Tests of Non-nested Hypotheses in Nonstationary Regressions With An Application to Modeling Industrial Production

John C. Chao
University of Maryland

Norman R. Swanson
Pennsylvania State University

this version: June 1998

JEL Classification: C14, C22, C53.

Keywords: Brownian motion, cointegration, conditional model, encompassing, J-test, mixture of normals.

Corresponding Author: Norman R. Swanson, Department of Economics, Pennsylvania State University, University Park, PA, 16802 (email: nswanson@psu.edu). The authors are grateful to the William Barnett and an anonymous associate editor and referee for useful comments and suggestions. Chao thanks the Office of Graduate Studies and Research at the University of Maryland, and Swanson thanks the Research and Graduate Studies Office at Pennsylvania State University and the National Science Foundation (grant number SBR-9730102) for research support.

Running Title: Non-nested Nonstationary Testing

Corresponding Author: Norman R. Swanson, Department of Economics, Pennsylvania State University, 521 Kern Graduate Building, University Park, PA 16802. Phone: (814) 865-2234. Fax: (814) 863-4775. Email: nswanson@psu.edu.

Abstract

In the context of I(1) time series, we provide some asymptotic results for the J-type test proposed by Davidson and MacKinnon (1981). We examine both the case where our regressor sets, x_{1t} and x_{2t}, are not cointegrated, and the case where they are. In the former case, it turns out that the OLS estimator of the weighting coefficient from the artificial compound model converges at rate T to a mixed normal distribution, and the associated t-statistic has an asymptotic standard normal distribution. In the latter case, we find that the J-test also has power against violation of weak exogeneity (with respect to the short-run coefficients of the null model), which is caused by correlation between the disturbance of the null model and that of the cointegrating equation linking x_{1t} and x_{2t}. Moreover, unlike the previous case, the OLS estimator of the weighting coefficient from the artificial compound model converges at $\sqrt{T}$ to an asymptotic normal distribution when the null model is correctly specified. In an empirical illustration we use the tests to examine an industrial production dataset for six countries.

1 Introduction

Contributions in the area of tests of non-nested hypotheses have been frequent in economics in the last two decades. In particular, the pathbreaking work of Pesaran (1974), who applied the generalized likelihood ratio test of Cox (1961, 1962) to select among alternative sets of regressors in the linear regression model, has set off an explosion of research activity which resulted in numerous publications in the 1970's and 1980's. An important contribution subsequent to that of Pesaran (1974) is the paper by Davidson and MacKinnon (1981), where alternatives to the generalized likelihood ratio test are developed. One advantage of the Davidson-MacKinnon tests (e.g. the J-, C-, and P-tests) is that they can be implemented using any standard linear regression package, as they involve the construction of standard t-test statistics from the associated artificial compound regression models. Modifications and generalizations of these tests have been proposed and analyzed in a number of papers, including Pesaran and Deaton (1978), Davidson and MacKinnon (1983), Fisher and McAleer (1981), MacKinnon, White, and Davidson (1983), Ericsson (1983), Godfrey (1983), Vuong (1989), and the references therein, to name only a few. Excellent surveys have been written by MacKinnon (1983), McAleer (1987), Davidson and MacKinnon (1993), and Gourieroux and Monfort (1994).

A parallel literature on encompassing also examines issues related to the choice between rival, possibly non-nested, hypotheses (or models). The encompassing principle entails testing whether one model explains important characteristics of rival models. Two of the most influential papers in this literature are Chong and Hendry (1986) and Mizon and Richard (1986). Chong and Hendry develop tests of forecast encompassing, while Mizon and Richard develop tests of full parameter encompassing and apply their procedures to testing non-nested hypotheses.

In spite of the extensive literature on non-nested testing, there has been little treatment of this subject in the context of time series regression with I(1) regressors. To the best of our knowledge, the only discussion of non-nested testing in the presence of I(1) variables appears in Ericsson (1992) and Clements and Hendry (1994), who suggest implementing Davidson- MacKinnon type tests using an artificial regression model that is "balanced" in terms of the orders of integration of the variables. However, no systematic analysis of the asymptotic properties of Davidson-MacKinnon type tests is given in these papers.

In this paper, we provide some asymptotic results for the J-type test proposed by Davidson and

MacKinnon (1981), in the context of I(1) time series. The examination of J-tests in the context of I(1) regressors should be of some interest to macroeconometricians, given that many macroeconomic time series are thought to be well modeled as integrated processes (see e.g. Nelson and Plosser (1982)). The J-tests which we propose can be used in a number of macroeconomic contexts. For example, in Bernanke, Bohn, and Reiss (1988) a number of models of investment demand are compared using a variety of non-nested tests. Their models of investment given lagged differences in income (the accelerator model) and investment given lagged values of the ratio of the market value of capital to its replacement cost (Tobin's Q model) both involve a mixture of I(1) and I(0) variables, and hence are treatable in the framework which we examine. Another example where our setup is appropriate is the comparison of new classical and Keynesian models, as in Dadkhah and Valbuena (1985), where I(1) regressors are used. Our approach should also be of interest to forecasters who are interested in selecting "optimal" single-equation forecasting models within the context of I(1) and possibly cointegrated variables.

The problem which we consider concerns testing two alternative hypotheses:

$$H_0: \; y_t = \beta' x_{1,t} + \Gamma_1(L)\Delta w_{t-1} + u_{1,t} \;\; \text{and} \;\; H_1: \; y_t = \phi' x_{2,t} + \Psi_1(L)\Delta w_{t-1} + \epsilon_{1,t},$$

where x_{1t} and x_{2t} are vector I(1) processes, y_t is a scalar I(1) process, and Δw_t is a vector of lagged differences. Our analysis centers around the examination of two different cases. In the first case, it is assumed that x_{1t} and x_{2t} are not cointegrated. In this case, it turns out that the OLS estimator of the weighting coefficient from the standard composite artificial model used to construct the J-test converges at rate T to a mixed normal distribution, given an assumption that the null model validly conditions on variables that are weakly exogenous with respect to the parameters of the model. In the more general case where endogeneity of the regressors is allowed under both the null and the alternative models (i.e. the case where valid conditioning does not occur with the initial specification of H_0 and H_1), we show that the problem of selecting among alternative sets of I(1) regressors is better formulated by deriving the correct conditional models via a suitable reparameterization. (See Phillips (1991) and Zivot (1994) for similar derivations of conditional models in the context of estimation and testing of cointegrating vectors.) Implementing the J-test on the basis of the correct conditional models results in the test statistic having the usual asymptotic standard normal distribution.

In the second case, we assume that $x_{1,t}$ and $x_{2,t}$ are cointegrated. In this case, our framework does not correspond to the testing framework examined by Davidson and MacKinnon (1981), as the null model can now be written as the alternative model with nonlinear restrictions on some of the coefficients, and hence, is nested within the alternative model. Thus, in this case our framework corresponds more closely to the standard encompassing setup where a completing model is defined, linking the alternative sets of regressors, $x_{1,t}$ and $x_{2,t}$, as in Lu and Mizon (1991) and Ericsson (1992). Here, out results depend crucially on the correlation between $u_{1,t}$ (i.e. the disturbance term of the null model given above), and the disturbance term of the cointegrating equation linking $x_{1,t}$ and $x_{2,t}$. When the correlation is zero, the OLS estimator of the weighting coefficient from the artificial compound model again converges in probability to zero and has an asymptotic normal distribution when scaled by $\sqrt{T}$. When this correlation is nonzero, direct implementation of the J-test using the models given under H_0 and H_1 leads to the rejection of H_0 with probability approaching one asymptotically. The latter is a sensible result, given that the null model in this case is not the valid conditional model.

In an empirical example, we illustrate the use of the tests by examining a dataset consisting of industrial production figures for six different countries, including the U.S., Canada, France, Germany, Japan, and the U.K. Models of U.S. industrial production are examined, and our findings suggest that: (i) When models with explanatory variable sets including either the U.K., or Canada and Japan are compared, the regression models with Canada and Japan as I(1) regressors are preferred. (ii) When models with explanatory variables sets including either France, Germany, and the U.K. or Canada and Japan are compared, both regression models explain U.S. industrial production, although the group of regressors consisting of Canada and Japan appears to be more important, as evidenced by a weighting coefficient of 0.788 (relative to a weighting coefficient of 0.262 when France, Germany, and the U.K. are added to the artificial regression).

The rest of the paper is organized as follows. Section 2 outlines the framework used in the paper. In Section 3 we examine the asymptotic properties of the J-test. Section 4 provides an empirical illustration. Section 5 concludes and offers recommendations for future research. Technical summaries of our models are provided in Appendix A, while all proofs are gathered in Appendix B. We use the symbols $\Rightarrow$, $\xrightarrow{p}$, and $\equiv$ to denote weak convergence of the associated probability measure, convergence in probability, and equality in distribution, respectively. Also, $I(d)$ signifies a time series that is integrated of order d, $BM(-\ )$ denotes a vector Brownian motion with covariance matrix - , P_X denotes the projection matrix onto the span of X, $M_X = I - P_X$, $\mathbf{0}_a$ denotes an ax1 zero vector, and $\mathbf{0}_{a\text{x}b}$ denotes an axb zero matrix.

2 J-Tests with I(0) and I(1) Variables

We wish to study the choice of regressor problem which can be formulated in terms of the competing hypotheses:

$$H_0: \ y_t = \beta' x_{1,t} + \Gamma_1(L)\Delta w_{t-1} + u_{1,t}, \ t = 1, ..., T \tag{1}$$

$$H_1: \ y_t = \phi' x_{2,t} + \Psi_1(L)\Delta w_{t-1} + \epsilon_{1,t}, \ t = 1, ..., T, \tag{2}$$

where $x_{1,t}$ and $x_{2,t}$ are k_1x1 and k_2x1 vector I(1) processes, respectively, y_t is a scalar I(1) process, $\Delta w_t = (\Delta y_t, \Delta x'_{1,t}, \Delta x'_{2,t})'$, and where $\Gamma_1(L)$ can be partitioned conformably with $\Delta w'_t$ as $(\Gamma_{11}(L), \Gamma_{12}(L), \Gamma_{13}(L))$, with $\Gamma_{1j}(L) = \sum_{k=1}^{p} \Gamma_{1j,k} L^k$ (for j=1,2,3). $\Psi_1(L)$ is defined in the same way as $\Gamma_1(L)$, and for convenience, is also partitioned conformably with $\Delta w'_t$ as $(\Psi_{11}(L), \Psi_{12}(L), \Psi_{13}(L))$. We assume that the disturbances, $u_{1,t}$ and $\epsilon_{1,t}$ are I(0) processes, so that under the null hypothesis, y_t and $x_{1,t}$ are cointegrated in the sense of Engle and Granger (1987), and likewise for y_t and $x_{2,t}$ under the alternative hypothesis. In this setup, the null and alternative hypotheses differ only with regard to the inclusion of the I(1) variables. On the other hand, lagged Δw_ts appear in both models, and are meant to capture the short-run dynamics of the models.

In general, the properties of non-nested tests of the competing hypotheses given by equations (1) and (2) depend crucially upon whether or not $x_{1,t}$ and $x_{2,t}$ are cointegrated. Hence, in this paper we consider alternative completing models, each of which specifies different linkages among $x_{1,t}$ and $x_{2,t}$. Further, these completing models (or, alternatively, the marginal models of $x_{1,t}$ and $x_{2,t}$) are maintained under both the null and the alternative hypothesis. In particular, we consider the following two cases:

Case 1: $x_{1,t}$ and $x_{2,t}$ Not Cointegrated

$$x_{1,t} = x_{1,t-1} + \Gamma_2(L)\Delta w_{t-1} + u_{2,t}, \ t = 1, ..., T \tag{3}$$

$$x_{2,t} = x_{2,t-1} + \Gamma_3(L)\Delta w_{t-1} + u_{3,t}, \ t = 1, ..., T, \tag{4}$$

where $\Gamma_i(L)$ (for i=2,3) can be partitioned conformably with $\Delta w'_t$ as $(\Gamma_{i1}(L), \Gamma_{i2}(L), \Gamma_{i3}(L))$, with $\Gamma_{ij}(L) = \sum_{k=1}^{p} \Gamma_{ij,k} L^k$ (for j=2,3).

Case 2: $x_{1,t}$ **and** $x_{2,t}$ **Cointegrated**

$$x_{1,t} = A'x_{2,t} + \Gamma_2(L)\Delta w_{t-1} + u_{2,t},\ \ t = 1, ..., T \tag{5}$$

$$x_{2,t} = x_{2,t-1} + \Gamma_3(L)\Delta w_{t-1} + u_{3,t},\ \ t = 1, ..., T, \tag{6}$$

where $\Gamma_i(L)$ (for i=2,3) is as defined in Case 1.

Case 1 assumes that there are no cointegrating restrictions between $x_{1,t}$ and $x_{2,t}$. In this case, the null model cannot be written as a restricted version of the alternative model. In this sense, the setup for Case 1 resembles that used by Pesaran (1974) and Davidson and MacKinnon (1981). However, our setup also differs from that studied in these papers in at least two ways. First, we consider regressions with I(1) as well as I(0) regressors. Second, our setup does not rule out the possibility that regressors under either hypothesis may be endogenous, as consistent parameter estimation is possible in regression involving I(1) regressors, even in the presence of endogeneity. In Case 2, on the other hand, the null model can be written as a restricted version of the alternative model. This arises because of the cointegration between $x_{1,t}$ and $x_{2,t}$. In this sense Case 2 corresponds more closely to the setup used for parameter encompassing tests in Lu and Mizon (1991) and Ericsson (1992). Again, however, our focus differs from these authors as we attempt to give a vigorous asymptotic analysis of the properties of the J-test in models with I(1) regressors which may or may not be weakly exogenous for parameters of interest in the null model. It should perhaps be noted that we do not consider the more general case where some components of $x_{1,t}$ and $x_{2,t}$ are cointegrated, while others are not. This is the subject of ongoing research.

