A joint test for a unit root and common factor restrictions in the presence of a structural break

Junsoo Lee a, Christine Amsler b,*

* Department of Economics, Vanderbilt University, Nashville, TN 37235, USA
b Department of Economics, Michigan State University, East Lansing, MI 48824, USA

Abstract

This paper provides evidence that the common factor restrictions (CFR) play an important role in testing for a unit root in the presence of a structural break. We first show that the CFR should not be ignored in the unit root testing procedure when a structural break is allowed for. Then we provide a Wald test statistic for the joint hypothesis of a unit root and the CFR. The empirical finding suggests that the missing common factor restrictions may result in a quite different and perhaps misleading inference. © 1997 Elsevier Science B.V.

Keywords: Structural break; Unit root; Common factor restrictions

JEL classification: C12; C15; C22

1. Introduction

The innovative work of Perron (1989) suggests that unit root tests are biased toward accepting the false null hypothesis of a unit root if a time series exhibits stationary fluctuations around a trend containing structural breaks. He proposed an alternative testing procedure for the unit root hypothesis in the presence of a structural break. His empirical findings from applying unit root tests allowing for a break are in favor of the alternative hypothesis of stationarity. Amsler and Lee (1995) (hereafter AL) further examine this issue, and using the suitably modified version of the unit root test proposed by Schmidt and Phillips (1992) (hereafter SP), obtain theoretical and empirical evidence different from Perron (1989). One noticeable difference between Perron's test and the AL test is that the common factor restrictions (CFR) are imposed in the AL test, but not in Perron's test. In
this paper, therefore, we examine the role of the CFR in testing for a unit root in the presence of a structural break. We use the Dickey and Fuller (1979) (hereafter DF) parameterization as do Perron and others, but we add the CFR.

We first show that the CFR should be included in the unit root testing procedure when a structural break is allowed for. We then provide a Wald statistic which tests the joint hypothesis of a unit root and the CFR in the presence of a structural break. To allow for autocorrelated errors, we present the Phillips and Perron (1988) type correction for the Wald test. The empirical evidence supports the finding of AL. The results in this paper may indicate that Perron’s results, favoring the stationarity alternative, do not come from the presence of a break, but from the way that the DF parameterization allows for a break.

The plan of the paper is as follows. In Section 2 we discuss the CFR. In Section 3 we derive a Wald test, and relevant asymptotic results. In Section 4 we provide simulation results, and apply the Wald test to the Nelson–Plosser data. Throughout this paper, ‘→’ indicates weak convergence as $T \rightarrow \infty$.

2. Common factor restrictions

We consider the following data generating process (DGP):

$$y_t = \delta' z_t + x_t, \quad x_t = \rho x_{t-1} + e_t, \quad (1)$$

where $z_t$ contains exogenous variables. The unit root hypothesis is $\rho = 1$, and this is the null hypothesis to be tested. This DGP is used by SP and AL. We consider $z_t = [1, t, D_t]'$ and $\delta = [\delta_1, \delta_2, \delta_3]'$ for the crash model, where $D_t = 1$ for $t \geq T_B + 1$ and zero otherwise, while $T_B$ stands for the time period when a structural change occurs. Then, Eq. (1) implies, depending upon whether $\rho$ is equal to 1 or not:

- Null: $y_t = \mu_0 + c \cdot B_t + y_t + v_t$ \quad (2a)
- Alternative: $y_t = \mu_1 + \beta \cdot t + (\mu_2 - \mu_1)D_t + v_t$, \quad (2b)

where $v_t$ is stationary and $B_t = 1$ for $t = T_B + 1$ and zero otherwise. This is so, since we have $\mu_0 = \delta_2$, $c = \delta_3$, $y_0 = \delta_1 + x_0$, and $v_t = e_t$ under the null; and $\mu_1 = \delta_1$, $\beta = \delta_2$, $(\mu_2 - \mu_1) = \delta_3$ and $v_t = x_t$ under the alternative. Furthermore, Eq. (1) implies:

$$y_t = d_1 y_{t-1} + d_2 z_t + d_3 z_{t-1} + e_t, \quad (3)$$

where $d_1 = \rho$, $d_2 = \delta$ and $d_3 = -\rho \delta$. Therefore, in Eq. (3) the coefficients of $y_{t-1}$, $z_t$ and $z_{t-1}$ are subject to a nonlinear restriction:

