4

Nonsense Regressions

The problem of nonsense (often also called spurious) relationships goes back to Yule (1926), and has more recently been studied by Granger and Newbold (1974) and Phillips (1986). It remains an important issue in econometric modelling, since it is not easy to distinguish causal from spurious relationships between economic variables. This chapter helps in developing skills to detect nonsense regressions, showing what diagnostic statistics may be misleading, and what kinds of information may be useful to discover such relationships. In particular, it shows that large R^2s and highly-significant t-ratios are insufficient for causality, and that alternative diagnostic statistics may provide evidence on spurious relations. It also illustrates that the worse model may fit better due to invalid marginalization, and that high R^2s and highly-significant t-ratios may be due to variables behaving as difference stationary. Hence, examining the properties of the individual data series may be helpful when constructing econometric models: this chapter shows that correlograms of, and unit root tests on, individual series may provide information on their stationarity status. It is also found that stationarity may rule out interpreting some equations as behavioural models. The statistics used below are analyzed by deriving their asymptotic properties. However, asymptotic approximations may not be accurate, and hence it is useful to examine their small-sample properties. Monte Carlo techniques are suitable for doing so. Hence, this chapter continues the task of previous chapters in teaching the design of Monte Carlo studies to look at finite-sample properties of estimators and t-ratios, now applied to spurious relationships between unrelated variables generated by $\mathsf{I}(1)$ processes. The usefulness of asymptotic results to guide the design is highlighted by computing finite-sample biases and t-statistics for just a few values in the parameter space. Support is found for the argument provided above: unit roots lead to highly-significant t-ratios because estimated standard errors underestimate the finite-sample standard deviations of estimates, and correlation coefficients between unrelated variables converge to random variables, not fixed values. Exercises 4.1 and 4.2 consider the detection of nonsense regressions using the UK-money data, then §4.3 looks at the empirical conditional and marginal distributions. Then §4.4 undertakes the Monte Carlo study of OLS estimators in nonsense regressions, and examines the rejection frequency of the t-test of no relation. Expectations processes are investigated in §4.5, both with and without unit roots in the DGP, then the last two sections consider the asymptotic properties of OLS in nonsense regressions, and in an autoregressive model when the DGP has a unit root with drift.

4.1 UK-money nonsense regressions

(1) Using the data set UKM1, cumulate R_t over time to create $G_t = \sum_{j=1}^{t} R_j$.
(2) Regress $\log M_t$ on $\log P_t$, and discuss the results.
(3) Regress $\log M_t$ on G_t. Discuss these results. What criteria, if any, would help to distinguish which of the preceding regressions was meaningful or nonsense?

4.1.1 Creating variables in *GiveWin*

Enter *GiveWin* and load the data file UKM1. Click on the calculator icon. Highlight the variable R from the list-of-variables' window. Select cum(var) from the window of functions. Click on the '$=$' key. Type G in the name-editor window and accept.

4.1.2 Detecting a relationship between money and prices

Let x_t denote $\log X_t$. We wish to elicit evidence on the existence of a relationship between the logarithms of money and prices from the regression:

$$m_t = \beta_0 + \beta_1 p_t + \epsilon_t.$$

OLS estimation in *GiveWin* yields the following results over $1963(1)$–$1989(2)$:

$$\widehat{m}_t = \underset{(0.03)}{10.9} + \underset{(0.02)}{1.04}\; p_t \tag{4.1}$$

$$\mathsf{R}^2 = 0.95 \;\; \widehat{\sigma} = 0.189 \;\; DW = 0.02 \;\; ADF\,(2) = -0.04.$$

A superficial analysis of these results suggests that a large proportion of the variance of m_t is explained by prices ($\mathsf{R}^2 = 0.95$), consistent with the 'highly significant' t-ratio on the price coefficient. However, can these results be regarded as evidence for prices determining money as against reflecting 'common trends'?

First, fig. 4.1 shows the fitted and actual values, the residuals, their correlogram, and their histogram with a non-parametric estimate of the density. It is clear that the relation is not really close, deteriorates over time, and the residuals are highly autocorrelated and non-normal. Such a model would not be useful for forecasting, and is far from being strong evidence of a causal link.

Direct investigation of the data also provides useful information. The correlograms of m and p in fig. 4.2, and the values of the *ADF* statistics shown in table 4.1, suggest that m is possibly $\mathsf{I}(1)$ whereas p behaves more like $\mathsf{I}(2)$ (t-values on the first lag of the regressand are shown in parentheses: a constant and linear deterministic trend were included). Thus, the null of a unit root is never rejected for the levels, is rejected in some cases for the first differences, and is always rejected for the second differences (check for yourself whether this continues to hold for longer lags; whether they are significant, and if a trend is needed, or justified). The correlogram for Δp dies out slowly, hence it

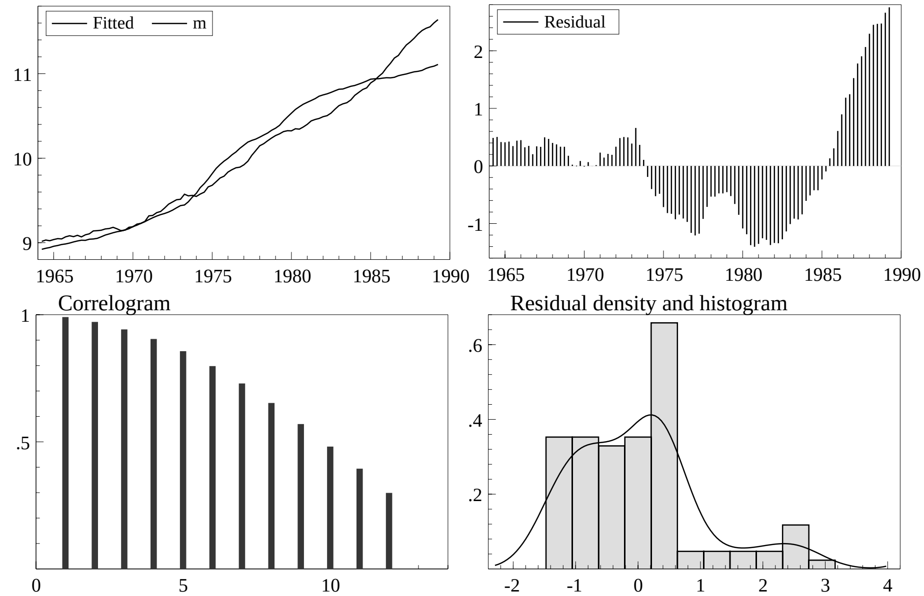

Figure 4.1 Model statistics for money on prices

is hard to detect that the root is less than unity – which it may not be. Indeed, roots are not inherent properties of time series, and the degree of integration may vary over time, although the recursive estimates of the roots shown in fig. 4.3b–d do not suggest that is the explanation here (the last must be significantly different from unity to reject $\mathsf{I}(3)$).

Table 4.1 Unit-root statistics for m and p

	m	Δm	$\Delta^2 m$	p	Δp	$\Delta^2 p$
ADF(0)	-0.96	-8.82^{**}	-19.08^{**}	-0.72	-3.22	-13.53
ADF(1)	-1.01 (1.29)	-5.19^{**} (-2.39)	-100^{**} (1.32)	-1.59 (14.36)	-2.46 (-2.32)	-9.71^{**} (1.95)

Hendry (1995a, p.129) reports Monte Carlo results on the behaviour of estimates when the variables in the regression are unrelated and cointegrated of different orders. According to those results, if m and p behaved as $\mathsf{I}(1)$ and $\mathsf{I}(2)$, respectively, we would expect a zero-mean R^2, but its variance increases with the sum of the orders of integration; an estimated standard error (ESE) which underestimates the Monte Carlo estimate of the finite-sample standard error (MCSD); and a large probability of rejecting the true null of $\beta_1 = 0$. Hence, a large value of R^2 and a highly-significant t-ratio cannot be interpreted to imply the existence of a causal relationship between m and p, in the absence of acceptable diagnostic-statistic results.

