1 Asymptotics of Polynomial Time Trend Estimation and Hypothesis Testing under Rank Deficiency

PETER C. B. PHILLIPS *Yale University, University of Auckland, and Singapore Management University*

1.1 Introduction

Under very general conditions Grenander and Rosenblatt (1954) proved a major theorem which showed that in regression with polynomial and trigonometric polynomial trend regressors and stationary errors simple application of ordinary least squares (OLS) is asymptotically equivalent to generalized least squares (GLS) and thereby efficient. This important property provided an asymptotic justification for the use of OLS in trend regression and made trend elimination by polynomial time trend regression a popular procedure in empirical research. A key condition underlying this Grenander–Rosenblatt (GR) theorem is that the regression errors are stationary and have positive spectral density at the origin, which ensures that the long-run variance of the errors is necessarily positive. A similar finding to the GR theorem holds in the case of trend regression with integrated regressors, so the result is not confined to deterministic regressors but includes pure cointegrated regression models in which the regressors are strictly exogenous (Phillips and Park, 1988). While these asymptotic properties, including the conditions for the limiting equivalence of OLS-GLS, are well established, there remain some interesting unexplored aspects of trend regression.

This chapter studies two such issues that can arise in estimation and inference with polynomial time trend regression: (i) problems in the limit theory that arise from a moving average (MA) error with a unit root that produces a zero error spectrum at the origin and leads to failure in standard invariance principle arguments and failure in the GR theorem; and (ii) problems of rank deficiency

This chapter provides an extended solution to a graduate take-home examination at Singapore Management University in April 2022. The examination problem provided a challenge similar to that of a research project, working from a familiar foundation towards arenas that are unfamiliar and that require some technical novelty and notational innovation to resolve. Such challenges are part of normal research and play an important role in graduate student apprenticeship training in econometrics. Research support is acknowledged from NSF Grant No. SES 18-50860, a Kelly Fellowship at the University of Auckland, and a Lee Kong Chian Fellowship at Singapore Management University.

and singularity that arise in trend regression when testing multiple hypotheses using Wald statistics. Both of these might be considered graduate student problems. They are not dealt with in graduate textbooks of econometrics but, as shown here, they are the type of problem that can easily arise in research when dealing with new models and nonstationary data. We focus on the simplest case of deterministic trend regression. But related issues do occur in cases of stochastic regressors (Phillips, 1987, 1995; Park and Phillips, 1988, 1989) as well as explosive model cases (Phillips and Magdalinos, 2013) and nonparametric kernel regression problems (Phillips et al., 2017) where unexpected singularities can occur in the asymptotic theory. A general treatment of rank deficiency and singularities in econometric testing is available in Magdalinos and Phillips (2018). In spite of the simplicity of the present deterministic trend model, considerable technical complexities arise in the general case, which require some innovations and notational aids to resolve adequately and to secure definitive results. These techniques should have wider applicability beyond the trend regression models of the present study.

This chapter is largely pedagogical in nature and is written specifically for this advanced textbook series. The development proceeds from a common model of deterministic trend regression where standard asymptotics hold to one where the limit theory and convergence rates of the usual OLS estimator and associated test statistics differ from the standard theory. To aid exposition a progression of examples are given that assist in building up the notation and techniques required to resolve the general case under study. It should therefore be accessible to graduate students and some undergraduates with advanced training in econometrics.

The trend regression model is given in the next section with key conditions that enable the asymptotic development given in Section 1.3 for estimation and in Section 1.4 for inference. Section 1.5 provides some final discussion. A brief appendix on Wiener stochastic integration is included for completeness to facilitate development of the limit theory.

1.2 Deterministic Trend Regression

We study limit theory in the simple linear trend regression model

$$y_t = x_t'\beta + u_t, \quad t = 0, 1, \ldots, n \tag{1.1}$$

where the deterministic trend regressor $x_t = (1, t, t^2, \ldots, t^m)'$ and the errors u_t are generated by the linear process $u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ with coefficients c_j satisfying the summability condition $\sum_{j=0}^{\infty} j|c_j| < \infty$ and innovations $\varepsilon_t \sim iid\left(0, \sigma^2\right)$.

Using the Phillips–Solo device (Phillips and Solo, 1992) gives the valid Beveridge–Nelson decomposition

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} = C(1)\varepsilon_t - (1-L)\tilde{C}(L)\varepsilon_t = C(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t, \tag{1.2}$$

where $\tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t$ and $\tilde{C}(L) = \sum_{j=0}^{\infty} \tilde{c}_j L^j$ with $\tilde{c}_j = \sum_{s=j+1}^{\infty} c_s$. The series representation $\tilde{\varepsilon}_t = \sum_{j=0}^{\infty} \tilde{c}_j \varepsilon_{t-j}$ converges almost surely in view of the summability condition $\sum_{j=0}^{\infty} j|c_j| < \infty$. The long-run behavior of u_t is primarily determined by the long-run moving average coefficient $C(1)$, which directly affects the value of the spectral density at the origin and is therefore instrumental in the long-run properties of the time series u_t.

The $(m+1) \times 1$ parameter vector $\beta = (\beta_0, \beta_1, \ldots, \beta_m)'$ in (1.1) is estimated by ordinary least squares (OLS) regression using $\hat{\beta} = \left(X'X\right)^{-1} X'y$ where $X' = [x_1, \ldots, x_n]$ and $y' = [y_1, \ldots, y_n]$. The remainder of the chapter considers the asymptotic theory of $\hat{\beta}$ and Wald tests based on $\hat{\beta}$ under different conditions that determine the long-run behavior of the equation error u_t. We look at two cases. The first is the standard model where $C(1) \neq 0$; the second is the degenerate unit root moving average case where $C(1) = 0$.

When $C(1) = 0$ it is clear from the decomposition (1.2) that the regression error u_t has the MA unit root form $u_t = -(1-L)\tilde{\varepsilon}_t$. Models with MA unit roots or near unit roots can arise in empirical work with time series data when there is overdifferencing or where multicointegration is present (Kheifets and Phillips, 2023; Phillips and Kheifets, 2024). The GR asymptotic equivalence theorem fails in such cases because the spectral density of u_t is $f_u(\lambda) = \frac{\sigma^2}{2\pi}|C(e^{i\lambda})|^2 = \frac{\sigma^2}{2\pi}|1 - e^{i\lambda}|^2 |\tilde{C}(e^{i\lambda})|^2$ and so at the origin $f_u(0) = 0$. As will be shown, the asymptotic theory for $\hat{\beta}$ changes considerably. Convergence rates of the components of $\hat{\beta}$ differ from when $C(1) \neq 0$ and conventional invariance principle asymptotics fail, causing difficulties with inference. A simple procedure for dealing with this failure is developed.

In both $C(1) \neq 0$ and $C(1) = 0$ cases, Wald tests of general hypotheses concerning β suffer from rank deficiency and singularity asymptotically that arise from the differing signal strengths in the nonstationary components of the regressor x_t and consequent differing convergence rates in the components of $\hat{\beta}$. These differences affect the Wald statistic limit theory and hypothesis testing. Related problems of rank deficiency in cointegrated regression and VARs were studied in Phillips (1995). Readers are referred to Magdalinos and Phillips (2018) for the first general treatment of matrix normalization problems in inference.

1.3 Limit Theory

1.3.1 The Standard Case: $C(1) \neq 0$

The $(m+1) \times 1$ parameter vector β in (1.1) is estimated by ordinary least squares regression using $\hat{\beta} = \left(X'X\right)^{-1} X'y$ where $X' = [x_1, \ldots, x_n]$ and $y' = [y_1, \ldots, y_n]$. The limit behavior of $\hat{\beta}$ follows by standard manipulations.

Let $D_n = \mathrm{diag}\{1, n, n^2, \ldots, n^m\}$. For $r \in [0,1]$, $D_n^{-1} x_{\lfloor nr \rfloor} \to X(r) = (1, r, r^2, \ldots, r^m)'$ as $n \to \infty$. It follows by Riemann integration that

$$n^{-1} D_n^{-1} X'XD_n^{-1} = n^{-1} \sum_{t=1}^{n} D_n^{-1} x_t x_t' D_n^{-1} \to \int_0^1 X(r)X(r)'dr \\ =: M, \quad \text{as } n \to \infty. \tag{1.3}$$

Under the given conditions, normalized partial sums of u_t satisfy the functional central limit law $n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} u_t \rightsquigarrow B_u(r)$ where B_u is Brownian motion with variance given by the long-run variance $\omega^2 = \sigma^2 C(1)^2$ of u_t. Setting $u' = [u_1, \ldots, u_n]$ we find using Wiener integral limit theory (see Lemma 1.6.1 in the Appendix) that

$$\frac{1}{\sqrt{n}} D_n^{-1} X'u = \sum_{t=1}^{n} D_n^{-1} x_t \frac{u_t}{\sqrt{n}} \rightsquigarrow \int_0^1 X(r) dB_u(r). \tag{1.4}$$

It follows directly from (1.3) and (1.4) that

$$\sqrt{n} D_n(\hat{\beta} - \beta) = \left(\frac{1}{n} D_n^{-1} X'XD_n^{-1}\right)^{-1} \frac{1}{\sqrt{n}} D_n^{-1} X'u \\ \rightsquigarrow \left(\int_0^1 X(r)X(r)'dr\right)^{-1} \left(\int_0^1 X(r)dB_u(r)\right) = \mathcal{N}\left(0, \omega^2 M^{-1}\right), \tag{1.5}$$

since $\int_0^1 X(r)dB_u(r) =_d \mathcal{N}\left(0, \omega^2 M\right)$.

1.3.2 The Degenerate Case: $C(1) = 0$

Using partial summation as in the proof of Lemma 1.6.1 we obtain

$$D_n^{-1} X'u = D_n^{-1} \sum_{t=1}^{n} x_t \Delta\tilde{\varepsilon}_t = D_n^{-1}(x_n \tilde{\varepsilon}_n - x_0 \tilde{\varepsilon}_0) - D_n^{-1} \sum_{t=1}^{n} \Delta x_t \tilde{\varepsilon}_{t-1} \\ \rightsquigarrow X(1)\tilde{\varepsilon}_\infty - e_1 \tilde{\varepsilon}_0, \tag{1.6}$$

where $D_n^{-1} x_n \to X(1)$ and $\tilde{\varepsilon}_n \rightsquigarrow \tilde{\varepsilon}_\infty =_d \tilde{\varepsilon}_t$ as $n \to \infty$, $e_1 = (1, 0, \ldots, 0)'$ and $D_n^{-1} x_0 = e_1$. Note also that $\Delta x_t \approx (0, 1, 2t, \ldots, mt^{m-1})'$ for large t and then

$\sqrt{n}D_n^{-1}\Delta x_{t=\lfloor nr\rfloor} \sim_a (0, \frac{1}{n^{1/2}}, \frac{2\lfloor nr\rfloor}{n^{3/2}}, \ldots, \frac{m\lfloor nr\rfloor^{m-1}}{n^{m-1/2}})' = O(\frac{1}{\sqrt{n}}) \to 0$ as $n \to \infty$ for all $r \in [0,1]$. It follows that

$$D_n^{-1}\sum_{t=1}^{n}\Delta x_t\tilde{\varepsilon}_{t-1} = \sum_{t=1}^{n}\sqrt{n}D_n^{-1}\Delta x_t\frac{\tilde{\varepsilon}_{t-1}}{\sqrt{n}} \to_p 0, \tag{1.7}$$

confirming (1.6). We deduce that

$$nD_n(\hat{\beta}-\beta) = \left(\frac{1}{n}D_n^{-1}X'XD_n^{-1}\right)^{-1}D_n^{-1}X'u \rightsquigarrow M^{-1}\left(X(1)\tilde{\varepsilon}_\infty - e_1\tilde{\varepsilon}_0\right), \tag{1.8}$$

so that $\hat{\beta}$ is consistent with convergence rate nD_n but no invariance principle applies, leading to difficulties with robust inference about β. However, observe that the convergence rate of $\hat{\beta}$ in (1.8) is nD_n, which is $O(\sqrt{n})$ faster than the convergence rate $\sqrt{n}D_n$ for the same estimator in the case where $C(1) \neq 0$. Thus, degeneracy in the long-run behavior of the equation error leads to an acceleration in convergence. But it does not lead to an invariance principle at least for this OLS estimator because its limit distribution depends on the distribution of $\tilde{\varepsilon}_t$, thereby causing the difficulties for inference.