$$d_1 d_2 + d_3 = 0. \quad (4)$$

This is the CFR. In general, if $z_t$ contains $k$ variables, there are $k$ CFRs. The time trend variable and the intercept do not generate a CFR; given intercept and time trend, the lagged time trend variable is redundant. The variable $D_t$ does generate a
CFR: the coefficient of $D_{t-1}$ must be equal to the negative of the coefficient of $y_{t-1}$ times the coefficient of $D_t$. Noticing that $B_t = \Delta D_t$, one can show that Eq. (3) is equivalent to the following nested equation which includes both the null and the alternative models, since Eq. (3) actually includes $D_t$ and $B_t$, and this is equivalent to inclusion of $D_t$ and $D_{t-1}$:

$$y_t = \alpha_1 + \alpha_2 t + \alpha_3 D_t + \alpha_4 B_t + \rho y_{t-1} + \epsilon_t.$$  \hfill (5)

The relationships between Eqs. (1) and (5) are: $\alpha_1 = \delta_1 (1-\rho) + \delta_2 \rho$, $\alpha_2 = \delta_2 (1-\rho)$, $\alpha_3 = \delta_3 (1-\rho)$, and $\alpha_4 = \rho \delta_3$. Then we have the slightly changed but equivalent form of the CFR:

$$\rho \alpha_3 - \alpha_4 (1-\rho) = 0.$$  \hfill (6)

There are two unit root restrictions in Eq. (5): $\rho = 1$ and $\alpha_2 = 0$. The first restriction transforms the CFR in Eq. (6) into $\alpha_3 = 0$. This means that although the CFR is nonlinear, it becomes linear when the unit root restriction is imposed. The validity of the CFR in its general form [Eq. (6)] does not depend on the unit root restriction. However, we cannot test the CFR in Eq. (6) independently of the unit root restriction, at least based on asymptotics, since $\alpha_4$, the coefficient of the regressor $B_t$ in Eq. (5), cannot be consistently estimated due to the fact that $B_t$ is asymptotically degenerate. Therefore, we settle for testing the joint restriction of the unit root and the CFR. Thus we wish to test:

$$H_0: \rho = 1, \quad \alpha_2 = \alpha_3 = 0.$$  \hfill (7)

Perron (1989) and others test only the first unit root restriction ($\rho = 1$), ignoring the two additional restrictions. The missing restrictions can potentially lead to quite different inference. It could well be that the unit root hypothesis is true, but that the other restrictions do not hold, or vice versa. It is typical of the DF regressions to include variables which are relevant under the alternative but not under the null. Here, the dummy variable ($D_t$) representing changed level and the time trend variable ($t$) are variables that belong in Eq. (2b) but not Eq. (2a), and thus are relevant in Eq. (5) under the alternative, but not under the null. The SP or AL test amounts to an LM test of $\rho = 1$, imposing the restriction $\alpha_2 = 0$ and the CFR. Perron's test is a test of $\rho = 1$, not imposing $\alpha_2 = 0$ or the CFR, $\alpha_3 = 0$.

3. Wald statistic

In this section, we develop a Wald statistic which considers the joint hypothesis Eq. (7). To do so, we define:

$$r(\psi) = (\rho - 1, \alpha_2, \alpha_3)' \quad \text{and} \quad R = \frac{dr(\psi)}{d\psi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$
where $\psi=(\rho, \alpha_2, \alpha_3, \alpha_4)'$ are the coefficients of the regressors $w_t=[y_{t-1}^*, t^*, D_t^*, \Delta D_t^*]$'. The asterisk (*) represents deviations from means. Then, the Wald test statistic is defined as:

$$\hat{\xi}_w = r(\hat{\psi})'[R\hat{\sigma}_c^2(W'W)^{-1}R]^{-1}r(\hat{\psi}),$$

where $W=(w_1, \ldots, w_T)'$ and $\hat{\psi}$ is the OLS estimate of $\psi$; and where $\hat{\sigma}_c^2$ is the usual error variance estimate from the regression of $y_t$ on $w_t$. We have the following result for the asymptotics of the Wald statistic.