Define $u_t = (u_{1,t}, u'_{2,t}, u'_{3,t})' = (u_{1,t}, u_{21,t}, ..., u_{2k_1,t}, u_{31,t}, ..., u_{3k_2,t})'$. Also, let $\Omega = E(u_t u_t')$ be the covariance matrix of u_t, and we can partition Ω conformably with $u_t = (u_{1,t}, u'_{2,t}, u'_{3,t})'$ as:

$$\Omega = \begin{pmatrix} \omega_{11} & \omega'_{21} & \omega'_{31} \\ \omega_{21} & \Omega_{22} & \Omega'_{32} \\ \omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix}$$

We assume the following conditions in our subsequent analysis.

Assumption 1: $u_t \equiv i.i.d.(\mathbf{0}, \Omega)$.

Assumption 2: $E(|u_{1,t}|^{2+\delta}) < \infty$, $E(|u_{2i,t}|^{2+\delta}) < \infty$, i=1, ..., k_1, and $E(|u_{3i,t}|^{2+\delta}) < \infty$, i=1, ..., k_2, for some $\delta > 0$.

3 Asymptotic Properties of the J-Test

3.1 Case 1: $x_{1,t}$ and $x_{2,t}$ Not Cointegrated

We begin by examining the consequences of using the single equation version of the J-Test developed by Davidson and MacKinnon (1981), which is based on the following artificial compound regression model:

$$y = X_1\beta(1-\alpha) + Z_1\Gamma_1'(1-\alpha) + \alpha\hat{y}_{H_1} + \xi_1, \tag{7}$$

where α is a scalar parameter, $y = (y_1, ..., y_T)'$, $X_1 = (x_{1,1}, ..., x_{1,T})'$ (a Txk_1 matrix), $Z_1 = (z_{1,1}, ..., z_{1,T})'$ (a Tx$p(k_1+k_2+1)$ matrix),
$\Gamma_1 = (\Gamma_{11,1}, \Gamma_{12,1}, \Gamma_{13,1}, \ldots, \Gamma_{11,p}, \Gamma_{12,p}, \Gamma_{13,p})$, and $\hat{y}_{H_1} = P_{(X_2,Z_1)}y$, where $X_2 = (x_{2,1}, ..., x_{2,T})'$ (a Txk_2 matrix), and $P_{(X_2,Z_1)}$ is the projection onto the span of the columns of X_2 and Z_1 (so that $\hat{y}_A$ is the fitted value of y under the alternative hypothesis). Note that $z_{1,t}$, $t = 1, ..., T$ are defined in Appendix A.I.

The null hypothesis given in Section 2 corresponds to the hypothesis that $\alpha = 0$. Davidson and MacKinnon (1981) propose testing this null hypothesis using a standard t-statistic. We begin by giving asymptotic results for $\hat{\alpha}_T$, the OLS estimator of α in equation (7), and asymptotic results for the associated t-statistic.

Theorem 3.1.1: Given Assumptions 1, 2, A1, and A2[1], under H_0:

$$\text{(i)}\quad T\hat{\alpha}_T \Rightarrow \frac{S_1+S_2}{S_3} \quad \text{as} \quad T \to \infty, \text{where} \tag{8}$$

$$S_1 = \beta'\int_0^1 B_2(r)B_3(r)'dr[\int_0^1 B_3(r)B_3(r)'dr]^{-1}\int_0^1 B_3(r)dB_{0.23}(r)'$$

$$-\beta'\int_0^1 B_2(r)B_3(r)'dr[\int_0^1 B_3(r)B_3(r)'dr]^{-1}\int_0^1 B_3(r)B_2(r)'dr$$

$$[\int_0^1 B_2(r)B_2(r)'dr]^{-1}\int_0^1 B_2(r)dB_{0.23}(r)'$$

$$S_2 = \beta'\int_0^1 B_2(r)B_3(r)'dr[\int_0^1 B_3(r)B_3(r)'dr]^{-1}[\int_0^1 B_3(r)dB_{2,3}(r)'\underline{\Omega}_{22}^{-1}\underline{\omega}_{21} + \omega_{31}]$$

$$-\beta'\int_0^1 B_2(r)B_3(r)'dr[\int_0^1 B_3(r)B_3(r)'dr]^{-1}\int_0^1 B_3(r)B_2(r)'dr$$

$$[\int_0^1 B_2(r)B_2(r)'dr]^{-1}[\int_0^1 B_2(r)dB_{2,3}(r)'\underline{\Omega}_{22}^{-1}\underline{\omega}_{21}+\omega_{21}]$$

$$S_3=\beta'\int_0^1 B_2(r)B_3(r)'dr[\int_0^1 B_3(r)B_3(r)'dr]^{-1}\int_0^1 B_3(r)B_2(r)'dr\beta$$

$$-\beta'\int_0^1 B_2(r)B_3(r)'dr[\int_0^1 B_3(r)B_3(r)'dr]^{-1}\int_0^1 B_3(r)B_2(r)'dr[\int_0^1 B_2(r)B_2(r)'dr]^{-1}$$

$$\int_0^1 B_2(r)B_3(r)'dr[\int_0^1 B_3(r)B_3(r)'dr]^{-1}\int_0^1 B_3(r)B_2(r)'dr\beta,$$

where, $J_4=(1|\mathbf{0}'_{k_1+k_2})$, $B_{2,3}(r)=(B_2(r)',B_3(r)')'$, $B_{0.2,3}(r)=J_4B_0(r)-\underline{\omega}'_{21}\underline{\Omega}_{22}^{-1}B_{2,3}(r)=\omega_{11.2}^{1/2}W(r)-\underline{\omega}'_{21}\underline{\Omega}_{22}^{-1}B_{2,3}(r)$, $\underline{\omega}_{21}=(\omega'_{21},\omega'_{31})'$, $\omega_{11.2}=\omega_{11}-\underline{\omega}'_{21}\underline{\Omega}_{22}^{-1}\underline{\omega}_{21}$, $W(r)$ is a standard univariate Brownian motion independent of $B_2(r)$ and $B_3(r)$, $B_2(r)$ and $B_3(r)$ are k_1 and k_2 dimensional Brownian motions, respectively, as defined in Appendix A.I, and

$$\boldsymbol{\Omega_{22}}=\begin{pmatrix}\Omega_{22} & \Omega'_{32}\\ \Omega_{32} & \Omega'_{33}\end{pmatrix}.$$

(ii) If in addition, $\omega_{21}=\mathbf{0}_{k_1}$ and $\omega_{31}=\mathbf{0}_{k_2}$, then

$$T\hat{\alpha}_T\Rightarrow\frac{S_1}{S_3}\ \text{ as }\ T\to\infty. \tag{9}$$

Theorem 3.1.2: Given Assumptions 1, 2, A1, A2, $\omega_{21}=\mathbf{0}_{k_1}$, and $\omega_{31}=\mathbf{0}_{k_2}$, under H_0:

$$t_{\hat{\alpha_T}}\Rightarrow N(0,1)\ \text{ as }\ T\to\infty, \tag{10}$$

where $t_{\hat{\alpha_T}}$ is the usual t-statistic for testing the null hypothesis that $\alpha_T=0$ in (7).

Remark 3.1.3: (a) Note that when both ω_{21} and ω_{31} are zero vectors, the asymptotic distribution of $\hat{\alpha}_T$ has the simpler form given by expression (9). Following arguments similar to those used in Phillips (1989), we can write this asymptotic distribution as a mixture of normals. More specifically, we write $S_1/S_3\equiv\int_{V>0}N(0,\omega_{11}V)dP(V)$, where $V=S_3^{-1}$ and P is its associated probability measure. Observe that this asymptotic distribution is symmetric around zero, which shows that $\hat{\alpha}_T$ is asymptotically median unbiased. Moreover, as was noted in Theorem 3.1.2, the usual t-statistic for testing the significance of α in this case has an asymptotic standard normal distribution. Hence, the usual t-statistic can be used for inference, without further modification.

(b) Note that when ω_{21} and/or ω_{31} are nonzero, $\hat{\alpha}_T$ still converges in probability to zero under the null hypothesis, but its asymptotic distribution has a second order bias which is captured by S_2 in Theorem 3.1.1. The presence of this bias term leads to difficulties when conducting statistical inference, as it involves unknown nuisance parameters.

The second order bias arises when ω_{21} and/or ω_{31} are nonzero because the single equation specification given under the null hypothesis does not constitute a valid conditional model of the full cointegrating system given by equations (1), (3), and (4). For example, assume that the parameters of interest are the parameters of the null model (1). Then $x_t = (x'_{1,t}, x'_{2,t})'$ is not in general weakly exogenous with respect to the parameters of interest (in the sense of Engle et al. (1983)) unless both ω_{21} and/or ω_{31} are null vectors. To see this, we first make the following simplifying assumption.

Assumption 1': $u_t \equiv i.i.d.N(\mathbf{0}, -\ )$[2].

Note that following a derivation similar to that given in Phillips (1991), Urbain (1992,1993), Johansen (1992,1995) and Zivot (1994), the conditional null model of y_t given the past and Δx_t under Assumption 1' can be written as:

$$y_t = \beta' x_{1,t-1} + \Pi_2 \Delta x_t + \Pi_1(L)\Delta w_{t-1} + v_{1.2,t},^{3} \tag{11}$$

where $\Pi_2 = \underline{\omega}'_{21}\underline{-}_{22}^{-1}$, and $\Pi_1(L) = \Gamma_1^*(L) - \underline{\omega}'_{21}\underline{-}_{22}^{-1}\underline{\Gamma}(L)$. We can partition $\Pi_1(L)$ conformably with $\Delta w_{t-1} = (\Delta y_{t-1}, \Delta x'_{1,t-1}, \Delta x'_{2,t-1})'$, as $(\Pi_{11}(L), \Pi_{12}(L), \Pi_{13}(L))$, with $\Pi_{1j}(L) = \sum_{k=1}^{p} \Pi_{1j,k} L^k$ (for j=1,2,3). Moreover, we define $v'_{1.2,t} = v'_{1,t} - \underline{v}'_{2,t}\underline{-}_{22}^{-1}\underline{\omega}_{21}$, where $v_{1,t} = u_{1,t} + \beta' u_{2,t}$ and $\underline{v}_{2,t} = (v'_{2,t}, v'_{3,t})' = (u'_{2,t}, u'_{3,t})'$. Now, comparing the conditional model of (11) with the marginal model for x_t given by (3) and (4)), we see that the coefficients of the conditional model will not be variation independent of the parameters of the marginal model unless ω_{21} and/or ω_{31} are null vectors, which is what was stated above.[4]

To overcome the problem of unknown nuisance parameters, we suggest reformulating the null and alternative hypotheses in terms of the appropriate conditional models, and implementing the J-test based on the artificial compound model constructed from the conditional models. Note that if we assume that (3) and (4)) describe the marginal model of x_t under the alternative hypothesis, then in the presence of nonzero correlation between $\epsilon_{1,t}$ and $(u'_{1,t}, u'_{2,t})'$, the conditional model under the alternative (analogous to the conditional null model) can be shown to take the form:

$$y_t = \phi' x_{2,t-1} + \Pi_4 \Delta x_t + \Pi_3(L)\Delta w_{t-1} + \epsilon_{1.2,t}, \tag{12}$$

where the coefficients associated with Δx_t and Δw_{t-1} are functions of the coefficients of $\Psi_1(L)$ from (2) and the covariances between $\epsilon_{1,t}$, $u_{2,t}$, and $u_{3,t}$, and are defined analogously to the coefficients which describe the null conditional model. Now, the competing non-nested hypotheses, (1) and (2), can be reformulated in terms of the conditional models (11) and (12). Thus, the artificial compound regression model for testing the null hypothesis, (11), against the alternative hypothesis, (12), in this context is given by:

$$y = X_{1,-1}\beta(1-\alpha^*) + Z_1^*\Pi^{*'}(1-\alpha^*) + \alpha^*\hat{y}^*_{H_1} + \xi_1^*, \tag{13}$$

where β and y are the same as above, α^* is a scalar parameter, $X_{1,-1} = (x_{1,0}, ..., x_{1,T-1})'$ is a Txk_1 matrix, $Z_1^* = (z^*_{1,1}, ..., z^*_{1,T})'$ is a Txm matrix with $m = (p+1)(k_1+k_2)+p$, $z^*_{1,t} = (\Delta x'_t, \Delta z'_{1,t})'$ $\Pi^* = (\Pi_2, \Pi_{11,1}, \Pi_{12,1}, \Pi_{13,1}, \ldots, \Pi_{11,p}, \Pi_{12,p}, \Pi_{13,p})$, and $\hat{y}^*_{H_1} = P_{(X_{2,-1},Z_1^*)}y$, where $X_{2,-1} = (x_{2,0}, ..., x_{2,T-1})'$ (a Txk_2 matrix). Note that the null hypothesis, (11), can be tested as the restriction that $\alpha^* = 0$ on the artificial compound model, (13). Moreover, note that by reformulating the test in terms of the competing hypotheses, (11) and (12), instead of the hypotheses (1) and (2), we have not changed the problem in any substantive way, since both formulations result in test procedures which select between the alternative sets of I(1) regressors, $x_{1,t}$ and $x_{2,t}$. However, a J-test based on (13) has the advantage that it leads to a null distribution which is free of nuisance parameters, as the next result shows.

Theorem 3.1.4: Given Assumptions 1, 2, A1, and A2, under H_0:

$$\text{(i)}\ \ T\hat{\alpha}^*_T \Rightarrow \frac{S_1^*}{S_3} \ \text{ as } \ T \to \infty, \tag{14}$$

where S_3 is as defined in Theorem 3.1.1, and
$S_1^* = \beta' \int_0^1 B_2(r)B_3(r)'dr \left(\int_0^1 B_3(r)B_3(r)'dr\right)^{-1} \int_0^1 B_3(r)dB^*_{0.2,3}(r)$, with $B^*_{0.2,3}(r) = \omega_{11.2}^{1/2}W(r)$, where $\omega_{11.2}$ is defined in Theorem 3.1.1.

$$\text{(ii)}\ \ t_{\hat{\alpha}^*_T} \Rightarrow N(0,1) \ \text{ as } \ T \to \infty, \tag{15}$$

where $t_{\hat{\alpha}^*_T}$ is the usual t-statistic for testing the null hypothesis that $\hat{\alpha}^*_T = 0$ in the regression given as equation (13).