If $\tilde{\varepsilon}_0 = 0$ then

$$nD_n(\hat{\beta}-\beta) \rightsquigarrow \left(\int_0^1 X(r)X(r)'dr\right)^{-1}X(1)\tilde{\varepsilon}_\infty = M^{-1}X(1)\tilde{\varepsilon}_\infty, \tag{1.9}$$

which has mean zero and variance $\mathbb{V}ar(\tilde{\varepsilon}_\infty)M^{-1}X(1)X(1)'M^{-1}$, which is singular of rank unity when $m > 0$. When $m = 0$ and $x_t = 1$, we have $X(r) = 1$ and the limit variance is simply $\mathbb{V}ar(\tilde{\varepsilon}_\infty)$. But when $m > 1$ there is a (nonrandom) direction where the convergence rate is faster than nD_n because of the singularity in (1.9). In particular, if $G_\perp$ is an $(m+1)\times m$ orthogonal complement of the vector $g = M^{-1}X(1)$ then it follows from (1.9) that $nG_\perp'D_n(\hat{\beta}-\beta) = o_p(1)$. This degenerate feature of the limit theory indicates that there may be another procedure that can accelerate convergence, which we now consider.

1.3.3 Use of Temporal Aggregation

Temporal aggregation of the model (1.1) leads to the equation

$$Y_t = X_t'\beta + U_t, \quad t = 1, \ldots, n, \tag{1.10}$$

where $Y_t = \sum_{s=1}^{t} y_s$, $X_t = \sum_{s=1}^{t} x_s = \sum_{s=1}^{t}(1, s, s^2, \ldots, s^m)'$ and when $C(1) = 0$ we have $U_t = \sum_{s=1}^{t}\Delta\tilde{\varepsilon}_s = \tilde{\varepsilon}_t - \tilde{\varepsilon}_0$, so that

$$Y_t = X_t'\beta + \tilde{\varepsilon}_t - \tilde{\varepsilon}_0, \quad t = 1, \ldots, n. \tag{1.11}$$

The parameter vector β in (1.10) is again estimated by OLS regression giving $\tilde{\beta} = \left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\tilde{Y}$ where $\tilde{X}' = [X_1, \ldots, X_n]$ and $\tilde{Y}' = [Y_1, \ldots, Y_n]$. Similarly, we define $\tilde{U}' = [U_1, \ldots, U_n] = [\tilde{\varepsilon}_1 - \tilde{\varepsilon}_0, \ldots, \tilde{\varepsilon}_n - \tilde{\varepsilon}_0]$.

To find the limit behavior of $\tilde{\beta}$ we use $F_n = nD_n$ and note that $F_n^{-1}X_{\lfloor nr \rfloor} = \frac{1}{n}\sum_{s=1}^{\lfloor nr \rfloor} D_n^{-1}x_s \to \tilde{X}(r) := \int_0^r X(s)ds = \int_0^r (1, s, s^2, \ldots, s^m)' ds = \left(r, \frac{r^2}{2}, \ldots, \frac{r^{m+1}}{m+1}\right)'$. It follows by Riemann integration that

$$\frac{1}{n}\sum_{t=1}^{\lfloor nr \rfloor} F_n^{-1}X_t = \frac{1}{n}\sum_{t=1}^{\lfloor nr \rfloor}\frac{1}{n}\sum_{s=1}^{t} D_n^{-1}x_s \to \int_0^r \tilde{X}(p)dp, \tag{1.12}$$

and

$$\frac{1}{n}F_n^{-1}\tilde{X}'\tilde{X}F_n^{-1} = n^{-1}\sum_{t=1}^{n} F_n^{-1}X_tX_t'F_n^{-1} \to \int_0^1 \tilde{X}(r)\tilde{X}(r)'dr =: \tilde{M}, \tag{1.13}$$

as $n \to \infty$. Next observe that

$$\begin{aligned}\frac{1}{n}F_n^{-1}\tilde{X}'\tilde{U} &= \frac{1}{n}\sum_{t=1}^{n} F_n^{-1}X_t\tilde{\varepsilon}_t - \frac{1}{n}\sum_{t=1}^{n} F_n^{-1}X_t\tilde{\varepsilon}_0 \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} F_n^{-1}X_t\frac{\tilde{\varepsilon}_t}{\sqrt{n}} - \frac{1}{n}\sum_{t=1}^{n} F_n^{-1}X_t\tilde{\varepsilon}_0 \\ &= O_p\left(\frac{1}{\sqrt{n}}\right) - \left(\frac{1}{n}\sum_{t=1}^{n} F_n^{-1}X_t\right)\tilde{\varepsilon}_0 \rightsquigarrow -\int_0^1 \tilde{X}(p)dp\ \tilde{\varepsilon}_0.\end{aligned} \tag{1.14}$$

To confirm (1.14) first note that under the stated summability conditions $\tilde{\varepsilon}_t = \tilde{C}(L)\varepsilon_t$ is zero mean stationary with long-run variance $\tilde{\omega}^2 = \sigma^2\tilde{C}(1)^2$ and so partial sums of $\tilde{\varepsilon}_t$ satisfy the functional law $n^{-1/2}\sum_{s=1}^{\lfloor nr \rfloor}\tilde{\varepsilon}_s \rightsquigarrow B_{\tilde{\varepsilon}}(r)$ where $B_{\tilde{\varepsilon}}$ is Brownian motion with variance $\tilde{\omega}^2 > 0$ provided $\tilde{C}(1) \neq 0$. It follows by using the summation by parts argument in Lemma 1.6.1 that $\sum_{t=1}^{n} F_n^{-1}X_t\frac{\tilde{\varepsilon}_t}{\sqrt{n}} \rightsquigarrow \int_0^1 \tilde{X}(r)dB_{\tilde{\varepsilon}}(r)$ so that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} F_n^{-1}X_t\frac{\tilde{\varepsilon}_t}{\sqrt{n}} = O_p\left(\frac{1}{\sqrt{n}}\right)$, which leads to (1.14). If $\tilde{C}(1) = 0$, then a further level of complexity arises that can be treated by similar methods but which we do not address in the present chapter. We deduce from (1.13) and (1.14) that

$$\begin{aligned}F_n(\tilde{\beta} - \beta) &= \left(\frac{1}{n}F_n^{-1}\tilde{X}'\tilde{X}F_n^{-1}\right)^{-1}\left(\frac{1}{n}F_n^{-1}\tilde{X}'\tilde{U}\right) \\ &\rightsquigarrow -\tilde{M}^{-1}\int_0^1 \tilde{X}(p)dp\ \tilde{\varepsilon}_0,\end{aligned} \tag{1.15}$$

so that $\tilde{\beta}$ is consistent with rate of convergence $F_n = nD_n$ but no invariance principle applies, leading to continuing difficulties with robust inference about β.

Remark 1.3.1 *We deduce that temporal aggregation of the data and estimation of* (1.10) *raises the convergence rate of OLS regression from the rate $\sqrt{n}D_n$ for $\hat{\beta}$ by $O(\sqrt{n})$ to the rate $F_n = nD_n$ for $\tilde{\beta}$. This rise in the convergence rate is explained by the rise in the signal strength associated with the signal matrix $\tilde{X}'\tilde{X}$ in the regression formula $(\frac{1}{n}F_n^{-1}\tilde{X}'\tilde{X}F_n^{-1})^{-1}(\frac{1}{n}F_n^{-1}\tilde{X}'\tilde{U})$ in* (1.15)*. Temporal aggregation is a simple mechanism for taking into account the MA unit root and zero long-run variance of u_t in the general model* (1.1) *when $C(1) = 0$, leading to* (1.10)*. This procedure may therefore be regarded as an important partial step toward implementing GLS. It is not a full implementation of GLS because the residual $\tilde{\varepsilon}_t$ in* (1.10) *is generally autocorrelated with spectrum $f_{\tilde{\varepsilon}}(\lambda) = \sigma^2|\tilde{C}(e^{i\lambda})|^2$ that is not necessarily flat. Thus, taking account of the MA unit root and the zero long-run variance of u_t leads to accelerated convergence and asymptotic normality. So the conventional OLS trend regression estimator $\hat{\beta}$ is asymptotically infinitely deficient by a $\sqrt{n}$ factor, relative to the partial use of GLS that leads to $\tilde{\beta}$, manifesting the failure of the GR asymptotic equivalence theorem.*

The rise in convergence rate that results from the use of OLS regression in the temporally aggregated model does not resolve difficulties concerning inference because the limit theory (1.15) involves the distribution of $\tilde{\varepsilon}_0$. This difficulty in the limit theory is addressed by an adjustment to the estimation technique, as explained in the next section.

1.3.4 The Effect of Intercept Adjustment

We propose to estimate the parameter vector β in (1.10) by OLS regression with a fitted intercept in the regression. The model may therefore be written in the form

$$Y_t = X_t'\beta + \tilde{\varepsilon}_t - \tilde{\varepsilon}_0 = \mu + X_t'\beta + \tilde{\varepsilon}_t,\ t = 1, \ldots, n \tag{1.16}$$

with (random) intercept $\mu = \tilde{\varepsilon}_0$. The presence of the intercept in the OLS regression leads to all variables in the model being demeaned. We use the daggered affix variable $A_t^\dagger$ to signify that the variable A_t is demeaned, so that $A_t^\dagger := A_t - \frac{1}{n}\sum_{s=1}^n A_s$. For the matrix of observations $A = [A_1, \ldots, A_n]'$ of A_t we correspondingly denote the observation matrix of demeaned variables $A_t^\dagger$ by $A^\dagger$.