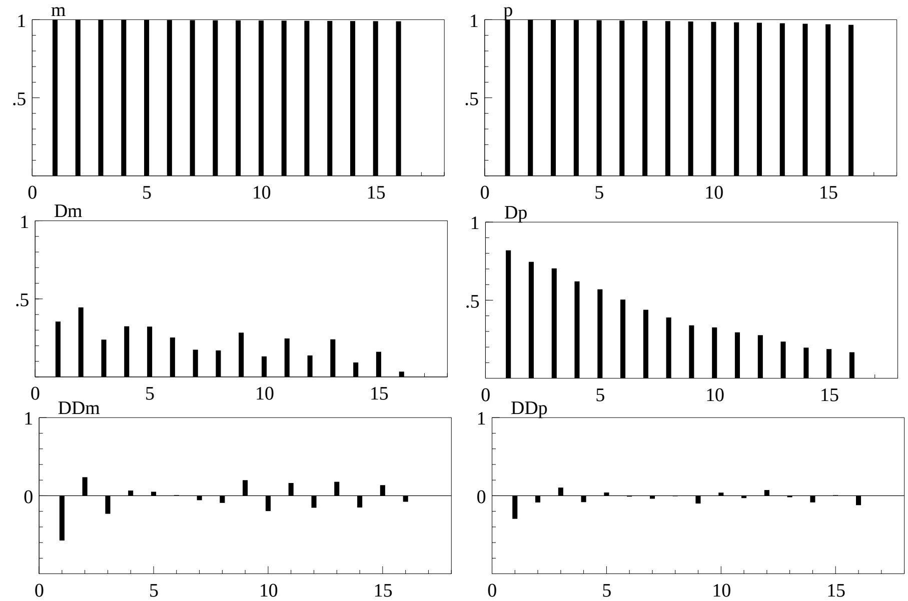

Figure 4.2 Correlograms of money and prices

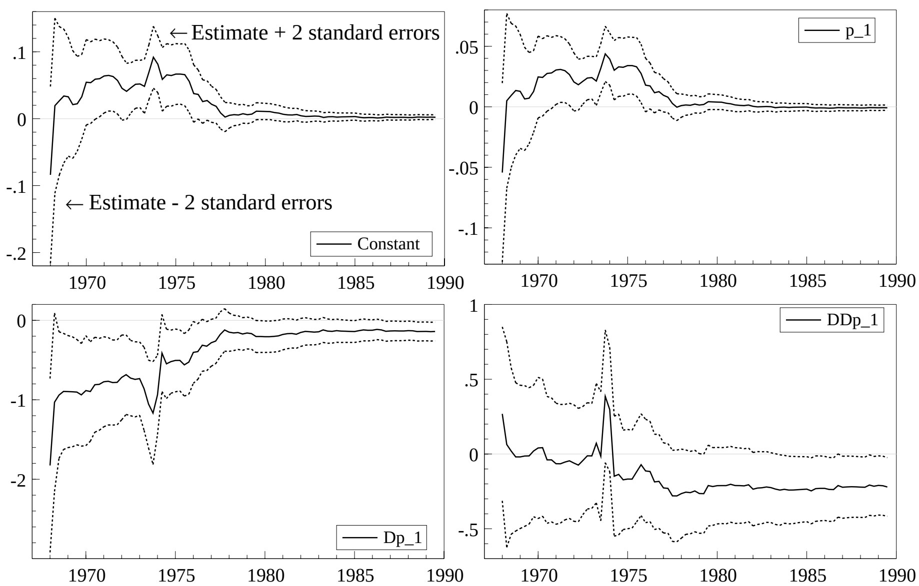

Figure 4.3 Recursive estimates of roots for prices

What would help us to distinguish a nonsense from a genuine relationship between money and prices? First, if m and p had genuinely different orders of integration, then they could not be directly related. Next, if m and p were unrelated $\mathsf{I}(2)$ series, then ϵ_t would, in general, also be $\mathsf{I}(2)$ and hence we may expect *DW* to converge to zero as T increases: Phillips (1986) shows that this is the case for $\mathsf{I}(1)$ unrelated series. Conversely, if m and p cointegrated to $\mathsf{I}(0)$, ϵ_t must be stationary, so that we expect *DW* to show a value away from zero. In addition, we may expect unit-root tests on residuals to show large negative values. Finally, if they cointegrated to I(1) (thereby removing the I(2) component), ϵ_t would be $\mathsf{I}(1)$ and additional variables would be needed to produce an $\mathsf{I}(0)$ error. In our empirical results, *DW* is very small, and *ADF* tests on the residuals in regression (4.1) do not allow us to reject the unit-root hypothesis, suggesting the lack of a relationship between money and prices. However, Hendry (1995a, ch.16) reports a congruent model of money on prices and other variables, so we have illustrated how difficult discriminating between genuine and nonsense regressions may be.

4.1.3 Detecting nonsense regressions between money and interest rates

Chapter 1.2 noted that $\log R_t$ does not have a clear interpretation (as asked in the original exercise), so we consider m_t on G_t. The correlograms in fig. 4.4 suggest that R_t becomes white noise after first differencing, supported by the values of the *DF* statistics in table 4.2.

Table 4.2 Unit-root statistics for R and G

	R	ΔR	G	ΔG	$\Delta^2 G$
ADF(0)	-2.02	-8.44^{**}	-0.09	-2.02	-8.44^{**}
ADF(1)	$-\underset{(2.42)}{2.25}$	-6.48^{**}	$-\underset{(26.1)}{1.04}$	$-\underset{(2.42)}{2.25}$	-6.48^{**}

If, in fact, R_t was generated by a random walk with no drift:

$$R_t = R_{t-1} + \eta_t,$$

then this would occur. However:

$$\Delta G_t = R_t,$$

and so G_t would be $\mathsf{I}(2)$. We wish to evaluate the existence of a relationship between m_t and G_t by estimating the equation:

$$m_t = \beta_0 + \beta_1 G_t + \epsilon_t.$$

We found in §4.1.2 that m behaves as $\mathsf{I}(1)$, so to proceed further, we investigate the conjecture that G_t is non-stationary. Figure 4.4(c)–(e) shows the correlograms for G up to its second difference, and suggests that $\Delta^2 G_t$ is stationary, supported by the values of the *DF* statistic in table 4.2. Since we do not anticipate drift in interest rates, no trend

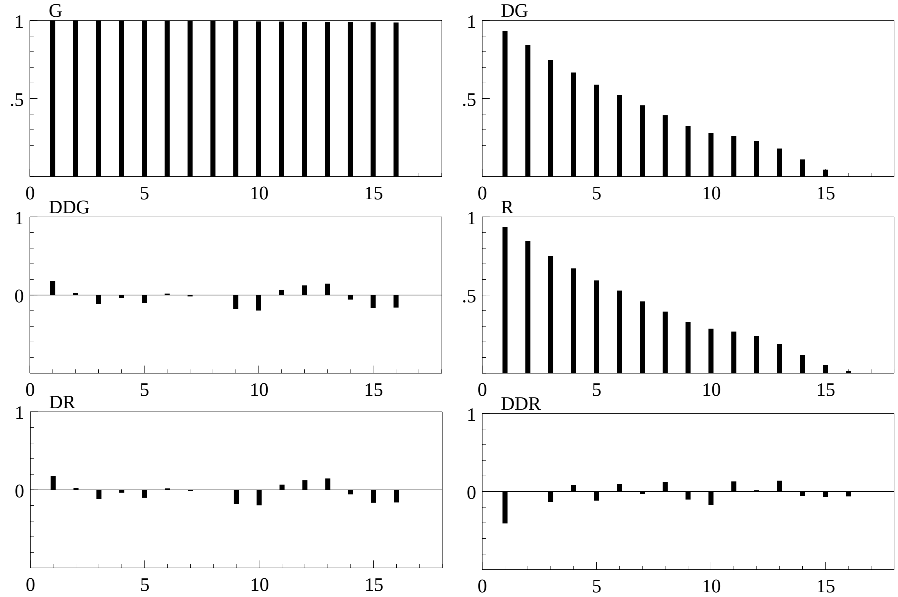

Figure 4.4 Correlograms of levels and cumulated interest rates

is included in these tests. Putting together the evidence provided by the correlograms and the tests, it seems that G is $\mathsf{I}(2)$. Thus, consider the estimated regression:

$$\widehat{m}_t = \underset{(0.03)}{8.79} + \underset{(0.005)}{0.26}\, G_t \tag{4.2}$$

$$\mathsf{R}^2 = 0.97 \;\; \widehat{\sigma} = 0.15 \;\; DW = 0.03 \;\; ADF\,(2) = -1.02.$$

As in (4.1), R^2 is large and β_1 is 'highly significant', but as in §4.1.2, we cannot rely on values of either the t-ratio or R^2. Finally, unit-root tests suggest that the residuals are non-stationary, and hence that there is no relationship between m and G. Hendry (1995a, ch.16) finds evidence of a relationship between money and interest rates rather than between money and cumulative interest rates. Thus, the results in §4.1.2 and those just obtained illustrate how easy it is to derive the same conclusions from situations of apparently valid and of non-existent relationships. Note that: the completely nonsense relation in (4.2) actually fits better than using the variable predicted by economic theory as the main determinant of money holdings (4.1); the regression of m on R delivers $\mathsf{R}^2 = 0.24$ only; and m_{t-1} is highly significant if added to any of these least-squares relations, usually raising R^2 to over 0.999.