The asymptotic distribution presented in expression (14) can also be given a Gaussian mixture representation, as follows: $S_1^*/S_3 \equiv \int_{V>0} N(0, \omega_{11.2}V)dP(V)$, where $V = S_3^{-1}$ and P is its associated probability measure. From this expression, it is apparent that $\hat{\alpha}_T$ is asymptotically median unbiased. Moreover, from the result given in part (ii) of Theorem 3.1.4, its t-statistic has an asymptotic standard normal distribution, and thus can be readily used for inference. It is well known from the estimation theory for cointegrated systems that estimators of cointegrating coefficients in a valid conditional model (e.g. in models such as (11) where there is valid conditioning on variables that are weakly exogenous (in the sense of Engle, Hendry, and Richard (1983)) with respect to both β, the long-run parameter vector, and Π^*, the matrix of short-run coefficients of the model) have such desirable properties as being asymptotically median unbiased with asymptotic distributions which are Gaussian mixtures (cf. Phillips (1991) and Phillips and Loretan (1991)). Our findings show that similar results can be obtained in the estimation of artificial regression models involving generated regressors, as long as such models are constructed from models which validly condition on variables which are weakly exogenous with respect to both the long-run and the short-run coefficients of the model. Note further that these results parallel those which we obtained earlier in the case where ω_{21} and ω_{31} were both assumed to be zero vectors (see Theorems 3.1.1(ii) and Theorem 3.1.2). This is not surprising since when $\omega_{21} = \mathbf{0}_{k_1}$ and $\omega_{31} = \mathbf{0}_{k_2}$, the null and the alternative hypotheses are already in conditional model form.

3.2 Case 2: $x_{1,t}$ and $x_{2,t}$ Cointegrated

In this subsection, we again begin by examining the consequences of using the single equation version of the J-Test, but under the assumption that the completing model has $x_{1,t}$ and $x_{2,t}$ cointegrated. The composite artificial regression model in this case is:

$$y = X_1\beta(1-\alpha) + \ddot{Z}_1\Gamma_1'(1-\alpha) + \alpha\hat{y}_{H_1} + \xi_1, \tag{16}$$

where β is a k_1x1 vector, α is a scalar, $y = (y_1, ..., y_T)'$, $X_1 = (x_{1,1}, ..., x_{1,T})'$ is a Txk_1 matrix, $\ddot{Z}_1 = (\ddot{z}_{1,1}, ..., \ddot{z}_{1,T})'$ is a Tx$p(k_1+k_2+1)$ matrix, $\hat{y}_{H_1} = P_{(X_2,\ddot{Z}_1)}y$, and $X_2 = (x_{2,1}, ..., x_{2,T})'$ is a Txk_2 matrix. Note that $\ddot{z}_{1,t}$, $t = 1, ..., T$ are defined in Appendix A.II. As in the previous subsection, we give asymptotic results for $\hat{\alpha}_T$, the OLS estimator of α in equation (16), and asymptotic results for the associated t-statistic.

Theorem 3.2.1:

(i) Given Assumptions 1, 2, A3, and A4[5], under H_0: (a) Suppose $Rank(A) = k_1 \leq k_2$, then:

$$\hat{\alpha}_T \xrightarrow{p} \frac{-\beta'\omega_{21}}{\beta'\text{-}_{22}\beta}, \tag{17}$$

and $t_{\hat{\alpha}_T}$ diverges as $T \to \infty$, for $\beta' \omega_{21} \neq 0$.

(b) Suppose $Rank(A) = k_2 < k_1$, then:

$$\hat{\alpha}_T \xrightarrow{p} \frac{l_1}{l_2} \tag{18}$$

and $t_{\hat{\alpha}_T}$ diverges as $T \to \infty$ for $l_1 \neq 0$, where, $l_1 = -\beta' A'(AA')^{-1} A[I_{k_1} - \text{-}\,_{22} A'_{\perp} (A_{\perp} \text{-}\,_{22} A'_{\perp})^{-1} A_{\perp}] \omega_{21}$ and $l_2 = \beta' A'(AA')^{-1} A \text{-}\,_{22} A'(AA')^{-1} A\beta - \beta' A'(AA')^{-1} A \text{-}\,_{22} A'_{\perp} (A_{\perp} \text{-}\,_{22} A'_{\perp})^{-1} A_{\perp} \text{-}\,_{22} A(AA')^{-1} A\beta$. Also, $A_{\perp}$ is a $(k_1 - k_2)\text{x}k_1$ matrix such that $AA'_{\perp} = \mathbf{0}_{k_2\text{x}(k_1-k_2)}$.

(ii) Assume in addition that $\omega_{21} = \mathbf{0}_{k_1}$, and: (a) Suppose $Rank(A) = k_1 \leq k_2$, then:

$$\sqrt{T}\hat{\alpha}_T \Rightarrow N(\mathbf{0}, \omega_{11}(\beta' \text{-}\,_{22}\beta)^{-1}) \quad \text{and} \quad t_{\hat{\alpha}_T} \Rightarrow \quad N(0,1) \quad \text{as} \quad T \to \infty, \tag{19}$$

(b) Suppose $Rank(A) = k_2 < k_1$, then:

$$\sqrt{T}\hat{\alpha}_T \Rightarrow N(\mathbf{0}, \omega_{11} l_2^{-1}) \quad \text{and} \quad t_{\hat{\alpha}_T} \Rightarrow \quad N(0,1) \quad \text{as} \quad T \to \infty, \tag{20}$$

where l_2 is defined above.

Remark 3.2.2: (a) In the case where $x_{1,t}$ and $x_{2,t}$ are cointegrated, the asymptotic behavior of $\hat{\alpha}_T$ and the associated t-statistic depends critically upon whether or not ω_{21} is a k_1 dimensional zero vector. When $\omega_{21} \neq \mathbf{0}$, $\hat{\alpha}_T$ does not converge to zero in probability under the null hypothesis (except in the unlikely case where $\beta'\omega_{21} = 0$), and the t-statistic rejects the null hypothesis with probability approaching unity asymptotically. This result arises because $\omega_{21} \neq \mathbf{0}$ implies a failure of valid conditioning of the null model, and the rejection which occurs is driven by this misspecification, and not by the validity of the alternative hypothesis. If one chooses to view the J-test as a specification test, as was the position taken by Davidson and MacKinnon (1981,1993), then this is a positive result since it implies that the application of the J-test in this context will have power in directions other than that suggested by the alternative model. (Note, however, that there are cases of failure of valid conditioning in the null model which may not be detected by the J-test, as will be discussed below in Remark 3.2.2(b).)

(b) In the case where $\omega_{21} = \mathbf{0}_{k_1}$, note that $\hat{\alpha}_T$ converges in probability to zero, and when scaled by $\sqrt{T}$ has an asymptotic normal distribution. Thus, the usual least squares t-statistic has an asymptotic standard normal distribution, and can be used to test non-nested hypotheses in the usual way. Interestingly, this statistic will have an asymptotic standard normal distribution regardless of whether $\omega_{31} = \mathbf{0}_{k_2}$, as long as it is true that $\omega_{21} = \mathbf{0}_{k_1}$. Since there is also a failure of valid conditioning in the null model when $\omega_{31} \neq \mathbf{0}_{k_2}$, this points to a potential problem in the application of the J-test in this case, as it is powerless to detect misspecification of this type. To give an intuitive explanation for why this phenomenon occurs, note first that when $\omega_{31} \neq \mathbf{0}_{k_2}$, X_2 is correlated with u_1. This in turn implies that both X_1 and the generated regressor, $\hat{y}_{H_1} = P_{(X_2,\ddot{Z}_1)}y$, which represents the vector of fitted values of the regression under the alternative model, are correlated with u_1. The former is due to the fact that X_1 is a function of X_2 (as a result of the cointegrating relation linking X_1 and X_2). Hence, this particular misspecification affects both the "null model" component and the "alternative model" component of the artificial compound regression model. Since the power of the J-test is likely to be greater the greater the difference between the null and the alternative models, it is unlikely to have much power in detecting misspecifications which affect both hypotheses in approximately the same way.

(c) In contrast to Case 1, reformulating the null and alternative hypotheses in terms of conditional models and applying the J-test to the conditional models does not lead to a sensible approach for selecting among alternative sets of rgeressors, when $\omega_{21} \neq \mathbf{0}_{k_1}$. To see this, note that the conditional models under the respective hypotheses can be written as:

$$y_t = \beta' x_{1,t} + c_1'(x_{1,t} - A' x_{2,t}) + c_2' \Delta x_{2,t} + c_3(L)\Delta w_{t-1} + \eta_{1,t}, \tag{21}$$

$$y_t = \phi' x_{2,t} + d_1'(x_{1,t} - A' x_{2,t}) + d_2' \Delta x_{2,t} + d_3(L)\Delta w_{t-1} + \eta_{2,t}, \tag{22}$$

where $\eta_{1,t}$ and $\eta_{2,t}$ are uncorrelated with the disturbances, $u_{1,t}$ and $u_{2,t}$, of the marginal models for $x_{1,t}$ and $x_{2,t}$ (as given by (5) and (6)), and where $c_3(L)$ and $d_3(L)$ are defined analogous to $\Gamma_1(L)$. Substituting (5) into (21), we see that in the case where $x_{1,t}$ and $x_{2,t}$ are cointegrated, (21) can be rewritten as:

$$y_t = \beta' A' x_{2,t} + (\beta' + c_1')(x_{1,t} - A' x_{2,t}) + c_2' \Delta x_{2,t} + c_3(L)\Delta w_{t-1} + \eta_{1,t}, \tag{23}$$

Note that (23) is what the null model, (21), predicts the alternative model, (22), should find, under the null hypothesis. Comparing (23) with (21), we find that in this case they have the same disturbance term so that even under the presumption that the null hypothesis is true, there is no variance dominance of the H_1 specification by the null model (in the sense of Lu and Mizon (1991) and Ericsson (1992)). In other words, even when H_0 is true, the conditional model, (22), under the alternative hypothesis would fit the data just as well as the conditional null model, (21). It follows that there is also no variance encompassing of the conditional model under the alternative by the conditional model under the null. Hence, to conduct a J-test of the null hypothesis would not be very meaningful here, as the J-test is implicitly a test of variance encompassing (see Davidson and MacKinnon (1981), Lu and Mizon (1991) and Ericsson (1992)). Indeed, an artificial compound model constructed from the conditional models, (21) and (22), would lead to an inoperable J-test in this context, since estimation of the model would result in asymptotic multicollinearity between the fitted values under the alternative, and the remaining regressors.

4 Empirical Illustration: Industrial Production

In order to illustrate how the tests discussed above can be applied, we examine a dataset consisting of industrial production indexes for seven different countries for the period 1970:1-1989:12. The countries are: USA (US_t), United Kingdom (UK_t), Japan (JA_t), Germany (GE_t), France (FR_t), and Canada (CA_t). Complete details of the dataset are in Filardo (1993). Using the notation above, let $y_t = US_t$. We consider two different examples. In the first example (Example 1), $x_{1,t} = (CA_t, JA_t)'$, and $x_{2,t} = UK_t$. This example is interesting, as it might be expected that the model with $x_{1,t}$ will be the preferred regression model, given the relative importance of Canada and Japan to U.S. trade.[7] In our second example (Example 2), we set $x_{1,t} = (CA_t, JA_t)'$, and $x_{2,t} = (UK_t, GE_t, FR_t)$. In this example, the J-test can be used as a test of whether or not long-run U.S. industrial production is better modelled using only the largest European economies, or only Canada and Japan. Of course, one possible outcome is that neither model is preferred, in which case we have evidence that all of these countries have an important long-run impact on U.S. industrial production. It should perhaps be noted that two examples which we examine can be justified within the context of choosing a forecasting model for U.S. industrial production, for example. In particular, given some belief in common world business cycle components (such as common stochastic trends), it might be argued that forecasting models of U.S. industrial production should contain information from "trading partners" such as Japan and Canada.

Before estimating various linear models and constructing J-test statistics, we first used the model selection approach of Corradi and Swanson (1997) to simultaneously choose between levels and logs and between I(1) and I(0). We found that all of the variables examined here are I(1) in levels around a linear deterministic trend. Then, using the AIC to select the number of lags to include in each model, cointegration was tested for. In both examples, there was evidence of cointegration (based on the Johansen trace test) between y_t and both $x_{1,t}$ and $x_{2,t}$. However, any evidence of cointegration between $x_{1,t}$ and $x_{2,t}$ was weak in all cases. For this reason, we assume that $x_{1,t}$ and $x_{2,t}$ are not cointegrated. Detailed unit root and cointegration test results as well as computer programs for selecting between levels and logs are available from the authors.[8]

Example 1: Assume weak exogeneity holds (see above discussion), so that equation (7) can be used to construct the J-test. In this case, when $x_{1,t}$ is used in the null model, and $x_{2,t}$ is used in the construction of $\hat{y}_{H_1}$, then $\hat{\alpha}_T = 0.047$ and $t_{\hat{\alpha}_T} = 1.766$. This suggests a failure to reject H_0 at a 5% level. Furthermore, when $x_{2,t}$ is used in the null model, and $x_{1,t}$ is used in the construction of $\hat{y}_{H_1}$, then $\hat{\alpha}_T = 0.965$ and $t_{\hat{\alpha}_T} = 41.81$. This suggests rejecting H_1. Thus, under the maintained assumption of weak exogeneity, the model with $x_{1,t}$ which includes CA_t and JA_t as regressors rather than UK_t, is preferred, regardless of which model is maintained under the null. Given that weak exogeneity is maintained under the null, the failure to reject H_0 can be viewed as evidence that weak exogeneity does indeed hold in the data. In order to further examine the validity of such a statement, we assumed that weak exogeneity does not hold, and ran the appropriate regression, equation (13). In this case, $\hat{\alpha}_T = 0.046$ and $t_{\hat{\alpha}_T} = 1.634$ when $x_{1,t}$ is used as the null model, and $\hat{\alpha}_T = 0.966$ and $t_{\hat{\alpha}_T} = 39.94$ when $x_{2,t}$ is used as the null model, in agreement with our above findings. Finally, notice that since weak exogeneity holds in this example, our results are unchanged if there is cointegration between $x_{1,t}$ and $x_{2,t}$ (see Theorem 3.1.1, 3.1.2, and 3.2.1(ii)).