With this notation in hand, the algebra and asymptotics in the aggregated model with a fitted intercept are determined with the following modifications of those given above. In particular, setting $X_t^{\dagger} := X_t - \frac{1}{n}\sum_{s=1}^{n} X_s$ and by taking limits as $n \to \infty$ and Riemann integration we have

$$F_n^{-1}X_{\lfloor nr \rfloor}^{\dagger} = \frac{1}{n}\sum_{s=1}^{\lfloor nr \rfloor} D_n^{-1}x_s - \frac{1}{n}\sum_{t=1}^{n}\frac{1}{n}\sum_{s=1}^{t} D_n^{-1}x_s$$
$$\to \tilde{X}(r) - \int_0^1 \tilde{X}(p)dp = \int_0^r X(s)ds$$
$$- \int_0^1 \int_0^p X(s)dsdp =: X^{\dagger}(r). \tag{1.17}$$

The limit function $X^{\dagger}(r)$ is the demeaned form of $\tilde{X}(r) = \int_0^r X(s)ds$ in the function space $C[0,1]$ and this function correspondingly satisfies $\int_0^1 X^{\dagger}(r)dr = 0$. We now obtain as $n \to \infty$

$$\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}F_n^{-1} = n^{-1}\sum_{t=1}^{n} F_n^{-1}X_t^{\dagger}X_t^{\dagger\prime}F_n^{-1} \to \int_0^1 X^{\dagger}(r)X^{\dagger}(r)'dr =: M^{\dagger}. \tag{1.18}$$

Next observe that $\sum_{t=1}^{n} X_t^{\dagger}\tilde{\varepsilon}_0 = \tilde{\varepsilon}_0 \sum_{t=1}^{n} X_t^{\dagger} = 0$ by virtue of the definition of the demeaned variable $X_t^{\dagger}$, thereby eliminating the effect of $\tilde{\varepsilon}_0$. It follows that

$$\frac{1}{\sqrt{n}}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{U}^{\dagger} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} F_n^{-1}X_t^{\dagger}\tilde{\varepsilon}_t - \frac{1}{\sqrt{n}}\sum_{t=1}^{n} F_n^{-1}X_t^{\dagger}\tilde{\varepsilon}_0$$
$$= \sum_{t=1}^{n} F_n^{-1}X_t^{\dagger}\frac{\tilde{\varepsilon}_t}{\sqrt{n}} \rightsquigarrow \int_0^1 X^{\dagger}(r)dB_{\tilde{\varepsilon}}(r), \tag{1.19}$$

where $B_{\tilde{\varepsilon}}$ is Brownian motion with variance $\tilde{\omega}^2 = \sigma^2\tilde{C}(1)$ in view of the functional law $n^{-1/2}\sum_{s=1}^{\lfloor nr \rfloor} \tilde{\varepsilon}_s \rightsquigarrow B_{\tilde{\varepsilon}}(r)$ indicated above, which holds provided $\tilde{C}(1) \neq 0$. The limit theory in (1.19) involves Wiener integration with respect to the process $B_{\tilde{\varepsilon}}$ and is validated, as in Lemma 1.6.1, by the bounded variation property of the demeaned trend function $X^{\dagger}(r)$ and the weak convergence of the standardized partial sum process $n^{-1/2}\sum_{s=1}^{\lfloor nr \rfloor} \tilde{\varepsilon}_s$.

Combining the limits in (1.18) and (1.19), we obtain the asymptotic behavior of the OLS estimator $\hat{\beta}^{\dagger} = \left(\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}\right)^{-1}\left(\tilde{X}^{\dagger\prime}\tilde{Y}^{\dagger}\right)$ after centering and scaling, namely,

$$\sqrt{n}F_n(\hat{\beta}^{\dagger} - \beta) = \left(\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}F_n^{-1}\right)^{-1}\left(\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{U}\right)$$

$$\rightsquigarrow \left(\int_0^1 X^\dagger(r)X^\dagger(r)'dr\right)^{-1}\int_0^1 X^\dagger(r)dB_{\tilde{\varepsilon}}(r)$$
$$=_d \mathcal{N}\left(0, \tilde{\omega}^2 M^{\dagger-1}\right). \tag{1.20}$$

As is apparent from (1.20), the OLS estimator $\hat{\beta}^\dagger$ obtained from the temporally aggregated model (1.16) with a fitted intercept in the regression satisfies an invariance principle and is asymptotically normal with a simple covariance matrix structure involving the inverse limiting signal matrix $M^{\dagger-1}$ and the scalar factor $\tilde{\omega}^2$, which is the long-run variance of the equation errors $\tilde{\varepsilon}_t$ in (1.16). It is further apparent that under the given conditions the spectral density $f_{\tilde{\varepsilon}}(\lambda) = \frac{\sigma^2}{2\pi}|\tilde{C}(e^{i\lambda})|^2 \in (0,\infty)$ of $\tilde{\varepsilon}_t$ satisfies the conditions in the GR theorem for the asymptotic equivalence of OLS and GLS. Hence, asymptotic efficiency in the regression (1.16) can be achieved by the use of OLS without any knowledge of the spectral density $f_{\tilde{\varepsilon}}(\lambda)$.

Consistent estimation of $\tilde{\omega}^2$ may be conducted by standard HAC methods applied to the residuals of the regression, namely,

$$\hat{U}_t^\dagger = Y_t - \hat{\mu}^\dagger - X_t'\hat{\beta}^\dagger = Y_t^\dagger - X_t^{\dagger\prime}\hat{\beta}^\dagger, \tag{1.21}$$

where the intercept estimate is $\hat{\mu}^\dagger = \frac{1}{n}\sum_{t=1}^n Y_t - \frac{1}{n}\sum_{t=1}^n X_t'\hat{\beta}^\dagger$. The nonparametric estimate has the usual form $\hat{\tilde{\omega}}^2 = \sum_{\ell=-L}^{L} k\left(\frac{\ell}{L}\right)\hat{\gamma}_{\hat{U}^\dagger}(\ell)$, where $\gamma_{\hat{U}^\dagger}(\ell) = \frac{1}{n}\sum_{1\le t,t+\ell\le n}\hat{U}_t^\dagger\hat{U}_{t+\ell}^\dagger$ and $k(\cdot)$ is a symmetric lag kernel such as the Bartlett or Parzen kernel. Since $\hat{\beta}^\dagger \to_p \beta$ it follows that $\hat{U}_t^\dagger = \tilde{\varepsilon}_t + o_p(1)$ and $\gamma_{\hat{U}^\dagger}(\ell) \to_p \gamma_{\tilde{\varepsilon}}(\ell) = \mathbb{E}(\tilde{\varepsilon}_t\tilde{\varepsilon}_{t+\ell})$ for all finite ℓ as $n \to \infty$. Then the kernel long-run variance estimator $\hat{\tilde{\omega}}^2 \to_p \tilde{\omega}^2$ under standard conditions on the kernel $k(\cdot)$ and the lag truncation parameter $L = L_n \to \infty$ as $n \to \infty$, which are henceforth assumed to hold.

With the limit theory (1.20) and consistent estimation of $\tilde{\omega}^2$ it might appear that inference about the regression coefficients β can be conducted using standard t-tests and Wald tests. However, the asymptotic distribution in (1.20) involves matrix normalization by F_n to account for the differing magnitudes in the trend regressors X_t. This matrix normalization and multiple hypotheses about the coefficients together lead to some additional complexities in the limit theory for Wald tests. A general treatment of hypothesis testing in regression under matrix normalization is developed in Magdalinos and Phillips (2018). The present discussion will focus on the potential deficiencies and their impact on the asymptotic theory induced by the presence of the deterministic trend polynomials X_t and multiple convergence rates in the regression (1.16).

Unlike Wald tests of general hypotheses, inference procedures for individual coefficients β_i follow directly from (1.20). For instance, to test null hypotheses such as $\mathcal{H}_0 : \beta_i = \beta_i^0$, standard t-tests can be employed using the limit theory

$$t_{\beta_i} = \frac{\hat{\beta}_i^\dagger - \beta_i^0}{\hat{\tilde{\omega}}^2 \left[\left(\tilde{X}^{\dagger\prime} \tilde{X}^\dagger \right)^{-1} \right]_{ii}} \rightsquigarrow \mathcal{N}(0,1), \text{ as } n \to \infty, \tag{1.22}$$

where $[A]_{ii}$ denotes the ith diagonal element of the matrix A. Tests of multiple hypotheses involving several of the coefficients β_i need to respect the different rates of convergence of the elements $\hat{\beta}_i$ that are evident in (1.20). Coefficient estimates with lower convergence rates then dominate the limit theory, which has some interesting features that are explored in Section 1.4.

1.3.5 Temporal Aggregation in the Standard Case Where $C(1) \neq 0$

We next consider the effect of estimating the parameter vector β in the temporally aggregated model (1.10) by OLS regression with a fitted intercept in the regression but in the case where $C(1) \neq 0$ and there is no degeneracy induced by an MA unit root. The model may therefore be written in the form

$$Y_t = X_t'\beta + U_t = \mu + X_t'\beta + U_t, \; t = 1, \ldots, n \tag{1.23}$$

with $\mu = 0$ and $U_t = \sum_{s=1}^t u_s$. The OLS estimator $\hat{\beta}^\dagger = (\tilde{X}^{\dagger\prime}\tilde{X}^\dagger)^{-1}(\tilde{X}^{\dagger\prime}\tilde{Y}^\dagger)$ has the same form as before but now the estimation error is $\hat{\beta}^\dagger - \beta = (\tilde{X}^{\dagger\prime}\tilde{X}^\dagger)^{-1}(\tilde{X}^{\dagger\prime}\tilde{U})$ and in place of (1.19) the sample covariance factor has the following limit:

$$\frac{1}{n^{3/2}} F_n^{-1} \tilde{X}^{\dagger\prime} \tilde{U} = \frac{1}{n} \sum_{t=1}^n F_n^{-1} X_t^\dagger \frac{U_t}{\sqrt{n}} \rightsquigarrow \int_0^1 X^\dagger(r) B_u(r) dr. \tag{1.24}$$

Upon centering and scaling, the limit distribution follows by standard methods giving

$$\begin{aligned} \frac{1}{\sqrt{n}} F_n(\hat{\beta}^\dagger - \beta) &= \left(\frac{1}{n} F_n^{-1} \tilde{X}^{\dagger\prime} \tilde{X}^\dagger F_n^{-1} \right)^{-1} \left(\frac{1}{n^{3/2}} F_n^{-1} \tilde{X}^{\dagger\prime} \tilde{U} \right) \\ &\rightsquigarrow M^{\dagger -1} \int_0^1 X^\dagger(r) B_u(r) dr \\ &=_d \mathcal{N}\left(0, \omega^2 M^{\dagger -1} \int_0^1 \int_0^1 X^\dagger(r) r \wedge s X^\dagger(s)' dr ds M^{\dagger -1} \right). \end{aligned} \tag{1.25}$$

The rate of convergence of $\hat{\beta}^\dagger$ is $\frac{1}{\sqrt{n}} F_n$ and is therefore slower by $O(n)$ than the MA unit root case where the error is $u_t = \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t$. Also, the variance matrix in the limit distribution (1.25) has a sandwich form, which arises from the fact that the temporally aggregated model (1.23) is a polynomial regression with a highly autocorrelated $I(1)$ error process U_t. The regressor in (1.23) is, for large t, $X_t = \sum_{s=1}^t x_s \sim_a (t, t^2, \ldots, t^{m+1})'$, whose component of lowest degree is the linear trend t which has the sample signal $\sum_{t=1}^n t^2 = O(n^3)$. This signal of $O(n^3)$ exceeds that of the $I(1)$ error U_t, which is $\sum_{t=1}^n U_t^2 = O_p(n^2)$), by order $O_p(n)$. It follows that the OLS estimator $\hat{\beta}^\dagger$ is consistent with the lowest convergence rate $O(\sqrt{n})$ for the coefficient β_0.

The fitted intercept $\hat{\mu}^\dagger$, on the other hand, is inconsistent. To show this, note that $\hat{\mu}^\dagger = \frac{1}{n}\sum_{t=1}^n Y_t - \frac{1}{n}\sum_{t=1}^n X_t'\hat{\beta}^\dagger$ and

$$\begin{aligned}\frac{1}{\sqrt{n}}\hat{\mu}^\dagger &= \frac{1}{n}\sum_{t=1}^n \frac{U_t}{\sqrt{n}} - \frac{1}{n}\sum_{t=1}^n X_t'F_n^{-1}\frac{1}{\sqrt{n}}F_n\left(\hat{\beta}^\dagger - \beta\right) \\ &\rightsquigarrow \int_0^1 B_u(r)dr - \int_0^1 X(r)'dr\left(\int_0^1 X^\dagger(r)X^\dagger(r)'dr\right)^{-1} \\ &\quad \times \int_0^1 X^\dagger(r)B_u(r)dr, \end{aligned} \tag{1.26}$$

so that $\hat{\mu}^\dagger = O_p(\sqrt{n})$ diverges as $n \to \infty$. The divergence of the fitted intercept is a consequence of the fact that the temporally aggregated model error U_t is an $I(1)$ time series whose signal exceeds that of the constant regressor associated with the intercept. The fitted equation therefore manifests a partially spurious regression property in which the intercept is not consistently estimated but the time polynomial regression coefficients are consistent because their signal is stronger, as discussed earlier.