Similar findings result for p_t and G_t: the cross plot of these shows an even closer relation ($\mathsf{R}^2 = 0.986$). Any comment?

4.2 Properties of empirical data

(1) Using the variable R_t from the data set UKM1, examine its correlogram, and compare the empirical results with what would be anticipated if R_t followed a first-order autoregressive process.
(2) Does the process seem to be stationary? If not, to what case does it conform?
(3) Fit a first-order autoregression to R_t and discuss the results. Are the residuals an innovation process? Are they approximately normally distributed?
(4) On the basis of all these results, how would you characterize R_t?
(5) Does it matter what sample period is selected? Is there any noticeable change after 1984?
(6) Repeat the analysis in (1)–(5) using ΔR_t.

4.2.1 Correlograms

Figure 4.4 suggests that R may have been generated by an $\mathsf{AR}(1)$ process with a coefficient close to 0.9, so the population value may be near unity.

4.2.2 Stationarity

Indeed, the information provided by the unit-root tests in table 4.2 suggest that R is non-stationary, and that there is a unit root in its process.

4.2.3 A first-order autoregression

The estimated $\mathsf{AR}(1)$ model over $1963(2)$–$1989(4)$ is:

$$\begin{aligned}
\widehat{R}_t &= \underset{(0.004)}{0.01} + \underset{(0.04)}{0.91}\ R_{t-1} \qquad (4.3)\\
\mathsf{R}^2 &= 0.84\ \ \widehat{\sigma} = 0.013\ \ DW = 1.66\ \ J = 0.85\\
\mathsf{F}_{ar}(4,101) &= 1.11\ \ \mathsf{F}_{\mathsf{arch}}(1,103) = 10.2^{**}\ \ \chi^2_{\mathsf{nd}}(2) = 3.22\\
\mathsf{F}_{het}(2,102) &= 3.50^{**}\ \ \mathsf{F}_{reset}(1,104) = 1.16.
\end{aligned}$$

In (4.3), the diagnostic tests are of the form $\mathsf{F}_{\mathsf{j}}\ (k, T-l)$ which denotes an F-test against the alternative hypothesis j for: k^{th}-order serial correlation (F_{ar}: see Godfrey, 1978), k^{th}-order autoregressive conditional heterocedasticity ($\mathsf{F}_{\mathsf{arch}}$: see Engle, 1982), heterocedasticity ($\mathsf{F}_{\mathsf{het}}$: see White, 1980a); functional-form mis-specification (F_{reset}: see Ramsey, 1969); and a chi-square test for normality ($\chi^2_{\mathsf{nd}}(2)$: see Doornik and Hansen, 1994). * and ** denote significance at the 5% and 1% levels respectively. J is the joint parameter-constancy test from Hansen (1992). The insignificant residual autocorrelation and normality tests suggest that they are uncorrelated and approximately normally distributed with zero mean. However, tests against heterocedasticity are significant indicating that residuals are not white noise, and hence are predictable in part from their own past. The hypothesis that the autoregressive coefficient is unity cannot be rejected under that null, an issue to which we now turn.

4.2.4 Characterization of the process generating R_t

The apparently contradictory results found in §4.2.1 and §4.2.2 can be reconciled. The estimated coefficient of lagged interest rates in §4.2.3 is close to 0.9, and the normality test suggests that the residuals are independently normally distributed with zero mean. If R is actually a random walk with no drift:

$$R_t = R_{t-1} + \epsilon_t, \quad \text{with} \quad \epsilon_t \sim \mathsf{IN}\left[0,1\right],$$

and $R_0 = 0$, then $R_t = \sum_{i=1}^{t} \epsilon_i = S_t$. The OLS estimator of α in the regression:

$$R_t = \alpha R_{t-1} + u_t,$$

is:[1]

$$\widehat{\alpha} = \left(\sum_{t=1}^{T} R_{t-1}^2 \right)^{-1} \sum_{t=1}^{T} R_{t-1} R_t = 1 + \left(\sum_{t=1}^{T} S_{t-1}^2 \right)^{-1} \sum_{t=1}^{T} S_{t-1} \epsilon_t,$$

so that:

$$\begin{aligned} T\left(\widehat{\alpha} - 1\right) &= \left(T^{-2} \sum_{t=1}^{T} S_{t-1}^2 \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} S_{t-1} \epsilon_t \right) \\ &\Rightarrow \left(\int_0^1 W\left(r\right)^2 \mathrm{d}r \right)^{-1} \int_0^1 W\left(r\right) \mathrm{d}W\left(r\right), \end{aligned}$$

and hence $\widehat{\alpha} \xrightarrow{P} 1$. When R_t is $\mathsf{I}(1)$, we may expect an estimated coefficient close to unity, so the results obtained from estimating an $\mathsf{AR}(1)$ model are compatible with R being a random walk. Although the diagnostic tests in (4.3) suggest that an $\mathsf{AR}(1)$ model is not a complete description of the DGP for interest rates, this is unlikely to overturn the finding of a near-unit root. Indeed, the heterocedasticity is sensibly interpreted as revealing higher variances at higher interest rates: models of $\Delta R_t / R_{t-1} \simeq \Delta \log R_t$ show little heterocedasticity (but see the earlier remarks on $\log R_t$).

4.2.5 Effect of the sample period on the results

Consider estimating the $\mathsf{AR}(1)$ model for the two sub-sample periods $1963(2)$–$1983(4)$ and $1984(1)$–$1989(4)$. The results are the following:

$$\widehat{R}_t = \underset{(0.004)}{0.01} + \underset{(0.04)}{0.91} R_{t-1}$$

[1] Strictly, $R_0 = 0$ is very unlikely, hence the intercept in the empirical model; it is suppressed in the theoretical analysis for simplicity, and does not materially affect its implications.

$$\begin{aligned}
\mathsf{R}^2 &= 0.85 \;\; \widehat{\sigma} = 0.014 \;\; DW = 1.53 \; J = 1.12^{*} \\
\mathsf{F}_{ar}(4,77) &= 1.65 \;\;\; \mathsf{F}_{arch}(1,79) = 13.4^{**} \;\; \chi^2_{\mathsf{nd}}(2) = 3.02 \\
\mathsf{F}_{het}(2,78) &= 4.84^{**} \;\; \mathsf{F}_{reset}(1,80) = 0.76
\end{aligned}$$

and:

$$\begin{aligned}
\widehat{R}_t &= \underset{(0.005)}{0.01} + \underset{(0.09)}{0.87} \; R_{t-1} \\
\mathsf{R}^2 &= 0.81 \;\; \widehat{\sigma} = 0.012 \;\; DW = 1.84 \;\; J = 0.92 \\
\mathsf{F}_{\mathsf{ar}}\,(4,18) &= 0.23 \;\;\; \mathsf{F}_{\mathsf{arch}}\,(1,19) = 1.34 \;\; \chi^2_{\mathsf{nd}}(2) = 0.86 \\
\mathsf{F}_{\mathsf{het}}\,(2,19) &= 3.86^{*} \;\; \mathsf{F}_{\mathsf{reset}}\,(1,21) = 0.14
\end{aligned}$$

Thus, there is no significant change in the results. In both sub-samples, the coefficients are close to 0.9, and residuals are normal but predictable in part from their own past. Given the introduction of interest-bearing retail sight deposits in 1984, and the resulting redefinition of R to a net rate (see Hendry and Ericsson, 1991b), accompanied by a sharp drop in its value, the 'consistency' of these estimates is actually surprising.

4.2.6 First differences of R_t

Figure 4.4b shows the correlogram for ΔR_t. Both, the correlogram and value of the *DF* statistic in table 4.2 suggest that ΔR_t is stationary. Such results appear consistent with the following estimated $\mathsf{AR}(1)$ model in first differences:

$$\widehat{\Delta R}_t = \underset{(0.001)}{0.0002} + \underset{(0.10)}{0.18} \; \Delta R_{t-1} \tag{4.4}$$

$$\begin{aligned}
\mathsf{R}^2 &= 0.03 \;\; \widehat{\sigma} = 0.013 \, DW = 2.0 \;\; J = 0.66 \\
\mathsf{F}_{ar}(4,100) &= 0.42 \;\;\; \mathsf{F}_{arch}(1,102) = 11.6^{**} \;\; \chi^2_{\mathsf{nd}}(2) = 6.7^{*} \\
\mathsf{F}_{het}(2,101) &= 4.95^{**} \;\; \mathsf{F}_{reset}(1,103) = 7.3^{**}
\end{aligned}$$

DW suggests that the residuals in (4.4) are stationary. However, the lagged coefficient is barely significant, and diagnostic statistics suggest that the residuals are not innovations, and not normally distributed, warning against regarding the $\mathsf{AR}(1)$ model in first differences as a complete representation of the DGP for interest rates.