Example 2: Assume weak exogeneity holds, so that equation (7) can be used to construct the J-test. In this case, when $x_{1,t}$ is used in the null model, and $x_{2,t}$ is used in the construction of $\hat{y}_{H_1}$, then $\hat{\alpha}_T = 0.239$ and $t_{\hat{\alpha}_T} = 6.962$. This suggests rejecting the model with $x_{1,t}$ at a 5% level. Alternatively, when $x_{2,t}$ is used in the null model, then $\hat{\alpha}_T = 0.810$ and $t_{\hat{\alpha}_T} = 25.58$. This suggests rejecting the model with $x_{2,t}$. Thus, under the maintained assumption of weak exogeneity, both regression models appear to explain the data. However, it is perhaps worth noting that $\hat{\alpha}_T$ is much larger when $\hat{y}_{H_1}$ is constructed using current values of $x_{1,t}$, suggesting that Canada and Japan influence U.S. industrial production more than do the three European countries which comprise $x_{2,t}$. In order to further examine these findings, we assumed that weak exogeneity does not hold, and ran the augmented model, equation (13). In this case, $\hat{\alpha}_T = 0.262$ and $t_{\hat{\alpha}_T} = 6.877$ when $x_{1,t}$ is used in the null model, and $\hat{\alpha}_T = 0.788$ and $t_{\hat{\alpha}_T} = 22.20$ when $x_{2,t}$ is used in the null. As in Example 1, these findings correspond with our finding when weak exogeneity is assumed to hold.

5 Conclusions

In this paper we have examined some asymptotic properties of the Davidson-MacKinnon J-test in the context of time series regression with I(1) variables. The work reported here is merely a starting point. A wide variety of further questions present themselves for subsequent research, both theoretical and empirical. On the theoretical side it remains to derive the properties of tests related to the J-test (e.g. encompassing and forecast encompassing tests) in the context of nonstationary regression. Also, it is of interest to analyze the properties of these tests when subsets of the set of regressors under the null (and/or alternative) hypothesis are themselves cointegrated and have some overlapping elements. On the empirical side, it remains to examine the finite sample properties of non-nested tests of hypotheses within the context of nonstationary regression, and to construct non-nested tests of alternative long-run cointegrating restrictions within macroeconomic models.

6 Appendix A: Technical Background

6.1 Appendix A.I: Case 1: $x_{1,t}$ and $x_{2,t}$ not Cointegrated

In this subsection, we give some alternative representations of the model presented in Section 2, Case 1, under the null hypothesis. Also, some preliminary Lemmas useful for proving our main results are given. Note first that the system given by expressions (1), (3), and (4) can alternatively be written in the error-correction form:

$$\Delta w_t = \Gamma^*(L)\Delta w_{t-1} + CBw_{t-1} + F_1 u_t, \tag{24}$$

where $u_t = (u_{1,t}, u'_{2,t}, u'_{3,t})'$. Let $m_1 = k_1 + k_2 + 1$, $m_2 = (p-1)\text{x}m_1 + 1$, $m_3 = (p-1)\text{x}m_1 + k_2 + 1$, and note that the parameter matrices in (24) are: $C' = -(1, \mathbf{0}'_{k_1}, \mathbf{0}'_{k_2})$, $B = (1, -\beta', \mathbf{0}'_{k_2})$,

$$\mathbf{F_1} = \begin{pmatrix} 1 & \beta' & \mathbf{0}'_{k_2} \\ \mathbf{0}_{k_1} & I_{k_1} & \mathbf{0}_{k_1\text{x}k_2} \\ \mathbf{0}_{k_2} & \mathbf{0}_{k_2\text{x}k_1} & I_{k_2} \end{pmatrix}, \text{and } \mathbf{\Gamma^*(L)} = \begin{pmatrix} \Gamma^*_{11}(L) & \Gamma^*_{12}(L) & \Gamma^*_{13}(L) \\ \Gamma^*_{21}(L) & \Gamma^*_{22}(L) & \Gamma^*_{23}(L) \\ \Gamma^*_{31}(L) & \Gamma^*_{32}(L) & \Gamma^*_{33}(L) \end{pmatrix},$$

where $\Gamma^*_{1j}(L) = \Gamma_{1j}(L) + \beta'\Gamma_{2j}(L)$, $\Gamma^*_{2j}(L) = \Gamma_{2j}(L)$, and $\Gamma^*_{3j}(L) = \Gamma_{3j}(L)$, for j=1,2,3. To ensure that Δw_t is I(1) with a Wold representation, Bw_{t-1} is an I(0) process, and w_t is an I(1) process, we impose the following conditions on (24):

Assumption A1: $|I_{m_1} - \Gamma^*(L)L| = 0$ implies $|L| > 1$.

Assumption A2: $C'_\perp(\Gamma^*(1) - I_{m_1})B_\perp$ is nonsingular, where $C_\perp$ and $B_\perp$ are $m_1\text{x}(m_1-1)$ matrices of full column rank, such that $C'_\perp C = \mathbf{0}_{m_1} = B'_\perp B'$.

Next, to analyze the statistical properties of this system, it is convenient to write it in the first order companion form:

$$z_{t+1} = H_1 z_t + G_1 F_1 u_t, \tag{25}$$

where $z_t = (\Delta w'_{t-1}, ..., \Delta w'_{t-p}, w'_{t-1}B)' = (z'_{1,t}, w'_{t-1}B)$, $G'_1 = (e'_p \quad I_{m_1}|B)$, and

$$\mathbf{H_1} = \begin{pmatrix} \Gamma^*_{.,1} & \Gamma^*_{.,2} & \cdots & \Gamma^*_{.,p-1} & \Gamma^*_{.,p} & C \\ I_{m_1} & \mathbf{0}_{m_1\text{x}m_1} & \cdots & \mathbf{0}_{m_1\text{x}m_1} & \mathbf{0}_{m_1\text{x}m_1} & \mathbf{0}_{m_1\text{x}m_1} \\ \mathbf{0}_{m_1\text{x}m_1} & I_{m_1} & \cdots & \mathbf{0}_{m_1\text{x}m_1} & \mathbf{0}_{m_1\text{x}m_1} & \mathbf{0}_{m_1\text{x}m_1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0}_{m_1\text{x}m_1} & \mathbf{0}_{m_1\text{x}m_1} & \cdots & I_{m_1} & \mathbf{0}_{m_1\text{x}m_1} & \mathbf{0}_{m_1\text{x}m_1} \\ B'\Gamma^*_{.,1} & B'\Gamma^*_{.,2} & \cdots & B'\Gamma^*_{.,p-1} & B'\Gamma^*_{.,p} & B'C+1 \end{pmatrix}.$$

Here,

$$\mathbf{\Gamma^*_{.,k}} = \begin{pmatrix} \Gamma^*_{11,k} & \Gamma^*_{12,k} & \Gamma^*_{13,k} \\ \Gamma^*_{21,k} & \Gamma^*_{22,k} & \Gamma^*_{23,k} \\ \Gamma^*_{31,k} & \Gamma^*_{32,k} & \Gamma^*_{33,k} \end{pmatrix},$$

for $k = 1, ..., p$, given that $\Gamma^*_{ij,k}$ (i,j = 1,2,3) are the coefficients of the lag polynomial $\Gamma^*_{ij}(L) = \sum_{k=1}^{p} \Gamma^*_{ij,k} L^k$. Given that z_t is $I(0)$ (by assumption), all of the eigenvalues of H_1 have modulus less than unity, the following moving average representation of (25) thus exists:

$$z_t = \Theta_1(L) G_1 F_1 u_{t-1}, \tag{26}$$

where $\Theta_1(L) = \sum_{j=0}^{\infty} H_1^j L^j$. It is now convenient to write Lemmas which are useful in our subsequent asymptotic analysis. Let $J_1 = (I_{pm_1}|\mathbf{0}_{pm_1})$, $J_2 = (\mathbf{0}_{k_1}|I_{k_1}|\mathbf{0}_{k_1 \mathrm{x} m_3})$ $J_3 = (\mathbf{0}_{k_2 \mathrm{x}(k_1+1)}|I_{k_2}|\mathbf{0}_{k_2 \mathrm{x} m_2})$, $J_4 = (1|\mathbf{0}'_{k_1+k_2})$, $J_5 = (\mathbf{0}_{k_1}|I_{k_1}|\mathbf{0}_{k_1 \mathrm{x} k_2})$, and $J_6 = (\mathbf{0}_{k_2}|\mathbf{0}_{k_2 \mathrm{x} k_1}|I_{k_2})$, and $\Lambda_1 = \sum_{j=0}^{\infty} H_1^j G_1 F_1 \text{-}\ F_1' G_1' H_1^{j'}$.

Lemma A.I.1: For the model discussed above with $u_t \equiv i.i.d.(\mathbf{0}, \text{-}\ )$, the following convergence results hold as $T \to \infty$:

(a) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \Rightarrow B_0(r) \equiv BM(\text{-}\ )$

(b) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} z_{1,t+1} \Rightarrow J_1 \Theta_1(1) G_1 F_1 B_0(r) \equiv B_1(r)$

(c) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \Delta x_{1,t} \Rightarrow J_2 \Theta_1(1) G_1 F_1 B_0(r) \equiv B_2(r)$

(d) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \Delta x_{2,t} \Rightarrow J_3 \Theta_1(1) G_1 F_1 B_0(r) \equiv B_3(r)$

(e) $\frac{1}{T} \sum_{t=1}^{T} z_t z_t' \xrightarrow{p} \Lambda_1$

(f) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (z_t \quad u_t) \Rightarrow N(\mathbf{0}, \Lambda_1 \quad \text{-}\ )$

Proof: Parts (a), (e), and (f) follow using the same arguments used by Toda and Phillips (1993) (see also Theorem 2.2 of Chan and Wei (1988)). Also, using the argument of Lemma 1 from Toda and Phillips (1993), we have that $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} z_t \Rightarrow \Theta_1(1) G_1 F_1 B_0(r)$. It follows that $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} z_{1,t+1} = J_1 \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} z_t \Rightarrow J_1 \Theta_1(1) G_1 F_1 B_0(r) \equiv B_1(r)$, as required for (b). Parts (c) and (d) follow using the same arguments.

Lemma A.I.2: For the model discussed above with $u_t \equiv i.i.d.(\mathbf{0}, \text{-}\ )$, the following convergence results hold as $T \to \infty$:

(a) $\frac{1}{T^2} \sum_{t=1}^{T} x_{1,t} x_{1,t}' \Rightarrow \int_0^1 B_2(r) B_2(r)' dr$

(b) $\frac{1}{T^2} \sum_{t=1}^{T} x_{1,t} x_{2,t}' \Rightarrow \int_0^1 B_2(r) B_3(r)' dr$

(c) $\frac{1}{T^2} \sum_{t=1}^{T} x_{2,t} x_{2,t}' \Rightarrow \int_0^1 B_3(r) B_3(r)' dr$

(d) $\frac{1}{T} \sum_{t=1}^{T} x_{1,t} z_{1,t}' \Rightarrow \int_0^1 B_2(r) dB_1(r)' + \Delta_1 + \Delta_2$

where $\Delta_1 = E(\Delta x_{1,t} z_{1,t}')$ and $\Delta_2 = \sum_{j=1}^{\infty} E(\Delta x_{1,t} z_{1,t+j}')$.

(e) $\frac{1}{T} \sum_{t=1}^{T} x_{2,t} z_{1,t}' \Rightarrow \int_0^1 B_3(r) dB_1(r)' + \Delta_3 + \Delta_4$

where $\Delta_3 = E(\Delta x_{2,t} z_{1,t}')$ and $\Delta_4 = \sum_{j=1}^{\infty} E(\Delta x_{2,t} z_{1,t+j}')$.

(f) $\frac{1}{T} \sum_{t=1}^{T} x_{1,t} u_{1,t} \Rightarrow \int_0^1 B_2(r) dB_0(r)' J_4' + \omega_{21}$

(g) $\frac{1}{T}\sum_{t=1}^{T} x_{2,t}u_{1,t} \Rightarrow \int_0^1 B_3(r)dB_0(r)'J_4' + \omega_{31}$

(h) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(z_{1,t}u_{1,t}) \Rightarrow N(\mathbf{0}, (J_1\Lambda_1 J_1' \quad J_4\text{-}\ J_4'))$

(i) $\frac{1}{T}\sum_{t=1}^{T} z_{1,t}z_{1,t}' \xrightarrow{p} J_1\Lambda_1 J_1'$

Proof: All results follow directly from Lemma A.I.1, the continuous mapping theorem, and arguments analogous to those used in Lemma 2.1 of Phillips and Park (1989).