1.4 Wald Statistic Inference

Before commencing a general treatment we illustrate the issues involved with some examples. These help in developing a notation that is suited for a complete discussion of the general case. In these examples and the subsequent treatment we continue to work with the temporally aggregated model in which $C(1) = 0$ and the model has a fitted intercept, as in (1.16). The analysis for the case where $C(1) \neq 0$ is entirely analogous, allowing for the different convergence rates of the estimated coefficients and different estimate of the residual long-run error variance.[1]

[1] In practical situations it is typically unknown, at least without testing, whether $C(1) = 0$ or $C(1) \neq 0$. Then an additional issue arises in the construction of the Wald test statistic, because the statistic depends on a consistent estimate of the variance matrix of the coefficient estimates, which in turn depends on whether $C(1) = 0$ or $C(1) \neq 0$. Readers are referred to Phillips and Kheifets (2024) for a systematic analysis of this issue in the case of a possibly multicointegrated system. That model involves stochastic trends but the ideas underlying the construction of the Wald statistic may be applied in the present context. A full analysis of that case is not presented here to avoid unnecessary lengthening of the chapter.

1.4.1 Example 1

We start with the case of a simple null hypothesis about all the coefficients of the form $\mathcal{H}_0: \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_m$. This hypothesis can be written as $\mathcal{H}_0: H\beta = h$ using the $m \times (m+1)$ matrix $H = [\iota_m, -I_m]$ and vector $h = 0$ with the m-vector $\iota_m = (1, 1, \ldots, 1)'$ and order m identity matrix I_m. The Wald statistic for testing $\mathcal{H}_0$ is then

$$\begin{aligned} W_n &= (H\hat{\beta}^\dagger - h)' \left[H \left(\tilde{X}^{\dagger\prime} \tilde{X}^\dagger \right)^{-1} H' \right]^{-1} (H\hat{\beta}^\dagger - h)/\hat{\tilde{\omega}}^2 \\ &= \hat{\beta}^{\dagger\prime} H' \left[H \left(\tilde{X}^{\dagger\prime} \tilde{X}^\dagger \right)^{-1} H' \right]^{-1} H\hat{\beta}^\dagger/\hat{\tilde{\omega}}^2. \end{aligned} \tag{1.27}$$

The analysis that follows proceeds with the case studied in Section 1.3.4 where $C(1) = 0$ and an intercept is included in the regression as in (1.16). Using the normalization matrix $F_n = nD_n = \mathrm{diag}(n, n^2, \ldots, n^{m+1})$, and setting $G_n = \mathrm{diag}(n^2, \ldots, n^{m+1})$, we have $G_n^{-1} = O(n^{-2})$ and observe that $HF_n^{-1} = [\iota_m, -I_m]F_n^{-1} = [n^{-1}\iota_m, -G_n^{-1}] = \frac{1}{n}[\iota_m, -nG_n^{-1}]$. Then

$$\begin{aligned} n^{3/2} HF_n^{-1} F_n(\hat{\beta}^\dagger - \beta) &= \left[\iota_m, -nG_n^{-1} \right] \sqrt{n} F_n(\hat{\beta}^\dagger - \beta) \\ &\sim_a \left[\iota_m, -nG_n^{-1} \right] \left(\int_0^1 X^\dagger(r) X^\dagger(r)' dr \right)^{-1} \int_0^1 X^\dagger(r) dB_{\tilde{\varepsilon}}(dr) \\ &\sim_a \mathcal{N}\left(0, \tilde{\omega}^2 \left[\iota_m, -nG_n^{-1} \right] M^{\dagger -1} \left[\iota_m, -nG_n^{-1} \right]' \right) \\ &\rightsquigarrow \mathcal{N}\left(0, \tilde{\omega}^2 \left[\iota_m, 0_{m\times m} \right] M^{\dagger -1} \left[\iota_m, 0_{m\times m} \right]' \right), \end{aligned} \tag{1.28}$$

as $n \to \infty$ because $nG_n^{-1} = O(n^{-1})$ and so $\left[\iota_m, -nG_n^{-1} \right] \to [\iota_m, 0_{m\times m}]$. Observe that the limit distribution given in (1.28) is singular because the transform matrix $[\iota_m, 0_{m\times m}]$ has rank $1 < m$. This rank deficiency affects the limit distribution of the Wald statistic and the degrees of freedom of the resulting chi-squared statistic. In particular, we have

$$\begin{aligned} &\hat{\beta}^{\dagger\prime} H' \left[H \left(\tilde{X}^{\dagger\prime} \tilde{X}^\dagger \right)^{-1} H' \right]^{-1} H\hat{\beta}^\dagger/\hat{\tilde{\omega}}^2 \\ &= \hat{\beta}^{\dagger\prime} H' \left[\frac{1}{n} HF_n^{-1} \left(\frac{1}{n} F_n^{-1} \tilde{X}^{\dagger\prime} \tilde{X}^\dagger F_n^{-1} \right)^{-1} F_n^{-1} H' \right]^{-1} H\hat{\beta}^\dagger/\hat{\tilde{\omega}}^2 \\ &= \frac{1}{n} \hat{\beta}^{\dagger\prime} \sqrt{n} F_n F_n^{-1} H' \left[\frac{1}{n} HF_n^{-1} \left(\frac{1}{n} F_n^{-1} \tilde{X}^{\dagger\prime} \tilde{X}^\dagger F_n^{-1} \right)^{-1} F_n^{-1} H' \right]^{-1} \\ &\quad \times HF_n^{-1} \sqrt{n} F_n \hat{\beta}^\dagger/\hat{\tilde{\omega}}^2 \end{aligned}$$

$$
\begin{aligned}
&\sim_a \hat{\beta}^{\dagger\prime}\sqrt{n}F_nF_n^{-1}H'\left[\left[\frac{1}{n}\iota_m, -G_n^{-1}\right]M^{\dagger-1}\left[\frac{1}{n}\iota_m, -G_n^{-1}\right]'\right]^{-1} \\
&\quad \times HF_n^{-1}\sqrt{n}F_n\hat{\beta}^{\dagger}/\tilde{\omega}^2 \\
&\sim_a \int_0^1 dB_{\tilde{\varepsilon}}(r)X^{\dagger\prime}(r)M^{\dagger-1}\left[\frac{1}{n}\iota_m, -G_n^{-1}\right]' \\
&\quad \times \left[\left[\frac{1}{n}\iota_m, -G_n^{-1}\right]M^{\dagger-1}\left[\frac{1}{n}\iota_m, -G_n^{-1}\right]'\right]^{-1} \\
&\times \left[\frac{1}{n}\iota_m, -G_n^{-1}\right]M^{\dagger-1}\int_0^1 X^{\dagger}(r)dB_{\tilde{\varepsilon}}(r)/\tilde{\omega}^2 \\
&\sim_a \int_0^1 dB_{\tilde{\varepsilon}}(r)X^{\dagger}(r)'M^{\dagger-1}\left[\iota_m, -nG_n^{-1}\right]' \\
&\times \left[\left[\iota_m, -nG_n^{-1}\right]M^{\dagger-1}\left[\iota_m, -nG_n^{-1}\right]'\right]^{-1} \\
&\times \left[\iota_m, -nG_n^{-1}\right]M^{\dagger-1}\int_0^1 X^{\dagger}(r)dB_{\tilde{\varepsilon}}(r)/\tilde{\omega}^2 \\
&= \int_0^1 dW_{\tilde{\varepsilon}}(r)X^{\dagger}(r)'M^{\dagger-1/2}P_{\mathcal{A}_n}M^{\dagger-1/2}\int_0^1 X^{\dagger}(r)dW_{\tilde{\varepsilon}}(r) \\
&= \mathcal{Z}'P_{\mathcal{A}_n}\mathcal{Z} \rightsquigarrow \mathcal{Z}'P_{\mathcal{A}}\mathcal{Z} =_d \chi_1^2. \qquad (1.29)
\end{aligned}
$$

Here $W_\varepsilon := (1/\tilde{\omega})B_\varepsilon$ is standard Brownian motion and $\mathcal{Z}'P_{\mathcal{A}_n}\mathcal{Z}$ is a quadratic form in the Gaussian vector $\mathcal{Z} := \left(\int_0^1 X^{\dagger}(r)X^{\dagger}(r)'dr\right)^{-1/2}\int_0^1 X^{\dagger}(r)dW_{\tilde{\varepsilon}}(r) =_d \mathcal{N}(0, I_m)$. The projection matrix $P_{\mathcal{A}_n} = \mathcal{A}_n\left(\mathcal{A}_n'\mathcal{A}_n\right)^{-1}\mathcal{A}_n'$ is deterministic of rank $m = \text{rank}\left(\left[\iota_m, -nG_n^{-1}\right]\right)$ for finite n. $P_{\mathcal{A}_n}$ projects onto the m-dimensional range space of the matrix $\mathcal{A}_n$ where

$$
\begin{aligned}
\mathcal{A}_n' &= \left[\iota_m, -nG_n^{-1}\right]M^{\dagger-1/2} = \left[\iota_m, O\left(\frac{1}{n}\right)\right]M^{\dagger-1/2} \\
&\quad \to [\iota_m, 0_{m\times m}]\,M^{\dagger-1/2} =: \mathcal{A}'. \qquad (1.30)
\end{aligned}
$$

The $(m+1)\times(m+1)$ dimensional limiting projection matrix $P_{\mathcal{A}} = \mathcal{A}\left(\mathcal{A}'\mathcal{A}\right)^{+}\mathcal{A}'$ has rank unity, where $\left(\mathcal{A}'\mathcal{A}\right)^{+}$ is the Moore–Penrose inverse[2] of $\mathcal{A}'\mathcal{A}$. The projector $P_{\mathcal{A}}$ can be simplified as follows. Let a' be the first row of the symmetric matrix $M^{\dagger-1/2} = \left(\int_0^1 X^{\dagger}(r)X^{\dagger}(r)'dr\right)^{-1/2}$. Then

$$
\mathcal{A}' = [\iota_m, 0]\,M^{\dagger-1/2} = \iota_m a', \qquad (1.31)
$$

[2] For an $n\times k$ matrix X of rank $k < n$ the Moore–Penrose inverse of XX' is $(XX')^{+} = X(X'X)^{-2}X'$ and the Moore–Penrose inverse of X is $X^{+} = X'(XX')^{+}$.

and

$$P_{\mathcal{A}} = a\iota_m'(\iota_m a' a \iota_m')^+ \iota_m a' = a(a'a)^{-1}a'\iota_m'(\iota_m\iota_m')^+\iota_m$$
$$= P_a \iota_m'(\iota_m\iota_m')^+\iota_m = P_a, \tag{1.32}$$

since the Moore–Penrose inverse $(\iota_m\iota_m')^+ = \iota_m(\iota_m'\iota_m)^{-2}\iota_m$ and thus

$$\iota_m'(\iota_m\iota_m')^+\iota_m = \iota_m'\iota_m(\iota_m'\iota_m)^{-2}\iota_m'\iota_m = 1.$$

It follows that $\mathcal{Z}'P_{\mathcal{A}}\mathcal{Z} = \mathcal{Z}'P_a\mathcal{Z} =_d \chi_1^2$, giving (1.29).