**Theorem 1.** Assume that $\{y_t\}$ are generated according to Eq. (1) under the null hypothesis of a unit root ($\rho=1$) with $z_t=(1, t, D_t)'$, and the innovations $\epsilon_t$ satisfy the regularity conditions of Phillips and Perron (1988) p. 336). Suppose that there occurs a structural break at time $T_B$, and that $T_B/T \rightarrow \lambda$ as $T \rightarrow \infty$. Then, the asymptotic distribution of the Wald statistic in Eq. (8) follows:

$$\hat{\xi}_w \rightarrow \frac{\sigma^2}{\sigma_c^2} \frac{\mathbf{H}}{g},$$

where

$$H = \begin{pmatrix}
(3\lambda^2 + 3\lambda - 1)/12 & h_1 + 0.5h_2 & 0.5h_1 + h_2/(12\lambda - 12\lambda^2) \\
h_1 + 0.5h_2 & h_2/(\lambda - \lambda^2) - h_3 & -h_1h_2/(\lambda - \lambda^2) - 0.5h_3 \\
0.5h_1 + h_2/(12\lambda - 12\lambda^2) & -h_1h_2/(\lambda - \lambda^2) - 0.5h_3 & (h^2_1 - h_3/12)/(\lambda - \lambda^2)
\end{pmatrix},$$

$$g = h_1 + h_1h_2 + h_3^2/(12\lambda - 12\lambda^2) + h_3(-3\lambda^2 + 3\lambda - 1)/12,$$

$$h_1 = \int_0^1 rW(r)dr - 0.5 \int_0^1 W(r)dr,$$

$$h_2 = \int_0^\lambda rW(r)dr - \int_0^1 W(r)dr,$$

$$h_3 = \int_0^1 W(r)^2 - [\int_0^1 W(r)dr]^2,$$

$$m=[f_1, f_2, f_3]'$$

$$f_1 = 0.5[W(1)^2 - \sigma_c^2/\sigma_c^2] - W(1) \int_0^1 W(r)dr,$$

$$f_2 = 0.5W(1) - \int_0^1 W(r)dr,$$

$$f_3 = W(\lambda) - \lambda W(1)$$

and $\sigma^2 = \lim T^{-1}(\Sigma e_t^2)$ and $\sigma_c^2 = \lim T^{-1}(\Sigma e_t^2)$; and where $W(r)$ is a Brownian motion defined on $[0, 1]$. 
Proofs of theorems are given in Appendix A. Theorem 1 shows that the asymptotic distribution of the Wald statistic depends on $\lambda$ and $\sigma^2 / \sigma_c^2$. We obtain $\sigma^2 / \sigma_c^2 = 1$ under iid errors so that we can tabulate the asymptotic critical values. We obtain them through a simulation of 10,000 replications with sample size $T = 1000$, and present them in Table 1. They do not vary significantly over different values of $\lambda$. This is rather a desirable property since the statistic would be less sensitive to $\lambda$ than otherwise. We used the Gauss software version 3.0 for all computations. The Gauss RNDNS procedure has been used to generate the pseudo-iid $N(0, 1)$ random numbers from an arbitrary seed.

To allow for autocorrelated innovations, we develop a 'corrected' Wald test for which a Phillips and Perron (1988) type correction is used. We define $s^2$ and $s^2(k)$ as in Equations (24) and (25) of SP, using the residuals $\hat{\epsilon}_t$ from an unrestricted regression on Eq. (5); these are consistent estimates of the error variance $\sigma^2_c$ and the long-run variance $\sigma^2$ in Eq. (9), respectively. Define the corrected Wald test statistic by:

$$\hat{\xi}_w^* = \frac{s^2}{s^2(k)} r^*(\hat{\psi}) [R\sigma^2_c(\mathbf{W^W})^{-1}\mathbf{R}]^{-1}r^*(\hat{\psi}),$$

where

$$r^*(\hat{\psi}) = r(\hat{\psi}) - \hat{\pi}i_3^T D_f^* [R(\mathbf{W^W})^{-1}\mathbf{R}]$$

$$\hat{\pi} = 0.5[s^2(k) - s^2]$$

$$i_3 = [1,0,0]'$$

$$D_f^* = \text{diagonal}[T,T^{3/2},T^{1/2}]$$.