6.2 Appendix A.II: Case 2: $x_{1,t}$ and $x_{2,t}$ Cointegrated

In this subsection, we give some alternative representations of the model presented in Section 2, Case 2, under the null hypothesis. Also, some preliminary Lemmas useful for proving our main results are given. Note first that the system given by expressions (1), (5), and (6) can alternatively be written in the error-correction form:

$$\Delta w_t = \bar{\Gamma}(L)\Delta w_{t-1} + \bar{C}\bar{B}w_{t-1} + F_2 u_t, \tag{27}$$

where $u_t = (u_{1,t}, u_{2,t}', u_{3,t}')'$. Let m_1-m_3 be as above, and $m_4 = m_2 + k_1 1$, and note that the parameter matrices in (27) are:

$$\bar{\mathbf{C}}' = -\begin{pmatrix} 1 & \mathbf{0}_{k_1}' & \mathbf{0}_{k_2}' \\ \mathbf{0}_{k_1} & I_{k_1} & \mathbf{0}_{k_1 \mathrm{x} k_1} \end{pmatrix} \quad \bar{\mathbf{B}} = \begin{pmatrix} 1 & \mathbf{0}_{k_1}' & -\beta' A' \\ \mathbf{0}_{k_1} & I_{k_1} & -A' \end{pmatrix}$$

$$\bar{\mathbf{F}}_\mathbf{2} = \begin{pmatrix} 1 & \beta' & \beta' A' \\ \mathbf{0}_{k_1} & I_{k_1} & A' \\ \mathbf{0}_{k_2} & \mathbf{0}_{k_2 \mathrm{x} k_1} & I_{k_2} \end{pmatrix}, \text{and } \bar{\mathbf{\Gamma}}(\mathbf{L}) = \begin{pmatrix} \bar{\Gamma}_{11}(L) & \bar{\Gamma}_{12}(L) & \bar{\Gamma}_{13}(L) \\ \bar{\Gamma}_{21}(L) & \bar{\Gamma}_{22}(L) & \bar{\Gamma}_{23}(L) \\ \bar{\Gamma}_{31}(L) & \bar{\Gamma}_{32}(L) & \bar{\Gamma}_{33}(L) \end{pmatrix},$$

where $\bar{\Gamma}_{1j}(L) = \Gamma_{1j}(L) + \beta'\Gamma_{2j}(L) + \beta' A'\Gamma_{3j}(L)$, $\bar{\Gamma}_{2j}(L) = \Gamma_{2j}(L) + A'\Gamma_{3j}(L)$, and $\bar{\Gamma}_{3j}(L) = \Gamma_{3j}(L)$, for j=1,2,3. To ensure that Δw_t is I(1) with a Wold representation, $\bar{B}w_{t-1}$ is an I(0) process, and w_t is an I(1) process, we impose the following conditions on (27):

Assumption A3: $|I_{m_1} - \bar{\Gamma}^*(L)L| = 0$ implies $|L| > 1$.

Assumption A4: $\bar{C}_\perp'(\bar{\Gamma}^*(1) - I_{m_1})\bar{B}_\perp$ is nonsingular, where $\bar{C}_\perp$ and $\bar{B}_\perp$ are m_1x$(m_1 - k_1 - 1)$ matrices of full column rank, such that $\bar{C}_\perp'\bar{C} = \mathbf{0}_{m_1} = \bar{B}_\perp'\bar{B}'$.

Again, it is convenient to write this system in the first order companion form:

$$\ddot{z}_{t+1} = H_2\ddot{z}_t + G_2 F_2 u_t, \tag{28}$$

where $\ddot{z}_t = (\Delta w'_{t-1}, ..., \Delta w'_{t-p}, w'_{t-1}\bar{B})' = (\ddot{z}'_{1,t}, w'_{t-1}\bar{B})$, $G'_1 = (e'_p \quad I_{m_1}|\bar{B})$, and

$$\mathbf{H_2} = \begin{pmatrix} \bar{\Gamma}_{.,1} & \bar{\Gamma}_{.,2} & \dots & \bar{\Gamma}_{.,p-1} & \bar{\Gamma}_{.,p} & \bar{C} \\ I_{m_1} & \mathbf{0}_{m_1 \mathrm{x} m_1} & \dots & \mathbf{0}_{m_1 \mathrm{x} m_1} & \mathbf{0}_{m_1 \mathrm{x} m_1} & \mathbf{0}_{m_1 \mathrm{x} m_1} \\ \mathbf{0}_{m_1 \mathrm{x} m_1} & I_{m_1} & \dots & \mathbf{0}_{m_1 \mathrm{x} m_1} & \mathbf{0}_{m_1 \mathrm{x} m_1} & \mathbf{0}_{m_1 \mathrm{x} m_1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{0}_{m_1 \mathrm{x} m_1} & \mathbf{0}_{m_1 \mathrm{x} m_1} & \dots & I_{m_1} & \mathbf{0}_{m_1 \mathrm{x} m_1} & \mathbf{0}_{m_1 \mathrm{x} m_1} \\ \bar{B}'\bar{\Gamma}_{.,1} & \bar{B}'\bar{\Gamma}_{.,2} & \dots & \bar{B}'\bar{\Gamma}_{.,p-1} & \bar{B}'\bar{\Gamma}_{.,p} & \bar{B}'\bar{C} + I_{k_1+1} \end{pmatrix}.$$

Here,

$$\bar{\mathbf{\Gamma}}^*_{\mathbf{.,k}} = \begin{pmatrix} \bar{\Gamma}_{11,k} & \bar{\Gamma}_{12,k} & \bar{\Gamma}_{13,k} \\ \bar{\Gamma}_{21,k} & \bar{\Gamma}_{22,k} & \bar{\Gamma}_{23,k} \\ \bar{\Gamma}_{31,k} & \bar{\Gamma}_{32,k} & \bar{\Gamma}_{33,k} \end{pmatrix},$$

for $k = 1, ..., p$, given that $\bar{\Gamma}_{ij,k}$ (i,j = 1,2,3) are the coefficients of the lag polynomial $\bar{\Gamma}_{ij}(L) = \sum_{k=1}^{p} \bar{\Gamma}_{ij,k} L^k$. Given that $\ddot{z}_t$ is $I(0)$ (by assumption), all of the eigenvalues of H_1 have modulus less than unity, the following moving average representation of (28) thus exists:

$$\ddot{z}_t = \Theta_2(L) G_2 F_2 u_{t-1}, \tag{29}$$

where $\Theta_2(L) = \sum_{j=0}^{\infty} H_2^j L^j$. It is now convenient to write Lemmas which are useful in our subsequent asymptotic analysis. Let $\Lambda_2 = \sum_{j=0}^{\infty} H_2^j G_2 F_2$- $F'_2 G'_2 H_2^{j'}$, $\bar{J}_1 = (I_{pm_1}|\mathbf{0}_{pm_1 \mathrm{x}(k_1+1)})$ and $\bar{J}_2 = (\mathbf{0}_{k_2 \mathrm{x}(k_1+1)}|I_{k_2}|\mathbf{0}_{k_2 \mathrm{x} m_4})$.

Lemma A.II.1: For the model discussed above with $u_t \equiv i.i.d.(\mathbf{0}, \text{-}\ )$, the following convergence results hold as $T \to \infty$:

(a) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} u_t \Rightarrow B_0(r) \equiv BM(\text{-}\ )$

(b) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \ddot{z}_{1,t} \Rightarrow \bar{J}_1 \Theta_2(1) G_2 F_2 B_0(r) \equiv B_4(r)$

(c) $\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \Delta x_{2,t-1} \Rightarrow \bar{J}_2 \Theta_2(1) G_2 F_2 B_0(r) \equiv B_5(r)$

(d) $\frac{1}{T} \sum_{t=1}^{T} \ddot{z}_t \ddot{z}'_t \xrightarrow{p} \Lambda_2$

(e) $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\ddot{z}_t \quad u_t) \Rightarrow N(\mathbf{0}, \Lambda_2 \quad \text{-}\ )$

Proof: The proof follows using the same arguments as those used for the proof of Lemma A.I.1.

Lemma A.II.2: For the model discussed above with $u_t \equiv i.i.d.(\mathbf{0}, \text{-}\ )$, the following convergence results hold as $T \to \infty$:

(a) $\frac{1}{T^2} \sum_{t=1}^{T} x_{2,t-1} x'_{2,t-1} \Rightarrow \int_0^1 B_5(r) B_5(r)' dr$

(b) $\frac{1}{T} \sum_{t=1}^{T} x_{2,t-1} u'_t \Rightarrow \int_0^1 B_5(r) dB_0(r)'$

(c) $\frac{1}{T} \sum_{t=1}^{T} x_{2,t-1} v'_t \Rightarrow \int_0^1 B_5(r) dB_0(r)' F'_2$ where $v_t = F_2 u_t$

(d) $\frac{1}{T}\sum_{t=1}^{T} x_{2,t-1}\ddot{z}_{1,t}' \Rightarrow \int_0^1 B_5(r)dB_4(r)' \ + \bar{\Delta}_1 + \bar{\Delta}_2$

where $\bar{\Delta}_1 = E(\Delta x_{2,t-1}\ddot{z}_{1,t}')$ and $\bar{\Delta}_2 = \sum_{j=1}^{\infty} E(\Delta x_{2,t-1}\ddot{z}_{1,t+j}')$.

(e) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\ddot{z}_{1,t}u_t') \Rightarrow N_0$ where $vec(N_0) \equiv N(\mathbf{0}, (\bar{J}_1\Lambda_2\bar{J}_1' \quad \text{-} \ ))$

(f) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\ddot{z}_{1,t}v_t' \Rightarrow N_1$ where $vec(N_1) \equiv N(\mathbf{0}, (\bar{J}_1\Lambda_2\bar{J}_1' \quad F_2\text{-}\ F_2'))$

(g) $\frac{1}{T}\sum_{t=1}^{T}\ddot{z}_{1,t}\ddot{z}_{1,t}' \xrightarrow{p} \bar{J}_1\Lambda_2\bar{J}_1'$

Proof: All results follow directly from Lemma A.I.1, the continuous mapping theorem, and arguments analogous to those used in Lemma 2.1 of Phillips and Park (1989).

7 Appendix B: Proofs

Proof of Theorem 3.1.1: (i) Anticipating the appropriate rate of convergence, we make use of the usual regression algebra, and write T times the OLS estimator of $\hat{\alpha}_T$ from equation (7) as:

$$T\hat{\alpha}_T = \frac{y'P_{(X_2,Z_1)}M_{(X_1,Z_1)}y/T}{y'P_{(X_2,Z_1)}M_{(X_1,Z_1)}P_{(X_2,Z_1)}y/T^2}, \tag{30}$$

where y, X_1, X_2, Z_1, and $P_{(X_2,Z_1)}$ are as defined in Section 3.1, and where $M_{(X_1,Z_1)}$ is a TxT matrix which projects onto the orthogonal complement of the span of the columns of X_1 and Z_1. First, consider the denominator of (30). Using Lemma A.I.2, the continuous mapping theorem, and following standard arguments, note that

$$y'P_{(X_2,Z_1)}M_{(X_1,Z_1)}P_{(X_2,Z_1)}y/T^2 = \frac{\beta' X_1'M_{Z_1}X_2}{T^2}\left(\frac{X_2'M_{Z_1}X_2}{T^2}\right)^{-1}\frac{X_2'M_{Z_1}X_1\beta}{T^2}$$

$$-\frac{\beta' X_1'M_{Z_1}X_2}{T^2}\left(\frac{X_2'M_{Z_1}X_2}{T^2}\right)^{-1}\frac{X_2'M_{Z_1}X_1}{T^2}\left(\frac{X_1'M_{Z_1}X_1}{T^2}\right)^{-1}$$

$$\frac{X_1'M_{Z_1}X_2}{T^2}\left(\frac{X_2'M_{Z_1}X_2}{T^2}\right)^{-1}\frac{X_2'M_{Z_1}X_1\beta}{T^2} + O_p(T^{-1}) \Rightarrow S_3, \tag{31}$$

where S_3 is as defined in the statement of Theorem 3.1.1. With respect to the numerator of (30), note again that by Lemma A.I.2 and the continuous mapping theorem:

$$y'P_{(X_2,Z_1)}M_{(X_1,Z_1)}y/T = \frac{\beta' X_1'M_{Z_1}X_2}{T^2}\left(\frac{X_2'M_{Z_1}X_2}{T^2}\right)^{-1}\frac{X_2'M_{Z_1}u_1}{T}$$

$$-\frac{\beta' X_1'M_{Z_1}X_2}{T^2}\left(\frac{X_2'M_{Z_1}X_2}{T^2}\right)^{-1}\frac{X_2'M_{Z_1}X_1}{T^2}\left(\frac{X_1'M_{Z_1}X_1}{T^2}\right)^{-1}\frac{X_1'M_{Z_1}u_1}{T} + O_p(T^{-1})$$

$$\Rightarrow \beta'\int_0^1 B_2(r)B_3(r)'dr\left(\int_0^1 B_3(r)B_3(r)'dr\right)^{-1}\left(\int_0^1 B_3(r)dB_0(r)'J_4' \;+\; \omega_{31}\right)$$

$$-\beta'\int_0^1 B_2(r)B_3(r)'dr\left(\int_0^1 B_3(r)B_3(r)'dr\right)^{-1}\int_0^1 B_3(r)B_2(r)'dr$$

$$\left(\int_0^1 B_2(r)B_2(r)'dr\right)^{-1}\left(\int_0^1 B_2(r)dB_0(r)'J_4' \;+\; \omega_{21}\right) \tag{32}$$

To decompose the limiting expression, (32), into S_1 and S_2, let $\tilde{w}_t = (u_{1,t}, \Delta x_t')'$, where $\Delta x_t = (\Delta x_{1,t}', \Delta x_{2,t}')'$, and note that by the same argument used in Lemma 1 of Toda and Phillips (1993), we can obtain the asymptotic result:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \tilde{w}_t \Rightarrow \tilde{B}(r) \equiv BM(\Sigma) \tag{33}$$

where $\tilde{B}(r)$ can be partitioned conformably with $\tilde{w}_t = (u_{1,t}, \Delta x_t')'$ as $(J_4 B_0(r), B_{2,3}(r)')'$, with $B_{2,3}(r) = (B_2(r)', B_3(r)')'$, and $B_0(r), B_2(r)$, and $B_3(r)$ defined as in Appendix A.I. In addition, $\Sigma = \sum_{j=0}^{\infty} E(\tilde{w}_t \tilde{w}_{t+j}')$, and Σ can be partitioned conformably with $\tilde{w}_t = (u_{1,t}, \Delta x_t')'$ as:

$$\underline{\mathbf{\Sigma}} = \begin{pmatrix} \sigma_{11} & \sigma_{21}' \\ \sigma_{21} & \Sigma_{22}' \end{pmatrix}, \tag{34}$$

where it is apparent that $\sigma_{11} = \omega_{11}$.