Consider partitioning the sample as in §4.2.5:

$$\begin{aligned}
\widehat{\Delta R}_t &= \underset{(0.002)}{0.001} + \underset{(0.11)}{0.21} \; \Delta R_{t-1} \\
\mathsf{R}^2 &= 0.03 \;\; \widehat{\sigma} = 0.013 \, DW = 2.0 \;\; J = 1.06^{*} \\
\mathsf{F}_{ar}(4,76) &= 0.66 \;\; \mathsf{F}_{arch}(1,78) = 6.24^{**} \;\; \chi^2_{\mathsf{nd}}(2) = 8.1^{*} \\
\mathsf{F}_{het}(2,77) &= 3.07 \;\; \mathsf{F}_{reset}(1,79) = 3.56
\end{aligned}$$

and:

$$\widehat{\Delta R}_t = \underset{(0.002)}{0.001} + \underset{(0.11)}{0.21}\ \Delta R_{t-1}$$

$$\begin{aligned}
\mathsf{R}^2 &= 0.002\ \ \widehat{\sigma} = 0.013\ DW = 2.0\ \ J = 0.59 \\
\mathsf{F}_{ar}(4,18) &= 0.14\ \ \mathsf{F}_{arch}(1,20) = 6.64^{**}\ \ \chi^2_{\mathsf{nd}}(2) = 2.1 \\
\mathsf{F}_{het}(2,19) &= 2.77\ \ \mathsf{F}_{reset}(1,21) = 15.2^{**}
\end{aligned}$$

The coefficients are within one standard error of each other, and so the general conclusion is the same for both sample periods, although the first period shows some signs of parameter non-constancy, and the second has some functional-form mis-specification.

4.3 Empirical conditional and marginal distributions

Consider the joint density of the three variables $\Delta\log M_t$, $\Delta\log P_t$, and $\Delta\log Y_t$ from the data-set UKM1.

(1) Estimate an autoregressive process for each variable, and discuss the results obtained. What order of lag length did you select and why? Are the results sensitive to the length chosen? Did you include an intercept? Does the sample period selected matter?
(2) To characterize the marginal distributions, regress each variable on two lagged values of itself and both other variables. Does the past of any variable help explain any other variable?
(3) Estimate a conditional model for $\Delta\log M_t$ given $\Delta\log P_t$ and $\Delta\log Y_t$, and one lagged value of each variable. Discuss your results, and interpret the evidence in light of the theoretical analysis in this chapter.

4.3.1 Autoregressions

Because the observations are quarterly, we consider a fifth-order autoregression with drift as the most general specification for each variable. Lags beyond second do not appear significant for explaining Δm_t. Lack of significance of higher-order lags cannot be associated with the non-orthogonality of regressors because correlations between them are low (the largest correlation is 0.46). Hence, in light of the evidence considered, the $\mathsf{AR}(2)$ in table 4.3 may be regarded as a valid reduction of the more general $\mathsf{AR}(5)$ model. No lags appear significant in a fifth-order autoregression for Δy_t in spite of regressor orthogonality (the largest correlation is 0.10), but the residuals appear non-normal. Non-orthogonality of inflation lags may cause higher lags to appear insignificant in an $\mathsf{AR}(5)$ as most correlations between regressors are above 0.7, and the smallest is 0.57. However, we can rewrite the model in terms of $\Delta^2 p_{t-i},\ i = 1,\ldots,4$

and Δp_{t-1} without loss of generality, so that correlations fall (in absolute value) below 0.3. All coefficients become highly significant, and close to one another, as are lags higher than five. Tests show that the residuals are not innovations, and are non-normally distributed. Standard errors are given in parentheses, and only significant diagnostic statistics are reported. The first three columns are 1963(4)–1989(2); the fourth is 1984(1)–1989(2); and the last two are 1963(3)–1973(4).

Table 4.3 Autoregressions for Δm, Δp and Δy

Regressand	Δm_t	Δp_t	Δy_t	Δm_t	Δp_t	Δy_t
Constant	$\underset{(0.003)}{0.011}$	$\underset{(0.001)}{0.003}$	$\underset{(0.002)}{0.007}$	$\underset{(0.010)}{0.044}$	$\underset{(0.068)}{0.006}$	$\underset{(0.003)}{0.012}$
Lag1	$\underset{(0.093)}{0.225}$	$\underset{(0.098)}{0.628}$	$\underset{(0.010)}{-0.076}$	$\underset{(0.221)}{-0.015}$	$\underset{(0.017)}{0.675}$	$\underset{(0.014)}{-0.313}$
Lag2	$\underset{(0.093)}{0.367}$	$\underset{(0.097)}{0.226}$	$\underset{(0.097)}{0.098}$			
$\widehat{\sigma}$	0.018	0.008	0.014	0.014	0.008	0.014
$\underset{(k,T-l)}{\mathsf{F}_{\mathsf{arch}}}$	–	$\underset{(4,92)}{2.9^{*}}$	–		$\underset{(2,36)}{16.7^{**}}$	
$\chi^2_{\mathsf{nd}}(2)$	–	9.3^{*}	10.1^{**}		9.4^{**}	7.4^{**}
$\underset{(k,T-l)}{\mathsf{F}_{\mathsf{het}}}$					$\underset{(2,37)}{5.4^{**}}$	
$\underset{(k,T-l)}{\mathsf{F}_{\mathsf{fun}}}$	–	$\underset{(5,94)}{4.2^{**}}$	–			
$\underset{(k,T-l)}{\mathsf{F}_{\mathsf{reset}}}$					$\underset{(1,39)}{13.0^{**}}$	$\underset{(1,39)}{5.9^{**}}$

The results are somewhat sample-period dependent as shown by columns 4–6: lags no longer appear significant for changes in money, the second lag has become insignificant in the inflation equation; and the first lag appears significant for changes in income. Nevertheless, the values of $\widehat{\sigma}$ are similar across periods, as are the main determinants of the dependent variables: this is despite picking the sub-samples to correspond to known changes in the UK economy which had important impacts on these variables (the introduction of own interest rates in 1984; the oil shocks of the 1970s on inflation; and the attempt to 'go for growth' in the mid-1970s).

4.3.2 Marginal models

The estimated marginal models for each variable are shown in table 4.4. The past changes of the other two variables do not appear relevant to any regressand. In practice, this is often interpreted as if the marginal distribution of each dependent variable is independent of the remaining two variables, and so, as if the remaining two variables do not Granger cause the dependent variable (see Granger, 1969). However, feedbacks from cointegrating vectors have not been tested, so no substantive conclusions can be drawn about causal links.

Table 4.4 Marginal distributions for Δm, Δp and Δy

Regressand	Δm_t	Δp_t	Δy_t
Constant	0.008 (0.005)	0.002 (0.002)	0.010 (0.004)
Δm_{t-1}	0.220 (0.10)	-0.021 (0.04)	0.059 (0.069)
Δm_{t-2}	0.360 (0.10)	0.040 (0.042)	0.146 (0.069)
Δp_{t-1}	-0.100 (0.230)	0.623 (0.10)	-0.269 (0.164)
Δp_{t-2}	0.230 (0.240)	0.246 (0.10)	-0.064 (0.169)
Δy_{t-1}	0.110 (0.14)	0.042 (0.06)	-0.218 (0.103)
Δy_{t-2}	-0.03 (0.14)	-0.002 (0.06)	0.000 (0.097)
$\widehat{\sigma}$	0.018	0.008	0.013
$\mathsf{F}_{\mathsf{arch}}\,(4,88)$	–	3.4^{*}	–
$\chi^2_{\mathsf{nd}}(2)$	–	9.3^{*}	17.5^{**}
$\mathsf{F}_{het}\,(12,83)$		2.9^{**}	
$\mathsf{F}_{fun}\,(27,68)$	–	4.2^{**}	–.

Standard errors are given in parentheses and only significant diagnostic statistics are provided.

4.3.3 Conditional models

We estimate models for each variable conditional upon the contemporaneous values of the other two variables over the sample period $1963\,(3)$–$1989\,(2)$. The results are provided in table 4.5. Only current inflation is barely significant for money, and vice versa.