**Theorem 2.** Assume that the conditions in Theorem 1 are satisfied. Then, the

---

1 To allow for autocorrelated innovations, one may alternatively use an augmented version of the Wald test by including the augmentation terms $\Delta y_{i-j}$, $j=1,\ldots,k$ in Eq. (5). However, simulation results (not reported here) show that this augmented test has a serious size distortion problem under strongly correlated errors in finite samples.
asymptotic distribution of the corrected Wald test statistic in Eq. (10) follows:

$$\hat{\xi}_{w}^{*} \sim m^{*} - m^{*}, \quad g$$

where $m^{*} = [f_{2}, f_{3}, f_{4}]$ with $f_{4}^{*} = 0.5W(1)^{2} - 0.5 - W(1)\int_{0}^{1}W(r) \, dr$; and where $f_{2}, f_{3}, H$ and $g$ are defined in Theorem 1.

The theorem indicates that the asymptotic distribution of the corrected Wald test statistic is free of the nuisance parameters $\sigma^{2}$ and $\sigma_{\xi}^{2}$, and it is the same as that of the uncorrected Wald test statistic under iid errors.

4. Monte Carlo simulation and empirical results

In this section, we examine the size and power of the Wald statistic by Monte Carlo simulation, and provide empirical results of applying the Wald statistic to the Nelson and Plosser (1982) data. The DGP for the crash model is Eq. (1) with the moving average error structure: $\epsilon_{t} = \eta_{t} + \theta \eta_{t-1}$. The parameter $\lambda$ for a break point is set at 0.5 for the simulation consisting of 5000 replications, and the asymptotic critical values in Table 1 are used for all simulations. In estimating the long-run error variance, the choice of the kernel is not critical and the Parzen kernel is employed for simulation. We use the optimal bandwidth parameter of Andrews (1991) as in Lee and Mossi (1996), and fixed truncation lags $(k)$, which vary according to the sample size. We selected two values, $(k_{1}$ and $k_{2}$); they are $(3, 8)$, $(4, 12)$ and $(6, 15)$ for sample sizes $T=50, 100$ and 500, respectively.

Table 2 contains size and power results for the Wald statistic. Experiment A considers the size of the test under the unit root null in the presence of MA errors, $\theta = -0.8, -0.5, 0, 0.5, 0.8$. When a structural change occurs ($\delta_{3} = 5$), the simulation results display about the same degree of size distortions as do the usual Phillips-Perron (Phillips and Perron, 1988) unit root tests not allowing for the break; see Schwert (1989) for comparison. Experiment B considers the size of the test with different magnitudes of a structural change. The sizes of the tests do not change much, which indicates that they successfully eliminate the effect of structural change. We report this result only for sample size $T=100$ and $\delta_{3}=10$.

Experiment C considers the power of the tests under iid errors (where virtually no significant size distortions are found) for different values of $\rho$. The truncation lag $k=0$ is used in this case. The initial value $y_{0} = x_{0} = 0$ is chosen. The power of the Wald test is 0.138, 0.466 and 0.858 when $\rho = 0.9, 0.8$ and 0.7, respectively. The Wald test is a little less powerful than the Perron test; in the same situations the power of the Perron test is simulated as 0.159, 0.529 and 0.907. This result is not different from our prior expectation that the test of the joint hypothesis is generally less powerful. Nonetheless the difference in power is not very significant. On the other hand, the power of the AL test is computed as 0.261, 0.728 and 0.998; the AL test is more powerful in the same situations than the Wald test or the Perron test.

The empirical results of the corrected Wald test applied to the Nelson–Plosser
data are provided in Table 3. The Fejer kernel is used for estimating the long-run error variance. We supply the results using the selected optimal bandwidth parameter (column 4) and fixed truncation lags, $k=4$ and $12$ (columns 5 and 6). The results indicate that we cannot reject the joint hypothesis of a unit root and the CFR at the 5% level for most of the series that we consider except the ‘industrial production’ series. These results are fairly similar to those of AL reproduced in column 7 and are dissimilar to those of Perron reproduced in column 8. It is not difficult to give the overall conclusion from the empirical results; namely, that most series appear to contain a unit root.