Observe from equation (25) in Appendix A.I that we can write MA representations for $\Delta x_{1,t}$ and $\Delta x_{2,t}$ as:

$$\Delta x_{1,t} = u_{2,t} + J_2 \Theta_1^*(L) G_1 F_1 u_{t-1} \tag{35}$$

$$\Delta x_{2,t} = u_{3,t} + J_3 \Theta_1^*(L) G_1 F_1 u_{t-1} \tag{36}$$

where $\Theta_1^*(L) = \Theta_1(L) - I_{pm_1+1} = \sum_{j=1}^{\infty} H_1^j L^j$, and all other notation is as defined in Appendix A.I. It follows from the independence of u_t that:

$$\sum_{j=0}^{\infty} E(\Delta x_{1,t} u_{1,t+j}) = \omega_{21} \tag{37}$$

$$\sum_{j=0}^{\infty} E(\Delta x_{2,t} u_{1,t+j}) = \omega_{31}, \tag{38}$$

so that $\sigma_{21} = (\omega_{21}', \omega_{31}')'$. Now, following the arguments of Phillips (1989), Lemma 3.1, note that $J_4 B_0(r)$ can be decomposed as:

$$J_4 B_0(r) = \sigma_{21}' \Sigma_{22}^{-1} B_{2,3}(r) + \sigma_{11.2}^{1/2} W(r) \tag{39}$$

where $W(r)$ is a standard Brownian motion independent of $B_{2,3}(r)$ and $\sigma_{11.2} = \sigma_{11} - \sigma_{21}' \Sigma_{22}^{-1} \sigma_{21}$. Set $B_{0.2,3}(r) = \sigma_{11.2}^{1/2} W(r)$ and it follows that we can write the limiting expression in (32) as the sum of S_1 and S_2. Finally, applying the continuous mapping theorem, we deduce, from (31), (32), and from the argument given above, that:

$$T\hat{\alpha}_T \Rightarrow \frac{S_1 + S_2}{S_3} \quad \text{as} \quad T \to \infty, \tag{40}$$

(ii) To prove this part, note that when $\omega_{21} = \mathbf{0}_{k_1}$ and $\omega_{31} = \mathbf{0}_{k_2}$, we have that $\sigma_{21} = (\omega_{21}', \omega_{31}')' = \mathbf{0}_{k_1+k_2}$. Hence, it follows immediately from part (i) that $S_2 = 0$, and so we have that:

$$T\hat{\alpha}_T \Rightarrow \frac{S_1}{S_3} \quad \text{as} \quad T \to \infty. \tag{41}$$

Note that here $B_{0.2,3}(r) = \sigma_{11} W(r) = \omega_{11} W(r)$, since $\sigma_{21} = (\omega_{21}', \omega_{31}')' = \mathbf{0}_{k_1+k_2}$.

Proof of Theorem 3.1.2: Write

$$t_{\hat{\alpha}_T} = \frac{\hat{\alpha}_T}{s(y' P_{(X_2,Z_1)} M_{(X_1,Z_1)} P_{(X_2,Z_1)} y)^{1/2}}, \tag{42}$$

so that $t_{\hat{\alpha}_T}$ is the usual t-statistic for testing the significance of α in the regression equation (7). Here,

$$s^2 = \frac{y' M_{(X_1, Z_1, \hat{y}_{H_1})} y}{T - k_1 - p m_1 - 1} \tag{43}$$

is the usual estimator of the variance of the error term in equation (7). Moreover, using Lemma A.I.2 and well-known arguments, note that $s^2 \xrightarrow{p} \omega_{11}$ as $T \to \infty$. Now, rewrite the t-statistic, (42), as:

$$t_{\hat{\alpha_T}} = \frac{T\hat{\alpha}_T}{s(y' P_{(X_2,Z_1)} M_{(X_1,Z_1)} P_{(X_2,Z_1)} y / T^2)^{1/2}}, \tag{44}$$

where the limiting expressions for $T\hat{\alpha}_T$ and for $y' P_{(X_2,Z_1)} M_{(X_1,Z_1)} P_{(X_2,Z_1)} y / T^2$ are given by equations (41) and (31), respectively. By the continuous mapping theorem, it thus follows that:

$$t_{\hat{\alpha_T}} \Rightarrow \frac{S_1}{\omega_{11}^{1/2} S_3^{1/2}} \quad \text{as} \quad T \to \infty. \tag{45}$$

Let $F_{2,3}$ be the σ-field generated by $\{B_{2,3}(r) : \; 0 \le r \le 1\}$. Then, following the arguments of Phillips (1989), Lemma 3.1, we have that

$$J_4 B_0(r)|_{F_{2,3}} \equiv \sigma_{21}' \Sigma_{22}^{-1} B_{2,3}(r) + \sigma_{11.2}^{1/2} W(r) \equiv \omega_{11}^{1/2} W(r) \equiv N(0, \omega_{11} r), \tag{46}$$

since $\sigma_{21} = \omega_{21}', \omega_{31}')' = \mathbf{0}_{k_1+k_2}$ and $\sigma_{11.2} = \sigma_{11} - \sigma_{21}' \Sigma_{22}^{-1} \sigma_{21} = \sigma_{11} = \omega_{11}$. Note that here, $\cdot|_{F_{2,3}}$ denotes the conditional distribution relative to $F_{2,3}$. It thus follows based on arguments similar to those used by Phillips (1989), Theorem 3.2, that:

$$\frac{S_1}{\omega_{11}^{1/2} S_3^{1/2}}|_{F_{2,3}} \equiv N(0,1) \tag{47}$$

Note that the distribution given by (47) does not depend on $B_2(r)$ and $B_3(r)$. Hence, it must also be the unconditional asymptotic distribution of $t_{\hat{\alpha_T}}$. Thus,

$$t_{\hat{\alpha_T}} \Rightarrow N(0,1) \text{ as } T \to \infty, \tag{48}$$

Proof of Theorem 3.1.4: As the proof of Theorem 3.1.4 is similar to that for Theorems 3.1.1 and 3.1.2 above, the arguments given here are abbreviated. To show part (i), first write T times the least squares estimator of α^* as:

$$T\hat{\alpha}_T^* = \frac{y' P_{(X_2, Z_1^*)} M_{(X_{1,-1}, Z_1^*)} y/T}{y' P_{(X_{2,-1}, Z_1^*)} M_{(X_{1,-1}, Z_1^*)} P_{(X_{2,-1}, Z_1^*)} y/T^2}, \tag{49}$$

where $y, X_{1,-1}, X_{2,-1}, and Z_1^*$ are as defined in Section 3.1, and where $M_{(X_{1-1}, Z_1^*)}$ is a TxT matrix which projects onto the orthogonal complement of the span of the columns of $X_{1,-1}$ and Z_1^*. We have multiplied by T in anticipation of the correct rate of the appropriate rate of convergence of $\hat{\alpha}_T^*$. First, consider the denominator of (30): It turns out that:

$$y' P_{(X_{2,-1}, Z_1^*)} M_{(X_{1,-1}, Z_1^*)} P_{(X_{2,-1}, Z_1^*)} y/T^2 \Rightarrow S_3. \tag{50}$$

Further details of this proof are in Chao and Swanson (1997). With respect to the numerator of (49), note that Lemma A.I.2 and the continuous mapping theorem yields that:

$$y' P_{(X_{2,-1}, Z_1^*)} M_{(X_{1,-1}, Z_1^*)} y/T \Rightarrow S_1^*, \tag{51}$$

where S_1^* is defined in the statement of Theorem 3.1.4. Finally, it follows from (50) and (51) and the continuous mapping theorem that:

$$T\hat{\alpha}_T^* \Rightarrow \frac{S_1^*}{S_3} \text{ as } T \to \infty, \tag{52}$$

To prove (ii), write:

$$t_{\hat{\alpha}_T^*} = \frac{T\hat{\alpha}_T^*}{s^*(y' P_{(X_{2,-1}, Z_1^*)} M_{(X_{1,-1}, Z_1^*)} P_{(X_{2,-1}, Z_1^*)} y/T^2)^{1/2}}, \tag{53}$$

where

$$s^{2^*} = \frac{y' M_{(X_{1,-1}, Z_1^*, \hat{y}_{H_1}^*)} y}{T - 2k_1 - k_2 - pm_1 - 1} \tag{54}$$

is the usual estimator of the variance of the error term in equation (13). Observe that $s^{2^*} \xrightarrow{p} \omega_{11.2}$ $= \underline{\omega}'_{21}\underline{\ }_{22}^{-1}\underline{\omega}_{21}$ as $T \to \infty$ under H_0, which can be shown using Lemma A.I.2 and the continuous mapping theorem. Then, (50), (52), and the continuous mapping theorem imply that:

$$t_{\hat{\alpha}_T^*} \Rightarrow \frac{S_1^*}{\omega_{11.2}^{1/2} S_3^{1/2}} \text{ as } T \to \infty. \tag{55}$$

Now, from (35) and (36),

$$\Delta x_{1,t} = u_{2,t} + J_2\Theta_1^*(L)G_1F_1u_{t-1} = v_{2,t} + J_2\Theta_1^*(L)G_1F_1u_{t-1}$$

$$\Delta x_{2,t} = u_{3,t} + J_3\Theta_1^*(L)G_1F_1u_{t-1} = v_{3,t} + J_3\Theta_1^*(L)G_1F_1u_{t-1},$$

where $\Theta_1^*(L) = \sum_{j=1}^{\infty} H_1^j L^j$, $v_{2,t} = u_{2,t}$, and $v_{3,t} = u_{3,t}$. It follows from the serial independence of $\{u_t\}$ and the definition of $v_{1.2,t}$ that: $E(\Delta x_{1,t}v_{1.2,t}) = \mathbf{0}_{k_1}$ and $E(\Delta x_{2,t}v_{1.2,t}) = \mathbf{0}_{k_2}$, Moreover, since $B_2(r)$ and $B_3(r)$ are, respectively, the limiting processes of the partial sum of $\Delta x_{1,t}$ and $\Delta x_{2,t}$ (see Lemma A.I.1), and $B^*_{0.2,3}$ is the limit process of the partial sum of $V_{1.2,t}$, it follows that $B_{2,3}(r) = (B_2(r)', B_3(r)')'$ and $B^*_{0.2,3}(r)$ are independent. Hence, let $F_{2,3}$ again be the σ-field generated by $\{B_{2,3}(r) : 0 \leq r \leq 1\}$, and it is apparent that:

$$B^*_{0.2,3}(r)|_{F_{2,3}} \equiv B^*_{0.2,3}(r) \equiv \omega_{11.2}^{1/2} = N(0, \omega_{11.2}^{1/2} r), \tag{56}$$

so that

$$\frac{S_1^*}{\omega_{11.2}^{1/2} S_3^{1/2}}|_{F_{2,3}} \equiv N(0,1) \tag{57}$$

Note that since the conditional distribution given by (57) does not depend on $B_2(r)$ and $B_3(r)$, it must also be the unconditional distribution. Thus, it follows from (55) that

$$t_{\hat{\alpha}_T^*} \Rightarrow N(0,1) \text{ as } T \to \infty, \tag{58}$$

Proof of Theorem 3.2.1: (i) To show case (a), where we suppose that $Rank(A) = k_1 \leq k_2$, write the least squares estimator of α in (16) as:

$$\hat{\alpha}_T = \frac{y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}y/T}{y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}P_{(X_2,\ddot{Z}_1)}y/T}, \tag{59}$$

where $y, X_1, X_2, \ddot{Z}_1$, are as defined after (16) in Section 3.2, and where the division by T of both the numerator and the denominator reflects our anticipation of the appropriate rate of convergence. Note first that the denominator of (59) can be rewritten as:

$$y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}P_{(X_2,\ddot{Z}_1)}y/T$$

$$= \frac{y'M_{\ddot{Z}_1}X_2}{T}[\left(\frac{X_2'M_{\ddot{Z}_1}X_2}{T}\right)^{-1} - \left(\frac{X_2'M_{\ddot{Z}_1}X_2}{T}\right)^{-1}\frac{X_2'M_{\ddot{Z}_1}X_1}{T}$$

$$\left(\frac{X_1' M_{\ddot{Z}_1} X_1}{T}\right)^{-1} \frac{X_1' M_{\ddot{Z}_1} X_2}{T} \left(\frac{X_2' M_{\ddot{Z}_1} X_2}{T}\right)^{-1}] \frac{X_2' M_{\ddot{Z}_1} y}{T} \tag{60}$$