However, these results cannot be taken substantively because inference is invalid when current-dated regressors are not weakly exogenous for the parameters in the equation. In fact, the diagnostic tests warn about the specification in all three equations, and additional tests on the money equation must be conducted to check weak exogeneity (see ch.5). In any case, each equation is merely a renormalization of the same set of variables, and leads to 18 regression and 3 error coefficients, as against 12 regression and 6 error coefficients in the joint distribution, or 15 regression and 3 error coefficients in a proper conditional factorization. Thus, such representations are ‘over-parameterized’ (with implicit cross-equation restrictions on the estimates), although any one column could be valid under suitable assumptions.

Table 4.5 Conditional distributions for Δm, Δp and Δy

Regressands	Δm_t	Δp_t	Δy_t
Constant	$\underset{(0.005)}{0.015}$	$\underset{(0.002)}{0.005}$	$\underset{(0.003)}{0.012}$
Δm_t	–	$\underset{(0.042)}{-\ 0.077}$	$\underset{(0.069)}{-\ 0.003}$
Δm_{t-1}	$\underset{(0.094)}{0.361}$	$\underset{(0.060)}{0.039}$	$\underset{(0.068)}{0.108}$
Δp_t	$\underset{(0.236)}{-\ 0.434}$	–	$\underset{(0.165)}{-\ 0.106}$
Δp_{t-1}	$\underset{(0.241)}{0.470}$	$\underset{(0.062)}{0.821}$	$\underset{(0.168)}{-\ 0.226}$
Δy_t	$\underset{(0.147)}{-\ 0.070}$	$\underset{(0.062)}{-\ 0.040}$	–
Δy_{t-1}	$\underset{(0.142)}{0.167}$	$\underset{(0.060)}{0.039}$	$\underset{(0.096)}{-\ 0.176}$
$\widehat{\sigma}$	0.019	0.008	0.013
$\mathsf{F}_{\mathsf{ar}}\,(4, 94)$	4.7^{**}	2.5^{*}	–
$\mathsf{F}_{\mathsf{arch}}\,(4, 90)$	–	4.0^{**}	–
$\chi^2_{\mathsf{nd}}(2)$	–	9.9^{**}	18.3^{**}
$\mathsf{F}_{het}\,(10, 87)$	–	2.1^{*}	–
$\mathsf{F}_{fun}\,(20, 77)$	–	5.2^{**}	–

4.4 Monte Carlo study of nonsense regressions

Using *PcNaive*, set up a Monte Carlo experiment to investigate the nonsense regressions problem when the following DGP holds:

$$\Delta y_t = \alpha + \epsilon_t \;\text{ where }\; \epsilon_t \sim \mathsf{IN}\left[0, \sigma^2_\epsilon\right] \;\text{ and }\; y_0 = 0; \tag{4.5}$$

$$\Delta z_t = \gamma + \nu_t \;\text{ where }\; \nu_t \sim \mathsf{IN}\left[0, \sigma^2_\nu\right] \;\text{ and }\; z_0 = 0;$$

when

$$\mathsf{E}\left[\epsilon_t \nu_s\right] = 0 \;\forall t, s. \tag{4.6}$$

In your experiment, consider $\left(\alpha, \gamma, \sigma^2_\epsilon, \sigma^2_\nu\right)$ in relation to the data being quarterly and in logs (α of 0.01 entails around 5 per cent p.a. growth, and $\sigma_\epsilon = 0.005$ is a 0.5 per cent equation standard error). The investigator postulates and estimates the equation:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 t + u_t \;\text{ where }\; u_t \underset{\widetilde{c}}{} \mathsf{IID}\left[0, \sigma^2_u\right], \tag{4.7}$$

and β_1 is interpreted as the derivative of y_t with respect to z_t ($\underset{\widetilde{c}}{}$ denotes 'is conjectured to be distributed as').

(1) How often is the hypothesis H_0: $\beta_1 = 0$ rejected when the sample size is 40?
(2) Does the rejection frequency change much as the sample increases up to 100?

(3) How do changes in the parameters of the DGP affect the rejection frequency (e.g. does a difference between α and γ matter; is their absolute size important; or their relative magnitudes compared to σ_ϵ and σ_ν respectively; etc.)?
(4) How would you systematically investigate this parameter dependence?
(5) Could the problem of nonsense regressions be detected? If so, briefly describe the behaviour of the statistics that you would use in this task.
(6) List all the features of the experiment that could matter in determining the outcome you obtained (for example, normality, independent errors, etc.). How can this issue be studied systematically?

4.4.1 Rejection frequency of the t-test

The t-ratio for testing H_0: $\beta_1 = 0$ is defined by:

$$\mathsf{t}_{\beta_1=0} = \frac{\widehat{\beta}_1}{\mathsf{ESE}\left[\widehat{\beta}_1\right]}. \tag{4.8}$$

To design a Monte Carlo study to investigate the finite-sample behaviour of this t-ratio, we will first derive its asymptotic properties as a guideline in designing our studies. The analysis draws on (among others) Stock (1987), Phillips (1986, 1987a, 1987b, 1988), Park and Phillips (1988, 1989), Phillips and Perron (1988), Chan and Wei (1988) and Banerjee and Hendry (1992): Banerjee, Dolado, Galbraith and Hendry (1993), Hendry (1995a) and Johansen (1995) provide more extensive treatments of the tools used here. We will obtain the asymptotic properties of numerator and denominator separately, and first derive the properties of $\widehat{\beta}_1$.

To do so, define $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ and $\mathbf{X} = (\boldsymbol{\iota} : \mathbf{z} : \mathbf{w})$ the matrix of regressors, with $\boldsymbol{\iota}$ being a $T \times 1$ vector of ones and $\mathbf{w}' = (1, 2, \ldots, T)$. Then, the OLS estimator of $\boldsymbol{\beta}$ is defined by:

$$\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}. \tag{4.9}$$

From (4.5), we can write:

$$\mathbf{z} = \mathbf{w}\gamma + \mathbf{s}_\nu,$$

where $\mathbf{s}_\nu = (S_{\nu,1}, ...S_{\nu,T})'$ and $S_{\nu,t} = \sum_{i=1}^{t} \nu_i$, so that $\mathbf{X}$ can be written as:

$$\mathbf{X} = (\boldsymbol{\iota} : \mathbf{w}\gamma + \mathbf{s}_\nu : \mathbf{w}) = (\boldsymbol{\iota} : \mathbf{s}_\nu : \mathbf{w}) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \gamma & 1 \end{pmatrix} = \mathbf{X}^*\mathbf{A},$$

where:

$$\mathbf{X}^* = (\boldsymbol{\iota} : \mathbf{s}_\nu : \mathbf{w}) \qquad \text{and} \qquad \mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \gamma & 1 \end{pmatrix}.$$

In addition, from (4.5):

$$\mathbf{y} = \mathbf{w}\alpha + \mathbf{s}_\epsilon = (\boldsymbol{\iota} : \mathbf{s}_\nu : \mathbf{w}) \begin{pmatrix} 0 \\ 0 \\ \alpha \end{pmatrix} + \mathbf{s}_\epsilon = \mathbf{X}^* \boldsymbol{\phi} + \mathbf{s}_\epsilon,$$

where $\boldsymbol{\phi}' = (0, 0, \alpha)$ and $S_{\epsilon,t} = \sum_{i=1}^{t} \epsilon_i$, so that we can write (4.9) as:

$$\mathbf{A}'\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{A}\widehat{\boldsymbol{\beta}} = \mathbf{A}'\mathbf{X}^{*\prime} \left(\mathbf{X}^*\boldsymbol{\phi} + \mathbf{s}_\epsilon\right),$$

and because $\mathbf{A}$ is non-singular $\forall\gamma$:

$$\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi} = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1} \mathbf{X}^{*\prime}\mathbf{s}_\epsilon. \tag{4.10}$$

Notice that:

$$\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \gamma\widehat{\beta}_1 + \widehat{\beta}_2 - \alpha \end{pmatrix}.$$

Let us now find out what the terms on the right-hand side of (4.10) converge to. First, denoting by $\mathbf{C}_T$ and $\mathbf{D}_T$ the matrices:

$$\mathbf{C}_T = \begin{pmatrix} T^{-\frac{1}{2}} & 0 & 0 \\ 0 & T^{-1} & 0 \\ 0 & 0 & T^{-3/2} \end{pmatrix} \text{ and } \mathbf{D}_T = \begin{pmatrix} T^{\frac{1}{2}} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & T^{-\frac{1}{2}} \end{pmatrix},$$