5. Concluding remarks

In this paper, we suggest that the common factor restrictions (CFR) be included in the unit root test procedure in the presence of a structural break. Then we develop a Wald test for the joint hypothesis of a unit root and the CFR. We observe that

---

**Table 2**

Size and power, 5% rejections

<table>
<thead>
<tr>
<th>Exp</th>
<th>$T$</th>
<th>$\rho$</th>
<th>$\delta_3$</th>
<th>$\theta$</th>
<th>Wald statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Auto $k_1$ $k_2$</td>
</tr>
<tr>
<td>A</td>
<td>50</td>
<td>1</td>
<td>5</td>
<td>-0.8</td>
<td>0.995 0.991 0.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.5</td>
<td>0.635 0.592 0.618</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.082 0.088 0.090</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.062 0.066 0.107</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.074 0.068 0.105</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1</td>
<td>5</td>
<td>-0.8</td>
<td>0.999 0.999 1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.5</td>
<td>0.666 0.639 0.753</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.062 0.069 0.060</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.046 0.044 0.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.043 0.055 0.066</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1</td>
<td>5</td>
<td>-0.8</td>
<td>0.999 0.998 1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.5</td>
<td>0.486 0.490 0.611</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.052 0.050 0.053</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.045 0.047 0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.045 0.048 0.034</td>
</tr>
<tr>
<td>B</td>
<td>100</td>
<td>1</td>
<td>10</td>
<td>-0.8</td>
<td>1.00 0.999 1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.5</td>
<td>0.660 0.673 0.771</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.067 0.069 0.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td>0.044 0.039 0.057</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.046 0.043 0.055</td>
</tr>
<tr>
<td>C</td>
<td>100</td>
<td>0.9</td>
<td>5</td>
<td>0</td>
<td>0.153 0.138*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8</td>
<td>0.484 0.466*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7</td>
<td>0.878 0.858*</td>
</tr>
</tbody>
</table>

*a$k=0$ is used for the power of the test.
### Table 3
Empirical Results

<table>
<thead>
<tr>
<th>Series</th>
<th>$T$</th>
<th>$T_B$</th>
<th>Wald statistic</th>
<th>AL$^b$</th>
<th>Perron$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Opt.$^a$ k=4</td>
<td>k=12</td>
<td></td>
</tr>
<tr>
<td>Real GNP</td>
<td>62</td>
<td>21</td>
<td>10.3 (3)</td>
<td>9.99</td>
<td>-3.15*</td>
</tr>
<tr>
<td>Nominal GNP</td>
<td>62</td>
<td>21</td>
<td>13.3 (4)</td>
<td>13.3</td>
<td>-2.58</td>
</tr>
<tr>
<td>Real per capita GNP</td>
<td>62</td>
<td>21</td>
<td>8.55 (3)</td>
<td>7.97</td>
<td>-2.96</td>
</tr>
<tr>
<td>Industrial production</td>
<td>111</td>
<td>70</td>
<td>18.3 (1)</td>
<td>19.0*</td>
<td>-2.64</td>
</tr>
<tr>
<td>Employment</td>
<td>81</td>
<td>40</td>
<td>10.3 (3)</td>
<td>10.3</td>
<td>-2.96</td>
</tr>
<tr>
<td>GNP deflator</td>
<td>82</td>
<td>41</td>
<td>13.6 (4)</td>
<td>13.6</td>
<td>-2.25</td>
</tr>
<tr>
<td>CPI</td>
<td>111</td>
<td>70</td>
<td>2.79 (7)</td>
<td>2.74</td>
<td>-2.73</td>
</tr>
<tr>
<td>Nominal wage</td>
<td>71</td>
<td>30</td>
<td>12.7 (4)</td>
<td>12.7</td>
<td>-3.04</td>
</tr>
<tr>
<td>Money stock</td>
<td>82</td>
<td>41</td>
<td>10.8 (7)</td>
<td>10.0</td>
<td>-3.18*</td>
</tr>
<tr>
<td>Velocity</td>
<td>102</td>
<td>61</td>
<td>6.35 (1)</td>
<td>6.51</td>
<td>-1.71</td>
</tr>
<tr>
<td>Interest rate</td>
<td>71</td>
<td>30</td>
<td>13.7 (0)</td>
<td>9.94</td>
<td>-1.12</td>
</tr>
</tbody>
</table>

$^a$Selected optimal bandwidth estimates are in the parentheses.
$^b$These are obtained from Amsler and Lee (1995), p. 365.
$^c$These are obtained from Perron (1989), p. 1383.
*Significant at the 5% level.
**Significant at the 2.5% level.
***Significant at the 1% level.