Making use of (27) and Lemma A.II.2, we can represent the components of (60) as follows:

$$\frac{X_2' M_{\ddot{Z}_1} y}{T} = \frac{X_{2,-1}' M_{\ddot{Z}_1} X_{2,-1} A\beta}{T} + r_1, \tag{61}$$

$$\frac{X_2' M_{\ddot{Z}_1} X_1}{T} = \frac{X_{2,-1}' \ddot{M}_{Z_1} X_{2,-1} A}{T} + R_2, \tag{62}$$

$$\left(\frac{X_2' M_{\ddot{Z}_1} X_2}{T}\right)^{-1} = \left(I_{k_2} - R_3 + O_p(\frac{1}{T^2})\right) \left(\frac{X_{2,-1}' M_{\ddot{Z}_1} X_{2,-1}}{T}\right)^{-1}, \text{and} \tag{63}$$

$$\left(\frac{X_1' M_{\ddot{Z}_1} X_1}{T}\right)^{-1} = \left(I_{k_1} - R_4 + O_p(\frac{1}{T^2})\right) \left(\frac{A' X_{2,-1}' M_{\ddot{Z}_1} X_{2,-1} A}{T}\right)^{-1}, \tag{64}$$

where

$$r_1 = \frac{u_3' M_{\ddot{Z}_1} X_{2,-1} A\beta}{T} + \frac{X_{2,-1}' M_{\ddot{Z}_1} v_1}{T} + \frac{u_3' M_{\ddot{Z}_1} v_1}{T},$$

$$R_2 = \frac{u_3' M_{\ddot{Z}_1} X_{2,-1} A}{T} + \frac{X_{2,-1}' M_{\ddot{Z}_1} v_2}{T} + \frac{u_3' M_{\ddot{Z}_1} v_2}{T},$$

$$R_3 = \left(\frac{X_{2,-1}' M_{\ddot{Z}_1} X_{2,-1}}{T}\right)^{-1} \left(\frac{u_3' M_{\ddot{Z}_1} X_{2,-1}}{T} + \frac{X_{2,-1}' M_{\ddot{Z}_1} u_3}{T} + \frac{u_3' M_{\ddot{Z}_1} u_3}{T}\right),$$

$$R_4 = \left(\frac{A' X_{2,-1}' M_{\ddot{Z}_1} X_{2,-1} A}{T}\right)^{-1} \left(\frac{A' X_{2,-1}' M_{\ddot{Z}_1} v_2}{T} + \frac{v_2' M_{\ddot{Z}_1} X_{2,-1} A}{T} + \frac{v_2' M_{\ddot{Z}_1} v_2}{T}\right),$$

Substituting expressions (61)-(64) into (60), we obtain:

$$y' P_{(X_2,\ddot{Z}_1)} M_{(X_1,\ddot{Z}_1)} P_{(X_2,\ddot{Z}_1)} y/T = \frac{\beta' u_2' u_2 \beta}{T} + O_p(\frac{1}{T}).\text{[6]} \tag{65}$$

It follows from Lemma A.II.2 and the Slutsky theorem that:

$$y' P_{(X_2,\ddot{Z}_1)} M_{(X_1,\ddot{Z}_1)} P_{(X_2,\ddot{Z}_1)} y/T \xrightarrow{p} \beta' \text{-}\ {}_{22}\beta \text{ as } T \to \infty \tag{66}$$

With respect to the numerator of (59), note that by straightforward algebra, we can write

$$y' P_{(X_2,\ddot{Z}_1)} M_{(X_1,\ddot{Z}_1)} y/T = \frac{y' M_{\ddot{Z}_1} X_2}{T} \left(\frac{X_2' M_{\ddot{Z}_1} X_2}{T}\right) [\frac{X_2' M_{\ddot{Z}_1} y}{T}$$

$$-\frac{X_2' M_{\ddot{Z}_1} X_1}{T} \left(\frac{X_1' M_{\ddot{Z}_1} X_1}{T} \right)^{-1} \frac{X_1' M_{\ddot{Z}_1} y}{T}] \tag{67}$$

Note further that with the exception of the factor $(X_1' M_{\ddot{Z}_1} y/T)$, all components of (67) arose in our examination of the denominator of $\hat{\alpha}_T$ (see (61)-(64)). Using (27), we can rewrite $(X_1' M_{\ddot{Z}_1} y/T)$ as

$$\frac{X_1' \ddot{M}_{Z_1} y}{T} = \frac{A' X_{2,-1}' \ddot{M}_{Z_1} X_{2,-1} A \beta}{T} + R_5, \tag{68}$$

where

$$R_5 = \frac{v_2' M_{\ddot{Z}_1} X_{2,-1} A \beta}{T} + \frac{A' X_{2,-1}' M_{\ddot{Z}_1} v_1}{T} + \frac{v_2' M_{\ddot{Z}_1} v_1}{T}$$

Substituting expressions (61)-(64) and (68) into (67), we obtain that:

$$y' P_{(X_2, \ddot{Z}_1)} M_{(X_1, \ddot{Z}_1)} y/T = \frac{-\beta' u_2' u_1}{T} + O_p(\frac{1}{T}), \tag{69}$$

so that applying Lemma A.II.2 and the Slutsky Theorem, it thus follows that

$$y' P_{(X_2, \ddot{Z}_1)} M_{(X_1, \ddot{Z}_1)} y/T \xrightarrow{p} -\beta' \omega_{21} \text{ as } T \to \infty \tag{70}$$

The result given by equation (17) in Theorem 3.2.1(i)(a) thus follows immediately from (66), (70), and the Slutsky Theorem. To show that the associated t-statistic diverges in the case where $\beta' E(u_2' u_1) = \beta' \omega_{21} \neq 0$, note that

$$t_{\hat{\alpha}_T} = \frac{\hat{\alpha}_T}{s} \left(y' P_{(X_2, \ddot{Z}_1)} M_{(X_1, \ddot{Z}_1)} P_{(X_2, \ddot{Z}_1)} y \right)^{-1/2}, \tag{71}$$

where $s^2 = \frac{y' M_{(X_1, \ddot{Z}_1, \hat{y}_{H_1})} y}{T - k_1 - p m_1 - 1}$ is the usual estimator of the variance of the error term in regression (16). Further,

$$s^2 \xrightarrow{p} \omega_{11} - \frac{(\beta' \omega_{21})^2}{\beta' \text{-}\ {}_{22} \beta} \text{as } T \to \infty \tag{72}$$

(While the limiting result given by (72) seems innocuous enough, its derivation is actually quite tedious due to a singularity induced by the cointegration between $x_{1,t}$ and $x_{2,t}$. For the sake of brevity, we have chosen to omit this derivation. However, complete details can be obtained from the authors upon request.) It follows from (66) and (72) that:

$$s \left(y' P_{(X_2, \ddot{Z}_1)} M_{(X_1, \ddot{Z}_1)} P_{(X_2, \ddot{Z}_1)} y \right)^{-1/2} = O_p(\frac{1}{\sqrt{T}}) \tag{73}$$

and hence the t-statistic given by (71) diverges for $\beta' \neq 0$, as $\hat{\alpha}_T \xrightarrow{p} -\frac{\beta' \omega_{21}}{\beta' \ \ {}_{22} \beta}$as $T \to \infty$.

To show case (b) of (i), note that in this case, A is not of full column rank, since $Rank(A) = k_2 \leq k_1$, so we define the matrix

$$A^* = (A'(AA')^{-1/2}|A'_\perp(A_\perp A'_\perp)^{-1/2}) = (A_1^*|A_2^*), \tag{74}$$

where $A_\perp$ is a $(k_1 - k_2)$xk_1 matrix such that $A_\perp A'_\perp = \mathbf{0}_{(k_1-k_2)\text{x}k_1}$. Note also that $A^{*'}A^* = I_{k_1} = A^*A^{*'}$, so that A^* is an orthogonal matrix. Again, we begin by examining the denominator of (59), which can be written as:

$$y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}P_{(X_2,\ddot{Z}_1)}y/T$$

$$= \frac{y'M_{\ddot{Z}_1}X_2}{T}[\left(\frac{X_2'M_{\ddot{Z}_1}X_2}{T}\right)^{-1} - \left(\frac{X_2'M_{\ddot{Z}_1}X_2}{T}\right)^{-1}\{X_2'M_{\ddot{Z}_1}X_1A^*D$$

$$\left(DA^{*'}X_1'M_{\ddot{Z}_1}X_1A^*D\right)^{-1}\left(DA^{*'}X_1'M_{\ddot{Z}_1}X_2\right)/T\}\left(\frac{X_2'M_{\ddot{Z}_1}X_2}{T}\right)^{-1}]\frac{X_2'M_{\ddot{Z}_1}y}{T}, \tag{75}$$

where D is a k_1xk_1 diagonal matrix of the form:

$$\mathbf{D} = \begin{pmatrix} \frac{1}{T} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{T}}I_{(k_1-k_2)} \end{pmatrix}.$$

Making use of (27) and Lemma A.II.2, note that after some algebra we obtain that:

$$\left(X_2'M_{\ddot{Z}_1}X_1A^*D\right)\left(DA^{*'}X_1'M_{\ddot{Z}_1}X_1A^*D\right)^{-1}\left(DA^{*'}X_1'M_{\ddot{Z}_1}X_2\right)/T$$

$$= \frac{X_{2,-1}'M_{\ddot{Z}_1}X_{2,-1}}{T} + \frac{u_3'M_{\ddot{Z}_1}u_3}{T} - \frac{(AA')^{-1}Au_2'M_{\ddot{Z}_1}u_2A'(AA')^{-1}}{T}$$

$$+\frac{(AA')^{-1}A'u_2'M_{\ddot{Z}_1}u_2A'_\perp}{T}\left(\frac{A'_\perp u_2'M_{\ddot{Z}_1}u_2A'_\perp}{T}\right)^{-1}\frac{A'_\perp u_2'M_{\ddot{Z}_1}u_2A'(AA')^{-1}}{T}$$

$$+\frac{u_3'M_{\ddot{Z}_1}X_{2,-1}}{T} + \frac{X_{2,-1}'M_{\ddot{Z}_1}u_3}{T} + O_p(\frac{1}{T}) = R_6 + O_p(\frac{1}{T}), \text{ say.} \tag{76}$$

Using (76), we can rewrite (75) as:

$$y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}P_{(X_2,\ddot{Z}_1)}y/T$$

$$= \left(\frac{\beta'A'X_{2,-1}'M_{\ddot{Z}_1}X_{2,-1}}{T} + r_1'\right)[(I_{k_2} - R_3 + O_p(\frac{1}{T^2}))\left(\frac{X_2'M_{\ddot{Z}_1}X_{2,-1}}{T}\right)^{-1}$$

$$-(I_{k_2} - R_3 + O_p(\frac{1}{T^2}))\left(\frac{X_2' M_{\ddot{Z}_1} X_{2,-1}}{T}\right)^{-1}(R_6 + O_p(\frac{1}{T}))$$

$$(I_{k_2} - R_3 + O_p(\frac{1}{T^2}))\left(\frac{X_2' M_{\ddot{Z}_1} X_{2,-1}}{T}\right)^{-1}]\left(\frac{X_{2,-1}' M_{\ddot{Z}_1} X_{2,-1} A\beta}{T} + r_1\right) \tag{77}$$

where the equality above follows from the substitution of expressions (61) and (63). More tedious but straightforward calculations then yields that:

$$y' P_{(X_2,\ddot{Z}_1)} M_{(X_1,\ddot{Z}_1)} P_{(X_2,\ddot{Z}_1)} y/T = \beta' A'(AA')^{-1} A(u_2' u_2/T) A'(AA')^{-1} A\beta$$

$$-\beta' A'(AA')^{-1} A(u_2' u_2/T) A_\perp' [A_\perp (u_2' u_2/T) A_\perp']^{-1} A_\perp (u_2' u_2/T)$$

$$A'(AA')^{-1} A\beta + O_p(\frac{1}{T}) \xrightarrow{p} l_2 \text{ as } T \to \infty. \tag{78}$$

To analyze the numerator of $\hat{\alpha}_T$ for this case, we write:

$$y' P_{(X_2,\ddot{Z}_1)} M_{(X_1,\ddot{Z}_1)} y/T = \frac{y' M_{\ddot{Z}_1} X_2}{T}\left(\frac{X_2' M_{\ddot{Z}_1} X_2}{T}\right)^{-1}[\left(\frac{X_2' M_{\ddot{Z}_1} y}{T}\right)$$

$$-X_2' M_{\ddot{Z}_1} X_1 A^* D\left(DA^{*'} X_1' M_{\ddot{Z}_1} X_1 A^* D\right)^{-1}\left(DA^{*'} X_1' M_{\ddot{Z}_1} y\right)/T]. \tag{79}$$

Again making use of (27) and Lemma A.II.2, note that:

$$\left(X_2' M_{\ddot{Z}_1} X_1 A^* D\right)\left(DA^{*'} X_1' M_{\ddot{Z}_1} X_1 A^* D\right)^{-1}\left(DA^{*'} X_1' M_{\ddot{Z}_1} y\right)/T$$

$$= \frac{X_{2,-1}' M_{\ddot{Z}_1} X_{2,-1} A\beta}{T} - \frac{(AA')^{-1/2} A u_2' M_{\ddot{Z}_1} u_2 A'(AA')^{-1} A\beta}{T}$$

$$+\frac{(AA')^{-1} A' u_2' M_{\ddot{Z}_1} u_2 A_\perp'}{T}\left(\frac{A_\perp' u_2' M_{\ddot{Z}_1} u_2 A_\perp'}{T}\right)^{-1}\frac{A_\perp' u_2' M_{\ddot{Z}_1} u_2 A'(AA')^{-1} A\beta}{T}$$

$$+\frac{u_3' M_{\ddot{Z}_1} X_{2,-1} A\beta}{T} + \frac{X_{2,-1}' M_{\ddot{Z}_1}(u_1 + u_2\beta + u_3 A\beta)}{T}$$

$$+\frac{u_3' M_{\ddot{Z}_1}(u_1 + u_2\beta + u_3 A\beta)}{T} + \frac{(AA')^{-1} A u_2' M_{\ddot{Z}_1}(u_1 + u_2\beta)}{T}$$

$$-\frac{(AA')^{-1} A u_2' M_{\ddot{Z}_1} u_2 A_\perp'}{T}\left(\frac{A_\perp u_2' M_{\ddot{Z}_1} u_2 A_\perp'}{T}\right)^{-1}\frac{A_\perp u_2' M_{\ddot{Z}_1}(u_1 + u_2\beta)}{T} + O_p(\frac{1}{T})$$

$$= R_7 + O_p(\frac{1}{T}), \text{ say.} \tag{80}$$

Using (61) and (63), we can rewrite (80) as:

$$y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}y/T$$

$$= \left(\frac{\beta' A' X'_{2,-1} M_{\ddot{Z}_1} X_{2,-1}}{T} + r'_1\right)[(I_{k_2} - R_3 + O_p(\frac{1}{T^2}))\left(\frac{X'_2 M_{\ddot{Z}_1} X_{2,-1}}{T}\right)^{-1}]$$

$$[\left(\frac{X'_{2,-1} M_{\ddot{Z}_1} X_{2,-1} A\beta}{T} + r_1\right) - (R_{14} + O_p(\frac{1}{T}))] \tag{81}$$

Further calculations show that (81) can be rewritten as:

$$y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}y/T$$

$$= -\beta' A'(AA')^{-1} A[I_{k_1} - (u'_2 u_2/T)A'_{\perp}(A_{\perp}(u'_2 u_2/T)A'_{\perp})^{-1}A_{\perp}](u'_2 u_2/T) + O_p(\frac{1}{T})$$

$$\xrightarrow{p} \ l_1 \text{ as } T \to \infty \tag{82}$$

Finally, it follows from (78), (82), and the continuous mapping theorem that as $T \to \infty$, $\hat{\alpha}_T \xrightarrow{p} \frac{l_1}{l_2}$, as required. When $l_1 \neq 0$, $\hat{\alpha}_T$ does not converge in probability to 0, and the divergence of the associated t-statistic follows from an argument entirely analogous to that given in case (a) above. For brevity, we thus omit the detAIls here.