(4.10) is:

$$\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi} = \mathbf{D}_T\mathbf{D}_T^{-1}\mathbf{C}_T \left(\mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{C}_T\right)^{-1} \mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{s}_\epsilon, \tag{4.11}$$

and because $\mathbf{D}_T^{-1}\mathbf{C}_T = T^{-1}\mathbf{I}_3$:

$$\mathbf{D}_T^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right) = \left(\mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{C}_T\right)^{-1} T^{-1}\mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{s}_\epsilon := \mathbf{B}_T^{-1}\mathbf{v}_T, \tag{4.12}$$

where $\mathbf{B}_T = \mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{C}_T$ and $\mathbf{v}_T = T^{-1}\mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{s}_\epsilon$. Next, $\mathbf{X}^{*\prime}\mathbf{X}^*$ is:

$$\begin{pmatrix} \boldsymbol{\iota}' \\ \mathbf{s}_\nu' \\ \mathbf{w}' \end{pmatrix} (\boldsymbol{\iota} : \mathbf{s}_\nu : \mathbf{w}) = \begin{pmatrix} T & \sum_{t=1}^{T} S_{\nu,t} & \frac{T(T+1)}{2} \\ \sum_{t=1}^{T} S_{\nu,t} & \sum_{t=1}^{T} S_{\nu,t}^2 & \sum_{t=1}^{T} tS_{\nu,t} \\ \frac{T(T+1)}{2} & \sum_{t=1}^{T} tS_{\nu,t} & \frac{T(T+1)(2T+1)}{6} \end{pmatrix},$$

so that:

$$\mathbf{B}_T = \begin{pmatrix} 1 & T^{-3/2}\sum_{t=1}^{T} S_{\nu t} & \frac{T+1}{2T} \\ T^{-3/2}\sum_{t=1}^{T} S_{\nu t} & T^{-2}\sum_{t=1}^{T} S_{\nu t}^2 & T^{-5/2}\sum_{t=1}^{T} tS_{\nu t} \\ \frac{T+1}{2T} & T^{-5/2}\sum_{t=1}^{T} tS_{\nu t} & \frac{(T+1)(2T+1)}{6T^2} \end{pmatrix}$$
$$\Rightarrow \begin{pmatrix} 1 & \sigma_\nu \int B_2(r)\,\mathrm{d}r & \frac{1}{2} \\ \sigma_\nu \int B_2(r)\,\mathrm{d}r & \sigma_\nu^2 \int B_2^2(r)\,\mathrm{d}r & \sigma_\nu \int rB_2(r)\,\mathrm{d}r \\ \frac{1}{2} & \sigma_\nu \int rB_2(r)\,\mathrm{d}r & \frac{1}{3} \end{pmatrix} := \mathbf{B},$$

where the integrals are between zero and unity, and:

$$\mathbf{v}_T = \begin{pmatrix} T^{-3/2}\sum_{t=1}^{T} S_{\epsilon,t} \\ T^{-2}\sum_{t=1}^{T} S_{\nu,t}S_{\epsilon,t} \\ T^{-5/2}\sum_{t=1}^{T} tS_{\epsilon,t} \end{pmatrix} \Rightarrow \sigma_\epsilon \begin{pmatrix} \int B_1(r)\,\mathrm{d}r \\ \sigma_\nu \int B_1(r)\,B_2(r)\,\mathrm{d}r \\ \int rB_1(r)\,\mathrm{d}r \end{pmatrix} := \mathbf{v}. \tag{4.13}$$

So, by the continuous mapping theorem:

$$\mathbf{D}_T^{-1}\left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right) = \begin{pmatrix} T^{-\frac{1}{2}}\widehat{\beta}_0 \\ \widehat{\beta}_1 \\ T^{\frac{1}{2}}\left(\gamma\widehat{\beta}_1 + \widehat{\beta}_2 - \alpha\right) \end{pmatrix} \Rightarrow \mathbf{B}^{-1}\mathbf{v}. \tag{4.14}$$

To derive $\widehat{\beta}_1$ from (4.14), notice that, apart from the scalar factor $\sigma_\nu^2 \det \mathbf{B}^{-1}$ then $\mathbf{B}^{-1}$ is (dropping the arguments and the differential elements from the integrals):

$$\begin{pmatrix} \frac{1}{3}\int B_2^2 - \left(\int rB_2\right)^2 & \frac{\frac{1}{2}\int rB_2 - \frac{1}{3}\int B_2}{\sigma_\nu} & \left(\int B_2\right)\left(\int rB_2\right) - \frac{1}{2}\int B_2^2 \\ \frac{\frac{1}{2}\int rB_2 - \frac{1}{3}\int B_2}{\sigma_\nu} & \frac{1}{12\sigma_\nu^2} & \frac{\frac{1}{2}\int B_2 - \int rB_2}{\sigma_\nu} \\ \left(\int B_2\right)\left(\int rB_2\right) - \frac{1}{2}\int B_2^2 & \frac{\frac{1}{2}\int B_2 - \int rB_2}{\sigma_\nu} & \int B_2^2 - \left(\int B_2\right)^2 \end{pmatrix} \tag{4.15}$$

with $\det \mathbf{B}$ given by:

$$\frac{\sigma_\nu^2}{12}\left[\int B_2^2 + 12\left(\int B_2\right)\left(\int rB_2\right) - 12\left(\int rB_2\right)^2 - 4\left(\int B_2\right)^2\right] = \sigma_\nu^2 h, \tag{4.16}$$

in an obvious notation. So, from (4.13)–(4.16) the second element in $\mathbf{B}^{-1}\mathbf{v}$ is:

$$\begin{aligned}\widehat{\beta}_1 \quad &\Rightarrow \quad \frac{\sigma_\epsilon}{\sigma_\nu h}\left(\left[\frac{1}{2}\int rB_2 - \frac{1}{3}\int B_2\right]\int B_1 + \frac{1}{12}\int B_1 B_2 \right. \\ &\qquad \left. + \left[\frac{1}{2}\int B_2 - \int rB_2\right]\int rB_1\right) \\ &= \quad \frac{\sigma_\epsilon}{\sigma_\nu h} g, \end{aligned} \tag{4.17}$$

again simplifying notation. Three main features are to be noticed from this result: (i) the usual rescaling factor $\sqrt{T}$ is not required; (ii) $\widehat{\beta}_1$ does not converge to a value of zero, corresponding to the absence of any causal link; and (iii) its asymptotic distribution is not a function of either α or γ.

Next, consider the denominator of the t-ratio in (4.8), namely:

$$\mathsf{ESE}\left[\widehat{\beta}_1\right] = \sqrt{\widehat{\sigma}_u^2 a_{22}},$$

where $\widehat{\sigma}_u^2 = \widehat{\mathbf{u}}'\widehat{\mathbf{u}}/(T-3)$, with $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$, and a_{22} denotes element $(2,2)$ of $(\mathbf{X}'\mathbf{X})^{-1}$. We analyze terms one at a time. First:

$$\widehat{\mathbf{u}} = \mathbf{X}^*\boldsymbol{\phi} + \mathbf{s}_\epsilon - \mathbf{X}^*\mathbf{A}\widehat{\boldsymbol{\beta}} = -\mathbf{X}^*\left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right) + \mathbf{s}_\epsilon,$$

so that:

$$\begin{aligned}\widehat{\mathbf{u}}'\widehat{\mathbf{u}} \quad &= \quad \left[-\mathbf{X}^*\left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right) + \mathbf{s}_\epsilon\right]'\left[-\mathbf{X}^*\left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right) + \mathbf{s}_\epsilon\right] \\ &= \quad \left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right)'\mathbf{X}^{*\prime}\mathbf{X}^*\left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right) - 2\left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right)'\mathbf{X}^{*\prime}\mathbf{s}_\epsilon + \mathbf{s}_\epsilon'\mathbf{s}_\epsilon \\ &= \quad -\left(\mathbf{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\phi}\right)'\mathbf{X}^{*\prime}\mathbf{s}_\epsilon + \mathbf{s}_\epsilon'\mathbf{s}_\epsilon.\end{aligned}$$