The empirical results from the Wald test are different from those that do not impose the CFR. More specifically, the results from the Wald test are fairly consistent with those from the LM test of Amsler and Lee (1995), which tests the unit root hypothesis while imposing the CFR, and are rather different from the results from Perron’s test which does not impose the CFR. Since the Amsler–Lee test is more powerful than either Perron’s test or the Wald test, when the CFR are true, these differences in results do not appear to come from a lack of power. Rather, failing to impose the CFR may lead to a quite different and perhaps misleading inference.

While we consider only a crash model in which a one time structural break occurs, the test can be extended to consider a changing growth model or a model with a sudden change in the level accompanied by a different growth path. We could also consider endogenizing the structural break, as in Banerjee et al. (1992) and Zivot and Andrews (1992). These extensions are not the focus of this paper, but remain as a future research.

**Acknowledgements**

We thank Peter Schmidt for helpful comments and suggestions. We are grateful for the comments of two anonymous referees.
Appendix A

Proofs of theorems

A.1. Proof of Theorem 1

Noticing that $RD_T^{-1} = D_T^* - 1$, for convergence rate matrices, $D_T = \text{diagonal}[T, T^{3/2}, T^{1/2}, I]$ and $D_T^* = \text{diagonal}[T, T^{3/2}, T^{1/2}]$, we obtain:

$$\xi_w = r(\hat{\psi}) D_T^* [R[D_T^{-1}(W'W)D_T^{-1}]^{-1} R']^{-1} D_T^* r(\hat{\psi})/\hat{\sigma}_c^2. \quad (A.1)$$

We decompose the matrix $W$ into submatrices $W_1$ and $W_2$, where $W_1$ and $W_2$ are $T \times 3$ matrix and $T \times 1$ vector with $t$th row (or observation) being $w_{1t} = \{y_{t-1}, t^*, D_t^*\}'$ and $w_{2t} = B_t^*$, respectively. Note that the dummy variable $B_t$ does not affect the asymptotic distribution of the test statistic, since the matrix $D_T^{-1}(W'W)D_T^{-1}$ is asymptotically block diagonal with respect to the sub-matrices, $D_T^{-1}(W_1'W_1)D_T^{-1} - 1$ and $W_2W_2$. Then, $R[D_T^{-1}(W'W)D_T^{-1}]^{-1} R'$ can be replaced with the sub-matrix $[D_T^{-1}(W_1'W_1)D_T^{-1}]^{-1}$. [But it does not necessarily mean that we can exclude $B_t$ in Eq. (5).] Also noting that under the null, $r(\psi) D_T^*$ is same as $r(D_T^*)$, where $r_1 = \{Y_{t-1} e_t^*, \Sigma_t e_t^*, \Sigma D_t^* e_t^*\}'$, we obtain,

$$\xi_w = (D_T^{-1}W_1'r_1) [D_T^{-1}(W_1'W_1)D_T^{-1}]^{-1} (D_T^{-1}W_1'r_1)/\hat{\sigma}_c^2. \quad (A.2)$$