This example illustrates a key asymptotic deficiency in the Wald test in which a test of m different hypotheses leads to a statistic whose limit distribution is chi-square with only a single degree of freedom. The dimension reduction that occurs in this limit theory of the Wald statistic is associated with the concentration of the range space of the projector $P_{\mathcal{A}_n}$ to the lower dimensional range space of the projector $P_{\mathcal{A}} = P_a$. The dimension reduction occurs in the limit distribution of the Wald statistic as the sample size $n \to \infty$ and thereby affects inference but it does not occur in the limit distribution of the estimated coefficient vector $\hat{\beta}^{\dagger}$ itself, as is clear from (1.20). More specifically, it is the matrix metric $H\left(\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}\right)^{-1}H'$, which is used to measure distance from the null hypothesis in the Wald statistic (1.27), that collapses in dimension as $n \to \infty$. This collapse in dimension at infinity is the result of the differing signal strengths in the signal matrix $\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}$ that determine the convergence rates of the elements of the estimated coefficient vector $\hat{\beta}$. It is captured by the replacement of this metric, upon appropriate matrix normalization, by the projector $P_{\mathcal{A}_n}$ which collapses to $P_{\mathcal{A}} = P_a$ as $n \to \infty$.

In the present case the null hypothesis is $\mathcal{H}_0 : \beta_0 = \beta_1 = \ldots, \beta_m$ and involves m restrictions on $m+1$ coefficients. But the limit distribution of the Wald test is simply χ_1^2 rather than the usual χ_m^2 where the number of degrees of freedom equals the number of restrictions. The reason for the reduced degrees of freedom in the limit distribution is that the estimates of the individual coefficients $\{\beta_1, \ldots, \beta_m\}$ of the trend terms all have faster rates of convergence given the stronger signal associated with their regressors. This difference means that the variances of the differentials in the regression coefficients are all dominated by the variance of $\hat{\beta}_0$. For instance, in the regression (1.16) for the temporally aggregated equation, the difference in the first two coefficients $\hat{\beta}_0 - \hat{\beta}_1$ satisfies

$$n^{3/2}(\hat{\beta}_0 - \hat{\beta}_1) = n^{3/2}(\hat{\beta}_0\beta_1) - n^{3/2}(\hat{\beta}_1 - \beta_1) = n^{3/2}(\hat{\beta}_0 - \beta_1) + O_p\left(\frac{1}{n}\right), \tag{1.33}$$

since $n^{5/2}(\hat{\beta}_1 - \hat{\beta}_1) = O_p(1)$. Recall that the regressor $X_t = \sum_{s=1}^t x_s = \sum_{s=1}^t (1, s, s^2, \ldots, s^m)'$ in (1.16). So the components of X_t corresponding to the coefficients β_0 and β_1 are $X_{1t} = t$ and $X_{2t} = \sum_{s=1}^{t} s = t(t+1)/2$. The convergence rate of $\hat{\beta}_0$ is then $n^{3/2}$, corresponding to the order of magnitude of the signal $\sum_{t=1}^{n} X_{1t}^2 = O(n^3)$ and the convergence rate of $\hat{\beta}_1$ is $n^{5/2}$ corresponding to the signal order $\sum_{t=1}^{n} X_{2t}^2 = O(n^5)$. Hence, the limit distribution in the Wald test is determined as if the coefficients of the higher-order trend terms were known. In effect, the greater rates of convergence of the higher-order coefficient estimates in the trend regression imply that their variances do not contribute to the limit distribution of the Wald statistic and only the variance of $\hat{\beta}_0$ figures in the limit theory. But $\hat{\beta}_0$ has only a single dimension and this is what leads to the χ_1^2 limit distribution in (1.29).

Remarks

(i) The reduction in dimension that occurs in the limit theory of the Wald statistic W_n in (1.29) from the rank m of H to unity matches the reduction in dimension of the limit theory of the transformed estimator (1.28) following matrix normalization. Importantly, in the derivation of this limit theory there is no reduction in dimension for finite n. The collapse in dimension occurs in the limit when $n \to \infty$. Notably, the singular covariance matrix $\tilde{\omega}^2 \left[\iota_m, 0_{m\times m}\right] \left(\int_0^1 X^\dagger(r) X^\dagger(r)' dr\right)^{-1} \left[\iota_m, 0_{m\times m}\right]'$ of the limit distribution in (1.28) of the normalized, transformed, and centered estimator $n^{3/2} H F_n^{-1} F_n(\hat{\beta}^\dagger - \beta)$ is consistently estimated by the matrix $\hat{\tilde{\omega}}^2 \left[\iota_m, -nG_n^{-1}\right] (\frac{1}{n} F_n^{-1} \tilde{X}^{\dagger\prime} \tilde{X}^\dagger F_n^{-1})^{-1} \left[\iota_m, -nG_n^{-1}\right]'$. This means that, although a generalized inverse is involved in our analysis of the limit distribution theory of the Wald statistic, it is not necessary to consider limits of generalized inverses as $n \to \infty$ in determining the limit theory of the Wald statistic as in the analyses of Stewart (1969), Puri et al. (1984), and Andrews (1987). The situation in this example where consistent estimation of the limiting covariance matrix is available is instead similar to the case studied in Vuong (1987).

(ii) As is apparent in the derivation leading to (1.29), the details of the dimensional reduction rely on the form of the null hypothesis $\mathcal{H}_0$ and in particular the form of the matrix H. The asymptotic theory can be complex and the final limit distribution of the Wald test is contingent on the interaction between the normalization matrix F_n and the hypothesis matrix H. This is the essence of the difficulties involved in matrix normalization.

1.4.2 Example 2

The next example gives a different scenario in the limit theory which displays no asymptotic deficiency but more complex matrix normalization. The null hypothesis is $\mathcal{H}_0 : H\beta = h$ with $h = 0$ and

$$H = H_{2\times m+1} = \begin{bmatrix} 1 & -2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix} =: \left[H_1, H_2, 0_{2\times(m-3)}\right], \tag{1.34}$$

with

$$H_1 = \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix},\ H_2 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, \tag{1.35}$$

so that the null hypothesis is simply $\mathcal{H}_0 \colon \beta_0 = 2\beta_1$ and $\beta_2 + \beta_3 = 0$.

We now partition the normalization matrix as $F_n = nD_n = \operatorname{diag}(n, n^2, \ldots, n^{m+1}) = \operatorname{blockdiag}[F_{1n}, F_{2n}, F_{3n}]$, where $F_{1n} = \operatorname{diag}(n, n^2)$, $F_{2n} = \operatorname{diag}(n^3, n^4)$, and $F_{3n} = \operatorname{diag}(n^5, \ldots, n^{m+1})$, so that $F_{3n}^{-1} = O(n^{-5})$. Next observe that

$$\begin{aligned} HF_n^{-1} &= \left[H_1, H_2, 0_{2\times(m-3)}\right] F_n^{-1} = \left[H_1 F_{1n}^{-1}, H_2 F_{2n}^{-1}, 0_{2\times(m-3)}\right] \\ &= \begin{bmatrix} \frac{1}{n} & -\frac{2}{n^2} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{n^3} & \frac{1}{n^4} & 0 & \cdots & 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{n^3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{2}{n} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \frac{1}{n} & 0 & \cdots & 0 \end{bmatrix} =: Q_n \times R_n, \end{aligned} \tag{1.36}$$

with $Q_n = \operatorname{diag}(n^{-1}, n^{-3})$ and $R_n = \begin{bmatrix} 1 & -\frac{2}{n} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \frac{1}{n} & 0 & \cdots & 0 \end{bmatrix}$ has dimension $2 \times (m+1)$ and rank 2. Then

$$\begin{aligned} &\sqrt{n} Q_n^{-1} H F_n^{-1} F_n(\hat{\beta}^\dagger - \beta) = R_n \sqrt{n} F_n (\hat{\beta}^\dagger - \beta) \\ &\sim_a R_n M^{\dagger -1} \int_0^1 X^\dagger(r) dB_{\tilde{\varepsilon}}(dr) \\ &=_d \mathcal{N}\left(0,\ \tilde{\omega}^2 R_n M^{\dagger-1} R_n'\right) \rightsquigarrow \mathcal{N}\left(0,\ \tilde{\omega}^2 R M^{\dagger -1} R'\right), \end{aligned} \tag{1.37}$$

where $R_n \to R = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & \cdots & 0 \end{bmatrix}$. Observe that the limit distribution of $\hat{\beta}^\dagger - \beta$ given in (1.37) is of full rank 2 but requires a more complex matrix normalization involving the normalization matrix Q_n than does that of the earlier result given in (1.28). With this limit theory for the estimator in hand we can obtain the limit distribution of the Wald statistic for testing $\mathcal{H}_0 \colon H\beta = h = 0$ as follows.

$$\begin{aligned} W_n &= \hat{\beta}^{\dagger\prime} H' \left[H \left(\tilde{X}^{\dagger\prime} \tilde{X}^\dagger\right)^{-1} H'\right]^{-1} H\hat{\beta}^\dagger / \hat{\tilde{\omega}}^2 \\ &= \hat{\beta}^{\dagger\prime} H' \left[\frac{1}{n} H F_n^{-1} \left(\frac{1}{n} F_n^{-1} \tilde{X}^{\dagger\prime} \tilde{X}^\dagger F_n^{-1}\right)^{-1} F_n^{-1} H'\right]^{-1} H\hat{\beta}^\dagger / \hat{\tilde{\omega}}^2 \end{aligned}$$

$$= \frac{1}{n}\hat{\beta}^{\dagger\prime}\sqrt{n}F_nR_n'Q_n'\left[\frac{1}{n}Q_nR_n\left(\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}F_n^{-1}\right)^{-1}R_n'Q_n'\right]^{-1}$$

$$\times\, Q_nR_n\sqrt{n}F_n\hat{\beta}^{\dagger}/\hat{\hat{\omega}}^2$$

$$= \hat{\beta}^{\dagger\prime}\sqrt{n}F_nR_n'\left[R_n\left(\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}F_n^{-1}\right)^{-1}R_n'\right]^{-1}$$

$$R_n\sqrt{n}F_n\hat{\beta}^{\dagger}/\hat{\hat{\omega}}^2 \rightsquigarrow \chi_2^2, \tag{1.38}$$

in view of (1.37) and the fact that $\mathrm{rank}(R) = 2$. Hence, in this case the distribution has the usual limiting chi-squared distribution with degrees of freedom that match the number of (independent) restrictions involved in the null hypothesis $\mathcal{H}_0\colon H\beta = h$ with the restriction matrix H given in (1.34).