Thus, from (4.11) and (4.14):

$$\begin{aligned}T^{-2}\widehat{\mathbf{u}}'\widehat{\mathbf{u}} \quad &= \quad -T^{-2}\mathbf{s}_\epsilon'\mathbf{X}^*\mathbf{C}_T\mathbf{B}_T^{-1}\mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{s}_\epsilon + T^{-2}\mathbf{s}_\epsilon'\mathbf{s}_\epsilon \\ &\Rightarrow \quad -\mathbf{v}'\mathbf{B}^{-1}\mathbf{v} + \sigma_\epsilon^2\int B_1^2,\end{aligned}$$

but (from direct, albeit tedious, multiplication):

$$\begin{aligned}\mathbf{v}'\mathbf{B}^{-1}\mathbf{v} \quad &= \quad \sigma_\epsilon^2 h^{-1}\left(\left[\frac{1}{3}\int B_2^2 - \left(\int rB_2\right)^2\right]\left[\int B_1\right]^2\right. \\ &\qquad + \left[\int rB_2 - \frac{2}{3}\int B_2\right]\int B_1B_2\int B_1 \\ &\qquad + \left[2\int B_2\int rB_2 - \int B_2^2\right]\int B_1\int rB_1\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{12}\left[\int B_1 B_2\right]^2 - \left[2\int rB_2 - \int B_2\right]\int rB_1 \int B_1 B_2 \\
&+\left[\int B_2^2 - \left(\int B_2\right)^2\right]\left[\int rB_1\right]^2\Bigg) \\
= \;& \sigma_\epsilon^2 h^{-1} f,
\end{aligned}
$$

(say) and so:

$$T^{-2}\widehat{\mathbf{u}}'\widehat{\mathbf{u}} \Rightarrow \sigma_\epsilon^2 \left(\int B_1^2 - h^{-1} f\right).$$

To find out what a_{22} converges to, write:

$$
\begin{aligned}
\left(\mathbf{X}'\mathbf{X}\right)^{-1} &= \mathbf{A}^{-1}\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\left(\mathbf{A}^{-1}\right)' \\
&= \mathbf{A}^{-1}\mathbf{C}_T\left(\mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{C}_T\right)^{-1}\mathbf{C}_T\left(\mathbf{A}^{-1}\right)',
\end{aligned}
$$

where:

$$\mathbf{C}_T^{-1}\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}'\mathbf{C}_T^{-1} = \left(\mathbf{C}_T\mathbf{X}^{*\prime}\mathbf{X}^*\mathbf{C}_T\right)^{-1} \Rightarrow \mathbf{B}^{-1}.$$

But, from the left-hand side of this expression, $T^2 a_{22}$ is element $(2,2)$ of $\mathbf{C}_T^{-1}\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}'\mathbf{C}_T^{-1}$, so it must converge to element $(2,2)$ of $\mathbf{B}^{-1}$:

$$T^2 a_{22} \Rightarrow \frac{1}{12h\sigma_\nu^2}.$$

Thus:

$$
\begin{aligned}
(T-3)^{\frac{1}{2}}\,\mathsf{ESE}\left[\widehat{\beta}_1\right] &= \sqrt{\left(T^{-2}\widehat{\mathbf{u}}'\widehat{\mathbf{u}}\right)\left(T^2 a_{22}\right)} \\
&\Rightarrow \sigma_\epsilon \sigma_v^{-1}\left[\frac{1}{12h}\left(\int B_1^2 - h^{-1} f\right)\right]^{\frac{1}{2}}.
\end{aligned}
\tag{4.18}
$$

Finally, the t-ratio converges to:

$$
\begin{aligned}
(T-3)^{-\frac{1}{2}}\,\mathsf{t}_{\beta_1=0} &= \frac{\widehat{\beta}_1}{(T-3)^{\frac{1}{2}}\,\mathsf{ESE}\left[\widehat{\beta}_1\right]} \\
&\Rightarrow \frac{g}{h}\left[\frac{1}{12h}\left(\int B_1^2 - h^{-1} f\right)\right]^{-\frac{1}{2}}
\end{aligned}
\tag{4.19}
$$

showing that the re-scaled t-ratio converges to a quantity which is not a function of α, γ, σ_ϵ^2 or σ_ν^2. Moreover, to obtain a limiting distribution, we had to normalize $\mathsf{t}_{\beta_1=0}$ by $(T-3)^{-1/2}$, so the un-normalized t-ratio diverges, making it increasingly likely that the null will be rejected as the sample size grows, despite the absence of any actual link between y and z, and implicit detrending (see Frisch and Waugh, 1933). Having derived this useful asymptotic result, and noticing from (4.16) that $\widehat{\beta}_1$ is not a function of either α or γ, we proceed to design a Monte Carlo study.

We consider a partial Monte Carlo study with only three experiments to get a first impression on how the properties of the t-ratio may vary in finite samples with the parameters in the DGP. The values assigned to the parameters are provided in table 4.6 below. We conducted a recursive Monte Carlo study with $M = 10,000$ replications for $T = [40, 100]$ to be able to analyze the effects of sample size on the rejection frequency.

Table 4.6 Monte Carlo design parameter values

Experiment	α	γ	σ_ϵ^2	σ_ν^2
1	0.0100	0.0025	0.005	0.001
2	0.0025	0.0100	0.005	0.001
3	0.0100	0.0025	0.001	0.005

Interpreting the data as quarterly, the values of α and γ imply 4% and 1% p.a. growth rates for y and z respectively. In the first two experiments, only α and γ change, whereas the third experiment exchanges the values of σ_ϵ^2 and σ_ν^2 relative to the first experiment. We used the same random number seed throughout to highlight the invariances predicted by the asymptotic theory.

Figure 4.5a–d shows the means of $\widehat{\beta}_1$ in each experiment for $T = 40, \ldots, 100$, the mean estimated standard errors of $\widehat{\beta}_1$, the mean values of the t-ratio, and the rejection frequencies of the t-test of H_0: $\beta_1 = 0$. From fig. 4.5a, the mean of $\widehat{\beta}_1$ is unaffected by altering α and γ, but changes with the other parameters in the DGP for experiments 1 and 3. This is simply the scaling factor σ_ϵ/σ_v in (4.16), and is precisely as predicted by the asymptotic results: as experiments 1 and 3 differ by a factor of 5 in their values of σ_ϵ/σ_v, the units of z are altered, and the two graphs are in fact identical when their means and scales are matched (the legends have been left to show that all three biases were plotted). Notice the tiny units: the mean biases at $T = 100$ are 0.004 (0.0095) in experiments 1 and 2, and 0.00084 (0.0019) in the third, so neither bias is significantly different from zero at the 5% level and the ratios are precisely 5.

The estimated standard errors of $\widehat{\beta}_1$ plotted in fig. 4.5b change with the relative values of σ_ϵ and σ_ν, and drop precisely in line with $1/\sqrt{T}$, both exactly as suggested by the asymptotic theory in (4.18): those for the first two experiments are identical, as are all three if matched for mean and scale. The mean t-values in fig. 4.5c are identical for all the parameters in the DGP in agreement with asymptotic theory, despite both numerator and denominator changing. Finally, the rejection frequencies of the t-test are identical in the three cases (again, the legend shows three lines were drawn).

There are several ways to understand the high rejection frequency of the correct null. The earlier asymptotic theory provides a formalization, but a some intuition is provided by fig. 4.6 which records one of the biases with ± 2MCSD and ± 2ESE. As

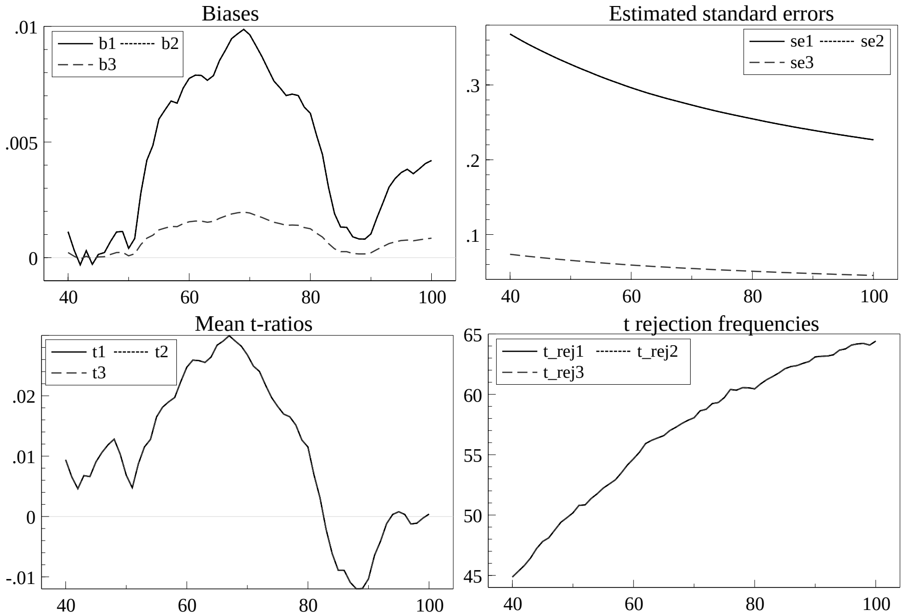

Figure 4.5 Results for the three Monte Carlo experiments

can be seen, the latter greatly under-estimates the former due to the neglected positive residual serial correlation. Thus, the wrong denominator is being used in the t-test, inflating its value.