Let

$$W_1'W_1 \equiv \begin{pmatrix} l_0 & l_1 & l_2 \\ l_1 & d_{11} & d_{12} \\ l_2 & d_{12} & d_{22} \end{pmatrix} = \begin{pmatrix} \Sigma y_{t-1}^2 & \Sigma y_{t-1} t^* & \Sigma y_{t-1} D_t^* \\ \Sigma t^* & \Sigma t^* D_t^* & \Sigma D_t^* \end{pmatrix}$$

and

$$(W_1'W_1)^{-1} \equiv \frac{1}{|D|} \begin{pmatrix} W_{11} & W_{12} & W_{13} \\ W_{12} & W_{22} & W_{23} \\ W_{13} & W_{23} & W_{33} \end{pmatrix},$$

where $|D|$ is the determinant. Then,

$$[D_T^{-1}(W_1'W_1)D_T^{-1}]^{-1} = \frac{1}{T^{-6}|D|} \begin{pmatrix} T^{-4} W_{11} & T^{-7/2} W_{12} & T^{-9/2} W_{13} \\ T^{-7/2} W_{12} & T^{-3} W_{22} & T^{-4} W_{23} \\ T^{-9/2} W_{13} & T^{-4} W_{23} & T^{-5} W_{33} \end{pmatrix}. \quad (A.3)$$
Now, consider the asymptotic results:

(a) \( T^{-2}l_0 = T^{-2}\Sigma y_i^*y_{i-1}^* \rightarrow \sigma^2 h_3 \)

(b) \( T^{-5/2}d_1 = T^{-5/2}\Sigma t^*y_{i-1}^* \rightarrow \sigma h_1 \)

(c) \( T^{-3/2}d_2 = T^{-3/2}\Sigma y_i^*D_t^* \rightarrow \sigma h_2 \)

(d) \( T^{-3}d_{11} = T^{-3}\Sigma t^*t^2 \rightarrow 1/12 \)

(e) \( T^{-2}d_{12} = T^{-2}\Sigma t^*D_t^* \rightarrow -0.5(1-\lambda)\lambda \)

(f) \( T^{-1}d_{22} = T^{-1}\Sigma D_t^*D_t^* \rightarrow (1-\lambda)\lambda \),

where \( h_1, h_2 \) and \( h_3 \) are defined in Theorem 1. Letting \( r_1 = [e_1, e_2, e_3]' \), we have additional asymptotic results:

(g) \( T^{-1}e_1 = T^{-1}\Sigma y_{i-1}^*e_{i-1}^* \rightarrow \sigma^2 f_1 \)

(h) \( T^{-3/2}e_2 = T^{-3/2}\Sigma t^*e^* \rightarrow \sigma f_2 \)

(i) \( T^{-1/2}e_3 = T^{-1/2}\Sigma D_t^*e^* \rightarrow \sigma f_3 \),

where \( f_1, f_2 \) and \( f_3 \) are defined in Theorem 1. Then, we obtain for each term in Eq. (A.3):

\[
T^{-6[D]} = -(T^{-2}d_{12})^2 T^{-2}l_0 + T^{-3}d_{11} T^{-1}d_{22} T^{-2}l_0 - T^{-1}d_{22}(T^{-5/2}l_1)^2
+ 2T^{-2}d_{12} T^{-5/2}l_1 T^{-3/2}l_2 - T^{-3}d_{11}(T^{-3/2}l_2)^2 \rightarrow -(1-\lambda)\lambda\sigma^2 g
\]

\[
T^{-4}W_{11} = -(T^{-2}d_{12})^2 + T^{-3}d_{11} T^{-1}d_{22} \rightarrow -(1-\lambda)\lambda\sigma^2 [(-3\lambda^2 + 3\lambda - 1)/(12\sigma^2)]
\]

\[
T^{-7/2}W_{12} \rightarrow -(1-\lambda)\lambda\sigma^2 [h_1/(\sigma + h_2/(2\sigma))]
\]

\[
T^{-9/2}W_{13} \rightarrow -(1-\lambda)\lambda\sigma^2 [h_1/(2\sigma) + h_2/(12(1-\lambda)\lambda\sigma)]
\]

\[
T^{-3}W_{22} \rightarrow -(1-\lambda)\lambda\sigma^2 [h_2^2/((1-\lambda)\lambda) - h_3]
\]

\[
T^{-4}W_{23} \rightarrow -(1-\lambda)\lambda\sigma^2 [-h_3/2 - h_1 h_2/((1-\lambda)\lambda)]
\]

\[
T^{-5}W_{33} \rightarrow -(1-\lambda)\lambda\sigma^2 [(h_1^2 - h_3/12)/((1-\lambda)\lambda)].
\]