1.4.3 Example 3

We now combine the features of Examples 1 and 2 in a third more complex case involving the null hypothesis $\mathcal{H}_0\colon H\beta = h$ with $h = 0$ and the $m \times (m+1)$ restriction matrix

$$H = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 \\ 2 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 3 & 0 & -2 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & 0 & 0 & \cdots & -1 \end{bmatrix}, \tag{1.39}$$

giving $q = m$ restrictions on β. So the null hypothesis is $\mathcal{H}_0\colon \beta_0 - \beta_1 = 0,\ 2\beta_0 + \beta_2 = 0,\ 3\beta_1 + 2\beta_3 = 0\ \beta_3 = \beta_4 = \cdots = \beta_m$ or equivalently $\bar{\mathcal{H}}_0\colon \beta_0 - \beta_1 = 0,\ \beta_0 + \frac{1}{2}\beta_2 = 0,\ \beta_1 + \frac{2}{3}\beta_3 = 0,\ \beta_3 = \beta_4 = \cdots = \beta_m$. For this equivalent null hypothesis $\bar{\mathcal{H}}_0\colon\ \bar{H}\beta = 0$ we have the normalized $m \times (m+1)$ restriction matrix

$$\bar{H} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & 1 & 0 & -\frac{2}{3} & \cdots & 0 \\ 0_{m-3} & 0_{m-3} & \iota_{m-3} & & -I_{m-3} & \end{bmatrix}$$

$$= \left[\begin{array}{cccccc} \iota_2 & \begin{matrix} -1 \\ 0 \end{matrix} & \begin{matrix} 0 \\ \frac{1}{2} \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} \cdots \\ \cdots \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \hline 0 & 1 & 0 & -\frac{2}{3} & \cdots & 0 \\ 0_{m-3} & 0_{m-3} & \iota_{m-3} & & -I_{m-3} & \end{array}\right]. \tag{1.40}$$

Setting $G_n = \mathrm{diag}(n^4, n^5, \cdots, n^{m+1})$ we now have

$$
\begin{aligned}
\bar{H}F_n^{-1} &= \left[\begin{array}{cccccc}
\multirow{2}{*}{$\frac{1}{n}\iota_2$} & -\frac{1}{n^2} & 0 & 0 & \cdots & 0 \\
 & 0 & \frac{1}{2n^3} & 0 & \cdots & 0 \\
\hline
0 & \frac{1}{n^2} & 0 & -\frac{2}{3n^4} & \cdots & 0 \\
0_{m-3} & 0_{m-3} & \frac{1}{n^3}\iota_{m-3} & & -G_n^{-1} &
\end{array}\right] \\
&= \left[\begin{array}{ccccc}
\frac{1}{n}I_2 & 0 & 0 & 0\cdots & 0 \\
0 & \frac{1}{n^2} & 0 & 0\cdots & 0 \\
0 & 0 & & \frac{1}{n^3}I_{m-3} &
\end{array}\right] \\
&\quad \times \left[\begin{array}{cccccc}
\multirow{2}{*}{ι_2} & -\frac{1}{n} & 0 & 0 & \cdots & 0 \\
 & 0 & \frac{1}{2n^2} & 0 & \cdots & 0 \\
\hline
0 & 1 & 0 & -\frac{2}{3n^2} & \cdots & 0 \\
0_{m-3} & 0_{m-3} & \iota_{m-3} & & -n^3 G_n^{-1} &
\end{array}\right] \\
&=: Q_n \times R_n,
\end{aligned} \tag{1.41}
$$

where Q_n is an $m \times m$ normalization matrix and R_n is an $m \times m+1$ matrix of rank m for finite n which captures the implied normalized form of the restriction matrix. Similar to (1.37) in Example 2 we can now obtain the limit theory for the appropriately normalized estimated error in the restrictions $\bar{H}(\hat{\beta}^\dagger - \beta)$, namely,

$$
\begin{aligned}
&\sqrt{n}Q_n^{-1}\,\bar{H}F_n^{-1}F_n(\hat{\beta}^\dagger - \beta) = R_n\sqrt{n}F_n(\hat{\beta}^\dagger - \beta) \\
&\sim_a R_n M^{\dagger-1}\int_0^1 X^\dagger(r)dB_{\tilde{\varepsilon}}(dr) \rightsquigarrow \mathcal{N}\left(0,\ \tilde{\omega}^2 R M^{\dagger-1} R'\right),
\end{aligned} \tag{1.42}
$$

since

$$
R_n \to R = \left[\begin{array}{ccccc}
\multirow{2}{*}{ι_2} & 0 & 0 & \cdots & 0 \\
 & 0 & 0 & \cdots & 0 \\
\hline
0 & 1 & 0 & \cdots & 0 \\
0_{m-3} & 0_{m-3} & \iota_{m-3} & & 0_{m-3\times m-2}
\end{array}\right]_{m\times(m+1)}. \tag{1.43}
$$

Observe that the limit matrix R has rank $3 < m$ and, correspondingly, the limit distribution in (1.42) is singular with deficient rank 3. As in Example 1, this singularity affects the limit distribution of the Wald statistic and the degrees of freedom of the resulting chi-squared statistic. The derivation follows the same lines but is more complex than that leading to (1.29) in Example 1 because there are now two sources of degeneracy, not one, and these are of different dimension. The Wald statistic has the following form (replacing H with $\bar{H}$):

$$
\begin{aligned}
&\hat{\beta}^{\dagger\prime}\,\bar{H}'\left[\bar{H}\left(\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}\right)^{-1}\bar{H}'\right]^{-1}\bar{H}\hat{\beta}^{\dagger}/\hat{\tilde{\omega}}^2\\
&=\hat{\beta}^{\dagger\prime}\,\bar{H}'\left[\frac{1}{n}\bar{H}F_n^{-1}\left(\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}F_n^{-1}\right)^{-1}F_n^{-1}\,\bar{H}'\right]^{-1}\bar{H}\hat{\beta}^{\dagger}/\hat{\tilde{\omega}}^2\\
&=\frac{1}{n}\hat{\beta}^{\dagger\prime}\sqrt{n}F_nF_n^{-1}\,\bar{H}'Q_n^{-1}\left[\frac{1}{n}Q_n^{-1}\,\bar{H}F_n^{-1}\left(\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}F_n^{-1}\right)^{-1}\right.\\
&\qquad\left.\times F_n^{-1}\,\bar{H}'Q_n^{-1}\right]^{-1}Q_n^{-1}\,\bar{H}F_n^{-1}\sqrt{n}F_n\hat{\beta}^{\dagger}/\hat{\tilde{\omega}}^2\\
&\sim_a\hat{\beta}^{\dagger\prime}\sqrt{n}F_nR_n'\left[R_nM^{\dagger-1}R_n'\right]^{-1}R_n\sqrt{n}F_n\hat{\beta}^{\dagger}/\tilde{\omega}^2\\
&\sim_a\int_0^1 dB_{\tilde{\varepsilon}}(r)X^{\dagger}(r)'M^{\dagger-1}R_n'\left[R_nM^{\dagger-1}R_n'\right]^{-1}R_nM^{\dagger-1}\\
&\qquad\times\int_0^1 X^{\dagger}(r)dB_{\tilde{\varepsilon}}(r)/\tilde{\omega}^2\\
&=\int_0^1 dW_{\tilde{\varepsilon}}(r)X^{\dagger}(r)'M^{\dagger-1/2}P_{\mathcal{A}_n}M^{\dagger-1/2}\times\int_0^1 X^{\dagger}(r)dW_{\tilde{\varepsilon}}(r)\\
&=\mathcal{Z}'P_{\mathcal{A}_n}\mathcal{Z}\rightsquigarrow\mathcal{Z}'P_{\mathcal{A}}\mathcal{Z}=_d\chi_3^2.
\end{aligned}\tag{1.44}
$$

As in (1.29) $W_\varepsilon := (1/\tilde{\omega})B_\varepsilon$ is standard Brownian motion and $\mathcal{Z}'P_{\mathcal{A}_n}\mathcal{Z}$ is a quadratic form in the Gaussian vector $\mathcal{Z} := M^{\dagger-1/2}\int_0^1 X^{\dagger}(r)dW_{\tilde{\varepsilon}}(r) =_d \mathcal{N}(0, I_m)$. The projection matrix $P_{\mathcal{A}_n} = \mathcal{A}_n\left(\mathcal{A}_n'\mathcal{A}_n\right)^{-1}\mathcal{A}_n'$ is deterministic of rank $m = \text{rank}(\mathcal{A}_n) = \text{rank}(R_n)$ for all finite n. $P_{\mathcal{A}_n}$ projects onto the m-dimensional range space of the $(m+1)\times m$ matrix $\mathcal{A}_n$, where in (1.44) we have

$$
\mathcal{A}_n' = R_nM^{\dagger-1/2} \to RM^{\dagger-1/2} =: \mathcal{A}',\tag{1.45}
$$

for which the limiting rank is $\text{rank}(\mathcal{A}) = 3$, the same as the matrix R. The $(m+1)$-dimensional limiting projection matrix $P_{\mathcal{A}} = \mathcal{A}\left(\mathcal{A}'\mathcal{A}\right)^{+}\mathcal{A}'$ therefore has rank 3. An explicit form of the projector $P_{\mathcal{A}}$ can be expressed as follows. Let a_1', a_2' and a_3' be the first, second and third rows of the symmetric matrix positive definite matrix $M^{\dagger-1/2} = \left(\int_0^1 X^{\dagger}(r)X^{\dagger}(r)'dr\right)^{-1/2}$. Then the $(m+1)\times m$ matrix $\mathcal{A}$ has the explicit form

$$
\mathcal{A} = M^{\dagger-1/2}R' = \begin{bmatrix} a_1\iota_2' & a_2 & a_3\iota_{m-3}' \end{bmatrix},\tag{1.46}
$$

in which there are only three linearly independent columns, which we collect in the $(m+1)\times 3$ matrix $A = [a_1, a_2, a_3]$. It follows directly that the projector $P_{\mathcal{A}}$ projects onto the range space of A and is therefore equivalent to the projector

P_A by virtue of uniqueness. We deduce that the limit distribution of the Wald statistic is $\mathcal{Z}'P_{\mathcal{A}}\mathcal{Z} = \mathcal{Z}'P_A\mathcal{Z} =_d \chi_3^2$, as given in (1.44).

1.4.4 The General Case

From these examples, we now proceed to a general treatment of the limit theory for Wald tests in trend polynomial regression. The analysis continues to use the time-aggregated regression model (1.16), again with $C(1) = 0$. We start by formulating a generalized version of (1.40) for the normalized restriction matrix $\bar{H}$ in which: (i) the rows of $\bar{H}$ are assembled so that the corresponding hypotheses are ordered according to the time polynomial power of the respective coefficients; and (ii) the first nonzero coefficient in each row, corresponding to the lowest power of the associated time polynomial, is set to unity, as in (1.40). With this ordering and normalization and recognizing that hypotheses can involve different elements of the coefficient vector β that relate to time polynomials of different degrees, we partition the $q \times (m+1)$ restriction matrix $\bar{H}$ of the null hypothesis to accord with these elements in rising polynomial degree form as follows:

$$\bar{H} = \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} & \cdots & h_{1,m+1} \\ 0 & h_{2,2} & h_{2,3} & \cdots & h_{2,m+1} \\ 0 & 0 & h_{3,3} & \cdots & h_{3,m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & h_{p,m+1} \end{bmatrix}_S =: \left[\bar{H}^{(1)}, \bar{H}^{(2)}, \bar{H}^{3}, \cdots, \bar{H}^{(m+1)} \right]. \quad (1.47)$$

In this representation there are p block rows indicated and $m+1$ columns. Each block diagonal entry of $\bar{H}$ is a vector of the form $h_{i,i} = c_i \iota_{q_i}$ in which $\iota_{q_i} = (1, 1, \ldots, 1)'$ has dimension q_i and $c_i = \mathbb{j}_{S_i}$ is the selector function of the block row i, which indicates whether that block row enters the matrix or not. In particular, block row i occurs in the matrix if the entry is $c_i = \mathbb{j}_{S_i} = 1$, which signifies that there are restrictions involving the coefficient β_{i-1} of t^{i-1}, namely, the lowest trend degree coefficient in block row i. On the other hand, if the entry $c_i = \mathbb{j}_{S_i} = 0$, then row i is removed from the matrix altogether as the selector $\mathbb{j}_{S_i} = 0$ signifies that there is no restriction involving β_{i-1} – so the ith-block row is removed entirely in the specification $\bar{H}_S$. The affix S on $\bar{H}_S$ in the upper block triangular matrix (1.47) means that selection may have taken place in the matrix and then $r = \sum_{i=1}^{p} \mathbb{1}\{c_i \neq 0\}$ is the actual number of block rows in the matrix $\bar{H}_S$. This arrangement implies that the q restrictions that comprise the rows of $\bar{H}$ are partitioned into r groups whose numbers sum to $q = \sum_{i=1}^{p} q_i \mathbb{1}\{c_i \neq 0\}$ in total. Upon selection of its relevant rows the restriction matrix $\bar{H} = \bar{H}_S$ is then of dimension $q \times (m+1)$ with full rank q.

