)^2 & \int (s-0.5)(\check{W}_1^\mu)^\top \\ \int (s-0.5)(\check{W}_1^\mu) & \int (\check{W}_1^\mu)(\check{W}_1^\mu)^\top \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 \\ \zeta_{1\perp}^\top & 0 \end{pmatrix} \cdot$$

$$\begin{pmatrix} \int \tilde{W}_1^\mu (\tilde{W}_1^\mu)^\top & \int (s-0.5) \tilde{W}_1^\mu \\ \int (s-0.5)(\tilde{W}_1^\mu)^\top & \int (s-0.5)^2 \end{pmatrix} \cdot$$

$$\begin{pmatrix} \int \tilde{W}_1^\mu (\tilde{W}_1^\mu)^\top & \int (s-0.5) \tilde{W}_1^\mu \\ \int (s-0.5)(\tilde{W}_1^\mu)^\top & \int (s-0.5)^2 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{W}_1^\mu \\ s-0.5 \end{pmatrix} =$$

$$\begin{pmatrix} \int (s-0.5)^2 & \int (s-0.5)(\check{W}_1^\mu)^\top \\ \int (s-0.5)(\check{W}_1^\mu) & \int (\check{W}_1^\mu)(\check{W}_1^\mu)^\top \end{pmatrix}^{-1} \begin{pmatrix} s-0.5 \\ \check{W}_1^\mu \end{pmatrix},$$

with $\check{W}_1^\mu = \check{W}_1 - \int \check{W}_1$. $\tilde{W}_1^\mu$ is defined analogously. The remainder of the proof follows the proof of case 1.1. □

Proof of Corollary 8.1. For $\phi(\varepsilon_t) = \varepsilon_t$, $\Phi_t = I_k$. Therefore, $R^{1.1} = I_k$. Moreover, the absolute values of the canonical correlations between $\zeta_{i\perp}^\top A_\perp^\top \varepsilon_t$ and $A_\perp^\top \varepsilon_t$ for $i = 1, 3, 4$, are equal to the square roots of the eigenvalues of the matrix $(\tilde{R}^{i.i})(\tilde{R}^{i.i})^\top$, with

$$\tilde{R}^{i.i} = (\zeta_{i\perp}^\top A_\perp^\top \Omega_{11} A_\perp \zeta_{i\perp})^{-1/2} \zeta_{i\perp}^\top A_\perp^\top \Omega_{11} A_\perp (A_\perp^\top \Omega_{11} A_\perp)^{-1/2}.$$

The result now follows by observing that $(\tilde{R}^{i.i})(\tilde{R}^{i.i})^\top = I$. □

Proof of Theorem 8.3. The proof is similar to that of Theorem 8.2 and follows rather directly from Lemma 8.9. For example, consider case 4.5. From Corollary 8.2 it follows that $\tau_c^{4.} = \tilde{C}(1)\delta$. Moreover, because the fourth data generating process is considered, δ can be written as $A\bar{\delta}$ for some vector $\bar{\delta}$ of length r. Considering the first r rows of δ, one obtains

$$\delta_1 = \alpha_{11}\bar{\delta} \;\Leftrightarrow\; -\alpha_{11}\tilde{\delta}_1 = \alpha_{11}\bar{\delta} \;\Leftrightarrow\; \bar{\delta} = -\tilde{\delta}_1.$$

So $\delta = -A\tilde{\delta}_1$. Substituting this into the expression for $\tau_c^{4.}$, it follows from the definition of $\tilde{C}(1)$ in Corollary 8.2, that

$$\tau_c^{4.} = \tilde{C}(1)\delta = (A^\top A)^{-1} A^\top (I - \Psi C) A \tilde{\delta}_1 = \tilde{\delta}_1.$$

As a result, $\tilde{M}_{2T}^{4.5} = \mathrm{diag}(0_{r\times r}, I_{k-r+2})$. Moreover, as $N^{4.5} = I_{k+2}$, the result now follows directly from Lemma 8.9. □

Lemma 8.11 *Given the assumptions stated in Lemma 8.4, only with* $|\alpha_{11}| = 0$ *instead of* $|\alpha_{11}| \neq 0$,

$$T^{1/2}\mathrm{vec}(\hat{\beta}_2 - \beta_2) = O_p(1).$$

Proof. Following the proof of Lemma 8.5 and the result from Lemma 8.4, one needs that the appropriate rows of

$$(\hat{\Pi}_{11}^{-1} \otimes ((\Lambda^b)^\top, I)(K_T^{a.b})^\top)$$

are $O(1)$. As $((\Lambda^b)^\top, I)(K_T^{a.b})^\top = O(T^{-1/2})$, it is therefore sufficient that $T^{1/2}\hat{\Pi}_{11}$ has a nondegenerate limiting distribution. This follows from Lemma 8.4. □

Proof of Theorem 8.4. From Lemma 8.4 it follows that

$$\sqrt{T}\mathrm{vec}(\hat{\Pi}_{11}^\top) \Rightarrow (\Phi_{11}^{-1} \otimes \Xi^{-1})(\int dW_{21} \otimes d\xi_1) = N(0, V_{\phi_1} \otimes \Xi^{-1}),$$

with $V_{\phi_1} = \Phi_{11}^{-1} E(\phi_{1.}(\varepsilon_t)\phi_{1.}(\varepsilon_t)^\top)\Phi_{11}^{-1}$ and $N(0, V)$ denoting a Gaussian random vector with mean zero and variance-covariance matrix V. Note that $\hat{\alpha}_{11} = \hat{\Pi}_{11}$ and

$$\sqrt{T}|\hat{\alpha}_{11}| = \sqrt{T}|\hat{\alpha}_{11}^\top| = \sqrt{T}\mathrm{vec}(|\hat{\alpha}_{11}^\top|\hat{\alpha}_{11}^{-1})^\top \mathrm{vec}(\hat{\alpha}_{11}^\top).$$

Further, $|\hat{\alpha}_{11}^\top|\hat{\alpha}_{11}^{-1}$ is the matrix of cofactors of $\hat{\alpha}_{11}^\top$, which converges in probability to the matrix of cofactors of $\alpha_{11}^\top$, say α_{11}^*. Therefore,

$$\sqrt{T}|\hat{\alpha}_{11}| \xrightarrow{d} N(0, \mathrm{tr}((\alpha_{11}^*)^\top \Xi^{-1} \alpha_{11}^* V_{\phi_1})).$$

It is obvious that $\hat{V}_{\phi_1} \xrightarrow{p} V_{\phi_1}$. Moreover,

$$T\left(\mathrm{tr}((\hat{\alpha}_{11}^\top)^{-1}\hat{V}_\xi^{-1}\hat{\alpha}_{11}^{-1}\hat{V}_{\phi 1})\right)^{-1} = T|\hat{\alpha}_{11}|^2\left(\mathrm{tr}((\hat{\alpha}_{11}^*)^\top \hat{V}_\xi^{-1}\hat{\alpha}_{11}^*\hat{V}_{\phi 1})\right)^{-1} \tag{8.25}$$

with $\hat{\alpha}_{11}^* = |\hat{\alpha}_{11}|\hat{\alpha}_{11}^{-1}$. The result is established if one can show that the denominator in (8.25) converges to $(\alpha_{11}^*)^\top \Xi^{-1}\alpha_{11}^* V_{\phi_1}$. Therefore, one needs to show that $\hat{V}_\xi$ converges to Ξ. This follows from the fact that under the present assumptions it is straightforward to show that $(\tilde{\beta}_2 - \beta_2) = o_p(T^{-1/2})$ (compare Engle and Granger (1987) and Johansen (1991)). □