Then we obtain:

\[
[D^+_O^{-1}(W_1'O_1)D^+_O^{-1}]^{-1} \rightarrow \frac{H^*}{g},
\]

where \( H^* \) is the same as \( H \) in Theorem 1 except that the first row of \( H^* \) is given as the first row of \( H \) multiplied by \( I_{O}^* = \text{diagonal}[1/\sigma^2, 1/\sigma, 1/\sigma] \). We also obtain:

\[
D^+_O^{-1}r_1 \rightarrow m^* = [\sigma^2 f_1, \sigma f_2, \sigma f_3]'.
\]

Now, note that \( H^* = I_{O} \cdot H \cdot I_{O} \) and \( I_{O}m^* = \sigma m \), where \( I_{O} = \text{diagonal}[1/\sigma, 1, 1] \). Then
it is directly shown from Eqs. (A.4) and (A.5) that:

\[
\xi_w \rightarrow \frac{1}{\sigma^2_c} m'' H^* \frac{H}{g} m' - m,
\]

which proves Theorem 1.

A.2. Proof of Theorem 2

Define \( f_1^* = f_1 - \gamma \), where \( \gamma = 0.5(\sigma^2 - \sigma^2_e)/\sigma^2 \). Also define \( m^* = [f_1^*, f_2, f_3]' \) such that \( m = m^* + \gamma i_3 \) where \( i_3 = (1, 0, 0)' \). Then the asymptotic distribution Eq. (A.6) of the Wald test under \( iid \) errors is expressed as:

\[
\xi_w \rightarrow m^* \frac{H}{g} m^*.
\]

Here, \( H, g \) and \( m^* \) do not depend on error variances, so that the above expression is free of the nuisance parameters. We want to show that the corrected Wald test follows the same asymptotic distribution in Eq. (A.7). First, we alternatively show the expression in Eq. (A.6) by:

\[
\frac{\sigma^2}{\sigma^2_c} m' \frac{H}{g} m - \frac{\sigma^2}{\sigma^2_c} (m^* + \gamma i_3)' \left( \frac{H}{g} m^* + \gamma i_3 \right) = \frac{\sigma^2}{\sigma^2_c} m^* \frac{H}{g} m^* + 2 \frac{\sigma^2}{\sigma^2_c} \left[ (\sigma^2 - \sigma^2_c)/2\sigma^2 \right] i_3 \frac{H}{g} m^* + \frac{\sigma^2}{\sigma^2_c} \left[ (\sigma^2 - \sigma^2_c)/2\sigma^2 \right]^2 i_3 \frac{H}{g} i_3.
\]

Now, we rewrite the corrected Wald test in Eq. (10) after some algebra as:

\[
\frac{\sigma^2}{\sigma^2_c} m^* \frac{H}{g} m^* \left[ [\sigma^2 - \sigma^2_c]/2\sigma^2 \right] i_3 \frac{H}{g} m^* \left[ [\sigma^2 - \sigma^2_c]/2\sigma^2 \right]^2 i_3.
\]

We note that the first term of the last equation is equivalent to the expression for the 'uncorrected' Wald test in Eq. (8), and its asymptotic distribution is given in Eq. (A.6). Also note that \( \pi \rightarrow \pi = \gamma \sigma^2 \). The asymptotic distribution of the second term is given after some algebra by:

\[
-2\pi i_3 \frac{H^*}{g} m^* \frac{1}{\sigma^2_c} (\sigma^2 - \sigma^2_c) i_3 \frac{H}{g} m^* \frac{1}{\sigma^2_c} - (\sigma^2 - \sigma^2_c) i_3 \frac{H}{g} [\sigma^2 - \sigma^2_c] i_3.
\]

The third term of Eq. (A.9) follows asymptotically:

\[
[(\sigma^2 - \sigma^2_c)/2\sigma^2]^2 \frac{\sigma^2}{\sigma^2_c} i_3 \frac{H}{g} i_3.
\]

Finally, after rearranging Eqs. (A.7), (A.10), (A.11), we can show that the asymptotic distribution of the whole expression in Eq. (A.9) is the same as that of the
distribution of the uncorrected Wald test statistic under iid errors [which is expressed in Eq. (A.7), and is free of nuisance parameters].

References