In the first example above where $\bar{H} = \bar{H}_S = [\iota_m, -I_m]$ there is only one block row in the restriction matrix $\bar{H}$ and so $r = 1$, $q_1 = q = m$, $c_1 = 1$, and $c_j = 0$ for all $j > 1$. In the second example presented earlier

$$\bar{H} = \bar{H}_S = \begin{bmatrix} 1 & -2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix},$$

and there are only two rows in the restriction matrix, so that $r = q = 2$, $q_1 = q_3 = 1$, $c_1 = c_3 = 1$, and $c_j = 0$ for all $j \notin \{1, 3\}$. In the third example presented earlier where

$$\bar{H} = \left[\begin{array}{cccccc} \iota_2 & \begin{matrix} -1 \\ 0 \end{matrix} & \begin{matrix} 0 \\ \frac{1}{2} \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} \cdots \\ \cdots \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \hline 0 & 1 & 0 & -\frac{2}{3} & \cdots & 0 \\ 0_{m-3} & 0_{m-3} & \iota_{m-3} & & -I_{m-3} & \end{array}\right], \tag{1.48}$$

there are three blocks of restrictions so that $r = 3, q_1 = 2, q_2 = 1, q_3 = m - 3$, $q = m$, and $c_1 = c_2 = c_3 = 1$.

In the general case with restriction matrix (1.47) the normalization matrix is $F_n = \text{diag}(n, n^2, \ldots, n^{m+1})$ and we write the effect of normalization by F_n^{-1} on the $q \times (m + 1)$ restriction matrix $\bar{H}$ in the following partitioned matrix form which uses the selection matrix notation

$$\begin{aligned}
\bar{H}F_n^{-1} &= \left[\bar{H}^{(1)}, \bar{H}^{(2)}, \bar{H}^{3}, \cdots, \bar{H}^{(m+1)} \right] F_n^{-1} \\
&= \left[\frac{1}{n}\bar{H}^{(1)}, \frac{1}{n^2}\bar{H}^{(2)}, \frac{1}{n^3}\bar{H}^{(3)}, \cdots, \frac{1}{n^{m+1}}\bar{H}^{(m+1)} \right] \\
&= \begin{bmatrix} \frac{1}{n}h_{1,1} & \frac{1}{n^2}h_{1,2} & \frac{1}{n^3}h_{1,3} & \cdots & \frac{1}{n^{m+1}}h_{1,m+1} \\ 0 & \frac{1}{n^2}h_{2,2} & \frac{1}{n^3}h_{2,3} & \cdots & \frac{1}{n^{m+1}}h_{2,m+1} \\ 0 & 0 & \frac{1}{n^3}h_{3,3} & \cdots & \frac{1}{n^{m+1}}h_{3,m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{n^{m+1}}h_{p,m+1} \end{bmatrix}_S \\
&= \begin{bmatrix} \frac{c_1}{n}I_{q_1} & 0 & 0 & \cdots & 0 \\ 0 & \frac{c_2}{n^2}I_{q_2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{c_3}{n^3}I_{q_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{c_p}{n^p}I_{q_p} \end{bmatrix}_S
\end{aligned}$$

$$\times \begin{bmatrix} c_1 \iota_{q_1} & \frac{1}{n} h_{1,2} & \frac{1}{n^2} h_{1,3} & \cdots & \frac{1}{n^{r-1}} h_{1,p} & \cdots & \frac{1}{n^m} h_{1,m+1} \\ 0 & c_2 \iota_{q_2} & \frac{1}{n} h_{2,3} & \cdots & \frac{1}{n^{r-2}} h_{2,p} & \cdots & \frac{1}{n^{m-1}} h_{2,m+1} \\ 0 & 0 & c_3 \iota_{q_3} & \cdots & \frac{1}{n^{r-3}} h_{3,p} & \cdots & \frac{1}{n^{m-2}} h_{3,m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_p \iota_{q_p} & \cdots & \frac{1}{n^{m-r+1}} h_{p,m+1} \end{bmatrix}_S$$

$$=: Q_{n,S^+} \times R_{n,S}. \tag{1.49}$$

In the matrix Q_{n,S^+} the affix S^+ signifies selection and retention of rows *and* columns according to the rule that $c_i = 1$ and the removal of those rows *and* columns for which $c_i = 0$. Thus, Q_{n,S^+} is a diagonal $q \times q$ matrix that is conformable with the $q \times (m+1)$ matrix $R_{n,S}$ defined by (1.49).

With this framework in the general case we can now obtain the limit theory as $n \to \infty$ for the appropriately normalized estimated error in the restrictions $\bar{H}(\hat{\beta}^\dagger - \beta)$, namely,

$$\begin{aligned} &\sqrt{n} Q_{n,S^+}^{-1} \bar{H} F_n^{-1} F_n (\hat{\beta}^\dagger - \beta) = R_{n,S} \sqrt{n} F_n (\hat{\beta}^\dagger - \beta) \\ &\sim_a R_{n,S} \left(\int_0^1 X^\dagger(r) X^\dagger(r)' dr \right)^{-1} \int_0^1 X^\dagger(r) dB_{\tilde{\varepsilon}}(dr) \\ &\rightsquigarrow \mathcal{N}\left(0,\ \tilde{\omega}^2 R_S M^{\dagger -1} R_S'\right), \end{aligned} \tag{1.50}$$

with

$$R_{n,S} \to R_S := \begin{bmatrix} c_1 \iota_{q_1} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & c_2 \iota_{q_2} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & c_3 \iota_{q_3} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & c_p \iota_{q_p} & \cdots & 0 \end{bmatrix}_S, \tag{1.51}$$

where R_S is a $q \times (m+1)$ matrix of rank $r = \sum_{i=1}^p \mathbb{1}\{c_i \neq 0\} \leq q = \sum_{i=1}^p q_i \mathbb{1}\{c_i \neq 0\}$. When $r < q$, the limit distribution given in (1.50) is singular of rank r.

As in Examples 1 and 3, rank deficiency in the limit transformation matrix R_S in (1.51) and singularity of the limiting normal distribution in (1.50) affects the limit distribution of the Wald statistic and the degrees of freedom of the resulting chi-squared statistic. Following the earlier approach we have developed above, the limit distribution of the Wald statistic $W_n = \hat{\beta}^{\dagger\prime} \bar{H}' \left[\bar{H} \left(\tilde{X}^{\dagger\prime} \tilde{X}^\dagger \right)^{-1} \bar{H}' \right]^{-1} \bar{H} \hat{\beta}^\dagger / \hat{\tilde{\omega}}^2$ for testing the null hypothesis $\mathcal{H}_0 : \ \bar{H}\beta = 0$ is

$$
\begin{aligned}
W_n &= \hat{\beta}^{\dagger\prime}\,\bar{H}'\left[\frac{1}{n}\,\bar{H}F_n^{-1}\left(\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}F_n^{-1}\right)^{-1}F_n^{-1}\,\bar{H}'\right]^{-1}\bar{H}\hat{\beta}^{\dagger}/\hat{\tilde{\omega}}^2 \\
&= \frac{1}{n}\hat{\beta}^{\dagger\prime}\sqrt{n}F_nF_n^{-1}\,\bar{H}'Q_{n,S}^{-1}\left[\frac{1}{n}Q_{n,S}^{-1}\,\bar{H}F_n^{-1}\left(\frac{1}{n}F_n^{-1}\tilde{X}^{\dagger\prime}\tilde{X}^{\dagger}F_n^{-1}\right)^{-1}\right. \\
&\qquad \left.\times F_n^{-1}\,\bar{H}'Q_{n,S}^{-1}\right]^{-1}Q_{n,S}^{-1}\,\bar{H}F_n^{-1}\sqrt{n}F_n\hat{\beta}^{\dagger}/\hat{\tilde{\omega}}^2 \\
&\sim_a \hat{\beta}^{\dagger\prime}\sqrt{n}F_nR_{n,S}'\left[R_{n,S}M^{\dagger-1}R_{n,S}'\right]^{-1}R_{n,S}\sqrt{n}F_n\hat{\beta}^{\dagger}/\tilde{\omega}^2 \\
&\sim_a \int_0^1 dB_{\tilde{\varepsilon}}(r)X^{\dagger}(r)'M^{\dagger-1}R_{n,S}'\left[R_{n,S}M^{\dagger-1}R_{n,S}'\right]^{-1}R_{n,S}M^{\dagger-1} \\
&\qquad \times\int_0^1 X^{\dagger}(r)dB_{\tilde{\varepsilon}}(r)/\tilde{\omega}^2 \\
&= \int_0^1 dW_{\tilde{\varepsilon}}(r)X^{\dagger}(r)'M^{\dagger-1/2}P_{\mathcal{A}_{n,S}}M^{\dagger-1/2}\int_0^1 X^{\dagger}(r)dW_{\tilde{\varepsilon}}(r) \\
&= \mathcal{Z}'P_{\mathcal{A}_{n,S}}\mathcal{Z} \rightsquigarrow \mathcal{Z}'P_{\mathcal{A}_{\mathcal{S}}}\mathcal{Z} =_d \chi_r^2.
\end{aligned}
\tag{1.52}
$$

Justification of (1.52) follows as before, albeit with the additional notational complications of this general case. $\mathcal{Z} = M^{\dagger-1/2}\int_0^1 X^{\dagger}(r)dW_{\tilde{\varepsilon}}(r) =_d \mathcal{N}(0, I_m)$ and $\mathcal{Z}'P_{\mathcal{A}_{n,S}}\mathcal{Z}$ is a Gaussian quadratic form. The projection matrix $P_{\mathcal{A}_{n,S}} = \mathcal{A}_{n,S}\left(\mathcal{A}_{n,S}'\mathcal{A}_{n,S}\right)^{-1}\mathcal{A}_{n,S}'$ has rank $q = \text{rank}(\mathcal{A}_{n,S}) = \text{rank}(R_{n,S})$ for all finite n and is deterministic. This matrix projects onto the q-dimensional range space of the $(m+1)\times m$ matrix $\mathcal{A}_{n,S}$, where in (1.52) we have

$$
\mathcal{A}_{n,S}' = R_{n,S}M^{\dagger-1/2} \to R_SM^{\dagger-1/2} =: \mathcal{A}_{\mathcal{S}}', \tag{1.53}
$$

for which the limiting rank is $\text{rank}(\mathcal{A}_S) = r$, the same as the rank of the matrix R_S. The $(m+1)$-dimensional limiting projection matrix $P_{\mathcal{A}_S} = \mathcal{A}_S\left(\mathcal{A}_S'\mathcal{A}_S\right)^{+}\mathcal{A}_S'$ therefore has rank r. As in the examples discussed above, an explicit form of the projector $P_{\mathcal{A}}$ can be found as follows. Use the expression for the limit matrix R_s given in (1.51), and let $\{a_i'\}_{i=1}^{m+1}$ be the rows of the positive definite matrix $M^{\dagger-1/2} = (\int_0^1 X^{\dagger}(r)X^{\dagger}(r)'dr)^{-1/2}$. Then the $(m+1)\times q$ matrix $\mathcal{A}_S$ can be written as

$$
\mathcal{A}_S = M^{\dagger-1/2}R_S' = \begin{bmatrix} c_1a_1\iota_{q_1}', & c_2a_2\iota_{q_2}', & \cdots, & c_pa_p\iota_{q_p}' \end{bmatrix}_S \tag{1.54}
$$

in which the selection operator S removes those columns for which $c_j = 0$ so that $\mathcal{A}_S$ has the stated dimension $(m+1)\times q$ with $q = \sum_{i=1}^{p} q_i \mathbb{1}\{c_i \neq 0\}$. What remains in the matrix $\mathcal{A}_S$ are r linearly independent columns which we assemble in the $(m+1)\times r$ matrix $A_S = \left[c_1a_1, c_2a_2, \cdots, c_pa_p\right]_S$, which is easily seen to have full rank r because the matrix $(\int_0^1 X^{\dagger}(r)X^{\dagger}(r)'dr)^{-1/2}$ is nonsingular.