(ii) To show case (a), where $E(u'_2 u_1) = \omega_{21} = \mathbf{0}_{k_1}$, we divide the numerator of $\hat{\alpha}_T$ given by (59) by $\sqrt{T}$ instead of by T. Then, following calculations similar to those used in (i)(a), we have (analogous to expression (69)) that:

$$y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}y/\sqrt{T} = \frac{-\beta' u'_2 u_1}{\sqrt{T}} + O_p(\frac{1}{\sqrt{T}}) \tag{83}$$

Note that $var(\beta' u_{2,t} u_{1,t}) = E(\beta' u_{2,t} u_{1,t})^2 = \omega_{11}(\beta' \text{-}\ {}_{22}\beta)$ by the uncorrelatedness of $u_{1,t}$ and $u_{2,t}$, where $u'_{2,t}$ and $u_{1,t}$ are the t^{th} row of the Txk_1 matrix u_2 and the t^{th} element of the Tx1 vector u_1, respectively. Hence, by the Lindeberg-Levy central limit theorem, we have that:

$$\frac{-\beta' u'_2 u_1}{\sqrt{T}} \Rightarrow N(0, (\beta' \text{-}\ {}_{22}\beta)) \text{ as } T \to \infty \tag{84}$$

On the other hand, the denominator of $\hat{\alpha}_T$ divided by T converges to $\beta' \text{-}_{22}\beta$ as in the proof of (i)(a) above (see expression (66)). Thus, it follows by the Cramer Convergence Theorem that

$$\sqrt{T}\hat{\alpha}_T = \frac{y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}y/\sqrt{T}}{y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}P_{(X_2,\ddot{Z}_1)}y/T} \Rightarrow N(0,(\beta' \text{-}_{22}\beta)^{-1}) \text{ as } T \to \infty \tag{85}$$

Moreover, define

$$s^2 = \frac{y'M_{(X_1,\ddot{Z}_1,\hat{y}^{H_1})}y}{T - k_1 - pm_1 - 1} \tag{86}$$

Then, it is easy to show using the usual argument that $s^2 \xrightarrow{p} \omega_{11}$. Thus, it follows from (87), the consistency of s^2, and the Cramer Convergence Theorem that:

$$t_{\hat{\alpha}_T} = \frac{\sqrt{T}\hat{\alpha}_T}{s\left(y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}P_{(X_2,\ddot{Z}_1)}y/T\right)^{1/2}} \Rightarrow N(0,1) \text{ as } T \to \infty \tag{87}$$

To show case (b), divide the numerator of $\hat{\alpha}_T$ in (59) by $\sqrt{T}$ instead of T, and calculations similar to those in part(i) case (b) yield that:

$$y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}y/\sqrt{T}$$

$$= -\beta'A'(AA')^{-1}A[I_{k_1} - (u_2'u_2/T)A_\perp'(A_\perp(u_2'u_/\sqrt{T})A_\perp')^{-1}A_\perp](u_2'u_2/T) + O_p(\frac{1}{\sqrt{T}}) \tag{88}$$

Applying the Weak Law of Large Numbers, the Slutsky Theorem, the Lindeberg-Levy central limit theorem, and the Cramer Convergence Theorem to (90), we find that:

$$\frac{y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}y}{\sqrt{T}} \Rightarrow N(0,\omega_{11}l_2) \text{ as } T \to \infty. \tag{89}$$

It then follows from (78), (89), and the Cramer Convergence Theorem that:

$$\sqrt{T}\hat{\alpha}_T = \frac{y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}y/\sqrt{T}}{y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}P_{(X_2,\ddot{Z}_1)}y/T} \Rightarrow N(0,\omega_{11}l_2^{-1}) \text{ as } T \to \infty. \tag{90}$$

Again, observe that $s^2 \xrightarrow{p} \omega_{11}$, as $T \to \infty$. Thus, it follows from (78), (89), and the Cramer Convergence Theorem that:

$$t_{\hat{\alpha}_T} = \frac{\sqrt{T}\hat{\alpha}_T}{s\left(y'P_{(X_2,\ddot{Z}_1)}M_{(X_1,\ddot{Z}_1)}P_{(X_2,\ddot{Z}_1)}y/T\right)^{1/2}} \Rightarrow N(0,1) \text{ as } T \to \infty \tag{91}$$

Footnotes

1. Assumptions A1 and A2 are stability conditions, as discussed in Appendix A.I.

2. Note that the normality assumption is not needed for our asymptotic analysis, but is given here to facilitate the derivation of the conditional model presented in equation (11).

3. The derivation of this conditional model is omitted. Interested readers are referred to Phillips (1991), Urbain (1992,1993), Johansen (1992,1995) and Zivot (1994) for details.

4. Please refer to Urbain (1992,1993) and Johansen (1992,1995) for a more complete discussion of weak exogeneity in a system with I(1) variables and cointegration.

5. Assumptions A1 and A2 are stability conditions, as discussed in Appendix A.II.

6. Note that expression (65) does not involve the higher order term $(\beta' A' X'_{2,-1} X_{2,-1} A\beta)/T$ since cointegration between $x_{1,t}$ and $x_{2,t}$ results in the annihilation of this term when the fitted value under the alternative is projected onto the orthogonal complement of $(X_1, \ddot{Z}_1)$. This explains why we divided the denominator of $\hat{\alpha}_T$ by T instead of T^2 as is usually done in regressions with I(1) regressors.

7. If the model with $x_{1,t}$ is preferred, then only short-run dynamics (i.e. lagged differences) of U.K. industrial production appear in the regression model for US_t. Long-run dynamics (i.e. cointegrating relations) do not enter into the "best" model for US_t. It is perhaps worth noting that this concept is related to separation in cointegration (Granger and Haldrup (1997)). However, in separation in cointegration, it is assumed that the short-run dynamics do not spill over between "models", while the long-run error-correcting terms do affect both "models", which is essentially the opposite of our setup.

8. It is worth noting that Terasvirta and Anderson (1992) find evidence of nonlinearity (i.e. smooth transition autoregression) for all of these countries except France, when using the growth rates of quarterly industrial production indices. This may be taken as evidence that our linear models are misspecified, thus invalidating our tests. However, Van Dijk, Franses, and Lucas (1996) find that much of the evidence of nonlinearity in Terasvirta and Anderson (1992) appears to be due to a small number of outliers. Although this is not surprising (Terasvirta and Anderson (1992) actually point out that the nonlinearity in their models is needed mainly to describe the responses of production to a few large negative shocks), we take this as evidence that our linear models are not inherantly misspecified.

8 References

Bernake, B., Bohn, H., and Reiss, P.C. (1988). Alternative Non-nested Specification Tests of Time-Series Investment Models. *Journal of Econometrics* **37**, 293-326.

Clements, M.P. and Hendry, D.F. (1994). Towards a Theory of Economic Forecasting. in: *Non-stationary Time-Series Analyses and Cointegration*, eds. Hargraves, C., Oxford University Press, Oxford, pp. 1-50.

Chao, J.C. and Swanson, N.R. (1997). Tests of Non-nested Hypotheses in Nonstationary Regressions with an Application to Modeling Industrial Production. Discussion Paper. Pennsylvania State University.

Chan, N.H. and Wei, C.Z. (1988). Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes. *Annals of Statistics* **16**, 367-401.

Chong, Y.Y. and Hendry, D.F. (1986). Econometric Evaluation of Linear Macro-economic Models. *Review of Economic Studies* **53**, 671-690.

Corradi, V. and Swanson, N.R. (1997). Choosing Between Levels and Logs in the Presence of Deterministic and Stochastic Trends. Discussion Paper. Pennsylvania State University.

Cox, D.R. (1961). Tests of Separate Families of Hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 105-123.

Cox, D.R. (1962). Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society, Series B* **24**, 406-424.

Dadkhah, K.M. and Valbuena, S. (1985). Non-nested Tests of New Classical vs Keynesian Models: Evidence from European Economies. *Applied Economics* **17**, 1083-1098.

Davidson, R. and MacKinnon, J.G. (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica* **48**, 781-793.

Davidson, R. and MacKinnon, J.G. (1983). Testing the Specification of Multivariate Models in the Presence of Alternative Hypotheses. *Journal of Econometrics* **23**, 301-313.

Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics.* Oxford University Press, New York.

ENGLE, R.F. AND GRANGER, C.W.J. (1987). Cointegration and Error: Representation, Estimation, and Testing. *Econometrica* **55**, 251-276.

ENGLE, R.F., HENDRY, D.F., AND RICHARD, J.-F. (1983). Exogeneity. *Econometrica* **51**, 277-304.

ERICSSON, N.R. (1983). Asymptotic Properties of Instrumental Variables Statistics for Testing Non-nested Hypotheses. *Review of Economic Studies* **50**, 287-304.

ERICSSON, N.R. (1992). Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration. *Journal of Policy Modeling* **14**, 465-495.

FILARDO, A. (1993). International Co-movements of Business Cycles. Discussion Paper. Federal Reserve Bank of Kansas City.

FISHER, G.R. AND MCALEER, M. (1981). Alternative Procedures and Associated Tests of Significance for Non-nested Hypotheses. *Journal of Econometrics* **16**, 103-119.

GODFREY, L.G. (1983). Testing Non-nested Models After Estimation by Instrumental Variables or Least Squares. *Econometrica* **51**, 355-365.

GOURIEROUX, C. AND MONFORT, A. (1994). Testing Non-nested Hypotheses. in: R.F. Engle and D. McFadden eds., *Handbook of Econometrics* **Vol. 4**, 2583-2637. Elsivier Science B.V., Amsterdam.

GRANGER, C.W.J. AND HALDRUP, N. (1992). Separation in Cointegrated Systems and Persistent-Transitory Decompositions. *Oxford Bulletin of Economics and Statistics* **59**, 449-463.

JOHANSEN, S. (1992). Cointegration in Partial Systems and the Efficiency of single Equation Analysis. *Journal of Econometrics* **52**, 389-402.

JOHANSEN, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, New York.

LU, M. AND MIZON, G.E. (1991). Forecast Encompassing and Model Evaluation. in: *Economic Structural Change: Analysis and Forecasting*, eds. Hackl, P. and Westlund, A.H. Springer-Verlag, Berlin, pp. 123-138.

MacKinnon, J.G. (1983). Model Specification Tests Against Non-nested Alternatives. *Econometric Reviews* **2**, 85-157.

MacKinnon, J.G., White, H. and Davidson, R. (1983). Tests for Model Specification in the Presence of Alternative Hypotheses: Some Further Results. *Journal of Econometrics* **21**, 53-70.

McAleer, M.J. (1987). Specification Tests for Separate Models: A Survey. in: *Specification Analysis in the Linear Model*, eds. King, M.L. and Giles, D.E.A. Routledge and Kegan Paul, London.

Mizon, G. and Richard, J.-F. (1986). The Encompassing Principle and Its Application to Non-nested Hypothesis Tests. *Econometrica* **54**, 657-678.

Nelson, C.R. and Plosser, C.I. (1982). Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications. *Journal of Monetary Economics* **10**, 139-162.

Pesaran, M.H. (1974). On the General Problem of Model Selection. *Review of Economic Studies* **41**, 153-171.

Pesaran, M.H. and Deaton, A.S. (1978). Testing Non-nested Nonlinear Regression Models. *Econometrica* **46**, 677-694.

Phillips, P.C.B. (1989). Partially Identified Econometric Models. *Econometric Theory* **5**, 181-240.

Phillips, P.C.B. (1991). Optimal Inference in Cointegrated Systems. *Econometrica* **59**, 283-306.

Phillips, P.C.B. and Loretan, M. (1991). Estimating Long-run Economic Equilibria. *Review of Economic Studies* **59**, 407-436.

Phillips, P.C.B. and Park, J.Y. (1989). Statistical Inference in Regressions with Integrated Processes: Part 2. *Econometric Theory* **5**, 95-131.

Terasvirta, T. and Anderson, H. (1992). Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. *Journal of Applied Econometrics* **7**, 119-136.

Toda, H.Y. and Phillips, P.C.B. (1993). Vector Autoregressions and Causality. *Econometrica* **61**, 1367-1393.

Urbain, J.-P. (1992). On Weak Exogeneity in Error-Correction Models. *Oxford Bulletin of Economics and Statistics* **54**, 187-207.

URBAIN, J.-P. (1993). *Exogeneity in Error-Correction Models.* Lecture Notes in Economics and Mathematical Systems No. 398. Springer-Verlag, Berlin, Germany.

VAN DIJK, FRANSES, P.H., AND LUCAS, A. (1996). Testing for Smooth Transition Autoregression in the Presence of Outliers , Econometric Institute Report 9622, Erasmus University.

VUONG, Q. (1989). Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses. *Econometrica* **57**, 307-334.

ZIVOT, E. (1994). Single Equation Conditional Error Correction Model Based Tests for Cointegration. Discussion Series 94-12, University of Washington.