It follows directly that the projector $P_{\mathcal{A}_S}$ projects onto the range space of the matrix A_S and is therefore equivalent to the projector P_{A_S} by uniqueness. The limit distribution of the Wald statistic W_n is then $\mathcal{Z}' P_{\mathcal{A}_S} \mathcal{Z} = \mathcal{Z}' P_{A_S} \mathcal{Z} =_d \chi_r^2$, as stated in (1.52).

1.5 Conclusion

As is apparent from the analysis in the last section, testing multiple hypotheses in models with trend regressors involves complexities that can affect the degrees of freedom of the limiting χ^2 distribution. Whereas the least squares coefficients themselves, upon suitable centering and matrix normalization, have a standard limiting nonsingular normal distribution, Wald statistics based on them can suffer from degeneracies that are induced by the nature of the hypotheses being tested. These arise through the algebraic interaction of the normalizing matrix and the matrix form of the hypotheses being tested. The heuristic explanation is that lower rates of convergence in estimated coefficients inevitably dominate the limit theory, so that hypotheses involving linear combinations of coefficients estimated at different rates end up in the asymptotic theory being dominated by the variation of the coefficients estimated at the lowest rates, a feature that can induce degeneracy in multiple hypothesis testing when several hypotheses are of this type. As we have seen, this degeneracy leads to some algebraic complexity in a general analysis. But the heuristics remain accurate even in such cases and should enable a straightforward computation of the nondegenerate component in the Wald statistic limit theory and the appropriate degrees of freedom in the limiting χ^2 distribution.

While these features of trend regression induce potential degeneracies in the limit theory of Wald statistics for multiple hypotheses, the finite sample distributions typically do not suffer from the same reductions in degrees of freedom, as the analyses above make clear. The higher-order terms that disappear in limit theory may have a considerable influence in finite samples. These effects and the adequacy of the asymptotic theory may be investigated by simulation experiments and formal asymptotic expansions to reveal their importance in practice. Such an investigation is left for future research.

1.6 Appendix

Lemma 1.6.1 *In the deterministic m-vector sequence* $a_{tn} = a\left(\frac{t}{n}\right)$, *the function* $a(\cdot)$ *is assumed to be of bounded variation and partial sums of* u_t *are assumed to satisfy the functional law* $\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor nr \rfloor} u_t \rightsquigarrow B_u(r)$ *where* B_u *is Brownian motion with variance* $\omega^2 = \mathbb{V}^{LR}(u_t) = 2\pi f_u(0) > 0$. *Then* $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_{tn} u_t \rightsquigarrow \int_0^1 a(r) dB_u(r) = \mathcal{N}\left(0, \omega^2 \int_0^1 a(r)a(r)' dr\right)$.

Proof The proof is by partial summation. Note that $\Delta(f_t g_t) = (\Delta f_t)g_t + f_{t-1}\Delta g_t$. Summing the left side of this equality gives $\sum_{t=1}^{n}\Delta(f_t g_t) = f_n g_n - f_0 g_0$ and summing the right side gives $\sum_{t=1}^{n}(\Delta f_t)g_t + \sum_{t=1}^{n} f_{t-1}\Delta g_t$, which leads to the following partial summation formula

$$\sum_{t=1}^{n}(\Delta f_t)g_t = f_n g_n - f_0 g_0 - \sum_{t=1}^{n} f_{t-1}\Delta g_t. \tag{1.55}$$

Define the partial sum process $S_t = \sum_{j=1}^{t} u_j$ with $S_0 = 0$. Applying partial summation as in (1.55) we have $\sum_{t=1}^{n} a(\frac{t}{n})u_t = \sum_{t=1}^{n} a(\frac{t}{n})\Delta S_t = a(1)S_n - \sum_{t=1}^{n} S_{t-1}\left(a(\frac{t}{n}) - a(\frac{t-1}{n})\right)$. Then upon standardization by $\sqrt{n}$ the following weak convergence holds as $n \to \infty$:

$$\begin{aligned}\sum_{t=1}^{n} a\left(\frac{t}{n}\right)\frac{u_t}{\sqrt{n}} &= a(1)\frac{S_n}{\sqrt{n}} - \sum_{t=1}^{n}\frac{S_{t-1}}{\sqrt{n}}\Delta a\left(\frac{t}{n}\right)\\ &\rightsquigarrow a(1)B_u(1) - \int_0^1 B_u(r)da(r),\end{aligned} \tag{1.56}$$

because $\frac{1}{\sqrt{n}}\sum_{t=1}^{\lfloor nr\rfloor} u_t \rightsquigarrow B_u(r)$ and $\sum_{t=1}^{n}\frac{S_{t-1}}{\sqrt{n}}\Delta a\left(\frac{t}{n}\right) \rightsquigarrow \int_0^1 B_u(r)da(r)$ by virtue of Riemann–Stieltjes integration which may be employed because the limit process $B_u(\cdot)$ is continuous almost surely and $a(\cdot)$ is of bounded variation by assumption. We may then define the Wiener stochastic integral $\int_0^1 a(r)dB_u(r)$ with respect to Brownian motion by virtue of the integration by parts formula

$$\begin{aligned}\int_0^1 a(r)dB_u(r) &= [a(r)B(r)]_0^1 - \int_0^1 B_u(r)da(r)\\ &= a(1)B_u(1) - \int_0^1 B_u(r)da(r).\end{aligned} \tag{1.57}$$

The functional $\int_0^1 a(r)dB_u(r)$ is Gaussian with zero mean and variance matrix $\omega^2\int_0^1 a(r)a(r)'dr$ since Brownian motion has independent increments and $\mathbb{E}(dB_u(s)dB_u(r)) = \omega^2 dr\mathbb{1}\{r = s\}$. The special case considered in the paper involves time polynomials of the form $a(\frac{t}{n}) = (\frac{t}{n})^k$ for integer $k \geq 0$ and then $\int_0^1 r^p dB_u(r) = \mathcal{N}\left(0, \omega^2\int_0^1 r^{2p}dr\right) = \mathcal{N}\left(0, \frac{\omega^2}{2p+1}\right)$. ■

Remark

(iii) More general versions of Lemma 1.6.1 hold and are proved in the same manner allowing for stochastic integration of bounded variation functions, including stochastic functions, and for functional laws to continuous stochastic processes other than Brownian motion. For instance, suppose the time series u_t is such that $\frac{1}{d_n}\sum_{t=1}^{\lfloor nr \rfloor} u_t \rightsquigarrow Y(r)$ where d_n is an increasing numerical sequence with $d_n \to \infty$ and $Y(r)$ is a limiting stochastic process with continuous sample paths almost surely. Then

$$\sum_{t=1}^{n} a\left(\frac{t}{n}\right)\frac{u_t}{d_n} = a(1)\frac{S_n}{d_n} - \sum_{t=1}^{n}\frac{S_{t-1}}{d_n}\Delta a\left(\frac{t}{n}\right)$$
$$\rightsquigarrow a(1)Y(1) - \int_0^1 Y(r)da(r) =: \int_0^1 a(r)dY(r). \tag{1.58}$$

With an appropriate choice of the normalization sequence d_n, (1.58) includes the case where $Y(r) = B_H(r)$ is fractional Brownian motion with Hurst parameter H. Further, if $P_t = \sum_{s=1}^{t} S_s$ then it follows by continuous mapping that a suitably normalized version of P_t satisfies the weak convergence $\frac{1}{nd_n}P_{\lfloor nr \rfloor} = \frac{1}{n}\sum_{s=1}^{t}\frac{S_s}{d_n} \rightsquigarrow \int_0^r Y(q)dq =: \bar{Y}(r)$, which is of bounded variation because $Y(q)$ is continuous. We therefore have

$$\frac{1}{nd_n}\sum_{t=1}^{n} P_t\frac{u_t}{d_n} = \frac{P(1)}{nd_n}\frac{S_n}{d_n} - \sum_{t=1}^{n}\frac{S_{t-1}}{d_n}\frac{1}{nd_n}\Delta P_t$$
$$\rightsquigarrow \int_0^1 Y(q)dqY(1) - \int_0^1 Y(r)d\bar{Y}(r)$$
$$= \bar{Y}(1)Y(1) - \int_0^1 Y(r)^2 dr =: \int_0^1 \bar{Y}(r)dY(r), \tag{1.59}$$

by Riemann–Stieltjes integration, thereby defining the Wiener stochastic integral $\int_0^1 \bar{Y}(r)dY(r)$ with respect to $Y(r)$.

Bibliography

Andrews, D. W. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory 3*(3), 348–358.

Grenander, U. and M. Rosenblatt (1954). Regression analysis of time series with stationary residuals. *Proceedings of the National Academy of Sciences of the United States of America 40*(9), 812.

Kheifets, I. and P. C. B. Phillips (2023). Fully modified least squares cointegrating parameter estimation in multicointegrated systems. *Journal of Econometrics*, 232(2), 300–319.

Magdalinos, T. and P. C. B. Phillips (2018). Wald testing with matrix normalization. Working Paper, *University of Southampton*.

Park, J. Y. and P. C. B. Phillips (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory 4*(3), 468–497.

Park, J. Y. and P. C. B. Phillips (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory 5*(1), 95–131.

Phillips, P. C. B. (1987, 12). Multiple regression with integrated time series. *Contemporary Mathematics 80*, 79–106.

Phillips, P. C. B. (1995). Fully modified least squares and vector autoregression. *Econometrica 63*, 1023–1078.

Phillips, P. C. B. and I. L. Kheifets (2024). High-dimensional IV cointegration estimation and inference. *Journal of Econometrics*, 238(2), 105622.

Phillips, P. C. B., D. Li, and J. Gao (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics 196*(1), 180–195.

Phillips, P. C. B. and T. Magdalinos (2013). Inconsistent var regression with common explosive roots. *Econometric Theory 29*(4), 808–837.

Phillips, P. C. B. and J. Y. Park (1988). Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors. *Journal of the American Statistical Association 83*(401), 111–115.

Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics 20*(2), 971–1001.

Puri, M. L., C. T. Russell, and T. Mathew (1984). Convergence of generalized inverses with applications to asymptotic hypothesis testing. *Sankhyā: The Indian Journal of Statistics, Series A 46*(2), 277–286.

Stewart, G. W. (1969). On the continuity of the generalized inverse. *SIAM Journal on Applied Mathematics 17*(1), 33–45.

Vuong, Q. H. (1987). Generalized inverses and asymptotic properties of Wald tests. *Economics Letters 24*(4), 343–347.