From fig. 4.5c, mean t-ratios are found not to diverge in these experiments, and in fact are very close to zero: this matches the mean value of g being almost zero here. However, fig. 4.5d entails that the null is rejected from 45% to 65% of the time, rising with T. Thus, the t-values must have large standard errors, confirmed by fig. 4.7 which shows the mean t-values ± 2SE for experiment 1. As Student's t has a standard error close to unity, and values outside ± 2 are usually judged 'significant', the region of ± 15 covered by t $\pm$ 2SE by the end of the sample is very large, leading to a high rejection frequency. Consequently, the excess rejection is not due to a 'substantive' relation being found where none exists, but to the statistic examined having a distribution that is dramatically different from that assumed. Indeed, under the null that $\beta_1 = 0$, (4.7) entails that $\{y_t\}$ is white noise around a trend, contradicting the fact that it is a random walk, and hence highly autoregressive. Before conducting inference, it is advisable to check that the assumptions of a test are not at odds with the empirical evidence.

4.4.2 Rejection frequency of t-test as the sample size increases

The graphs in fig. 4.5d showed that rejection of the correct null becomes more likely as the sample size increases, reaching a maximum of about 65% when $T = 100$. This is slightly against ones intuition that the lack of connection between these 'unrelated'

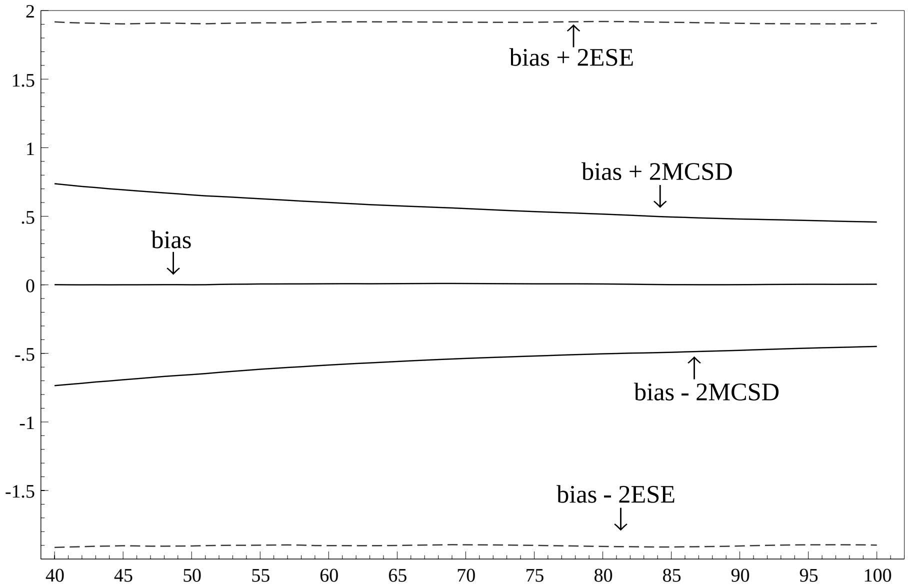

Figure 4.6 Mean biases with bias$\pm 2MCSD$ and bias$\pm 2SE$

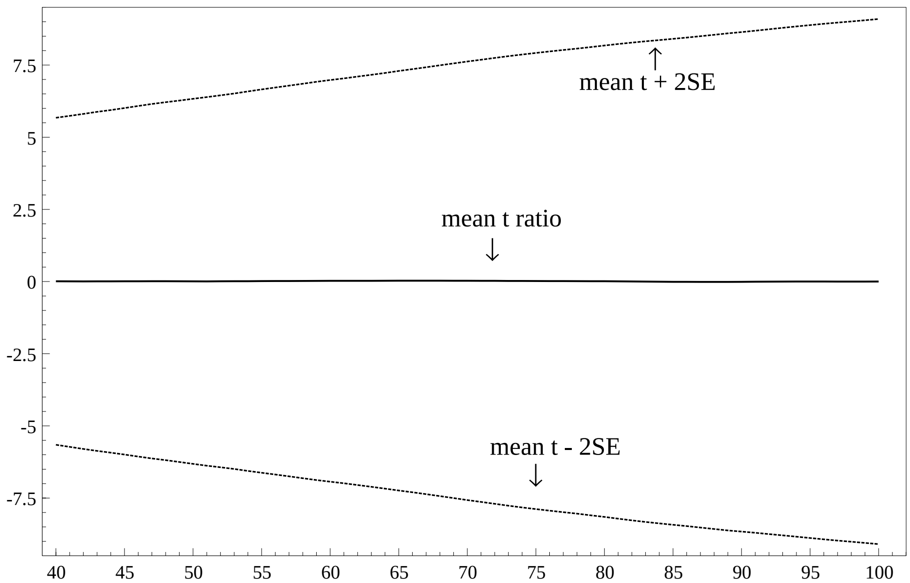

Figure 4.7 t-ratios for experiment 1 with $\bar{t} \pm 2SE$

series should be more easily detected in larger samples. The converse holds for the reason shown in (4.19), in that t needs to be scaled by $1/\sqrt{T}$ to converge in distribution.

However, rejection does not increase very rapidly with sample size. Moreover, in a certain sense, the series share common stochastic and deterministic trends (the mean value of $\widehat{\beta}_2$ is always highly significantly different from zero).

4.4.3 Effect of changing parameters

This question has been answered in §4.4.1: most features are invariant to the design variables of the DGP, but vary with T, and change in scale as the error variances alter.

4.4.4 Systematic parameter dependence

A systematic analysis of parameter dependence requires designing a serious Monte Carlo study, and the construction of response surfaces to help reduce specificity, and to summarize the results. Nevertheless, all the predictions of the analysis have been borne out so far. Thus, using common random numbers as an inter-experiment control has proved highly successful.

4.4.5 Detecting nonsense regressions

Large R^2s and small values of *DW* are common features of nonsense regressions. Figure 4.6 shows the mean $\widehat{\beta}_1 \pm 2\mathsf{MCSD}$, and $\widehat{\beta}_1 \pm 2\mathsf{ESE}$. Neither the mean of $\widehat{\beta}_1$, nor its ESE, change much with the sample size. In addition, as noted above, the MCSD is much larger than the ESE, illustrating that ESE may be underestimating $\widehat{\beta}_1$'s finite-sample variance. Finally, table 4.7 shows results from the experiments at the sample size $T = 100$, revealing large R^2s and small *DW*s.

Table 4.7 Monte Carlo values of t, R^2 and *DW*

Experiment	Mean t	t rejection	Mean R^2	Mean *DW*	*DW* reject
				$\rho = 0$	$\rho = 0$
1	-0.0015	63.73	0.69	0.26	100
2	0.0327	64.78	0.53	0.26	100
3	-0.031	64.01	0.91	0.26	100

4.4.6 Dependence of results on assumptions

None of the simulation results contradicted the analytical predictions; but that was because the DGP matched that in the Monte Carlo. However, the asymptotic findings would be unaltered by inducing autocorrelated errors on the DGP equations, yet this could alter the finite-sample outcomes. Indeed, Banerjee, Dolado, Hendry and Smith (1986) found a low correspondence between asymptotics and finite samples when the errors on their models were highly autocorrelated, so the close match above may depend on the assumption of white-noise errors. When limiting distributions are invariant to a wide range of alternative properties of the DGP, one should expect the match of asymptotic to finite-sample results to change from good to bad at different points in the parameter space.

4.5 Expectations processes

(1) In the second-order autoregressive process defined by:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \text{ where } \epsilon_t \sim \mathsf{IN}\left[0, \sigma_\epsilon^2\right], \tag{4.20}$$

derive the conditions on (ϕ_1, ϕ_2) under which (4.20) is stationary.

(2) Consider the following data-generation process for inflation:

$$\begin{aligned} i_t &= \rho + \mathsf{E}_t\left[\pi_{t+1}\right] + v_t \\ \pi_t &= \mu_0 + \mu_1 \pi_{t-1} + \mu_2 \pi_{t-2} + e_t, \end{aligned} \tag{4.21}$$

where i_t is the nominal interest rate, ρ is the real interest rate, π_t is the rate of inflation, $\mathsf{E}_t[\pi_{t+1}]$ is the conditional expectation of π_{t+1} given information available at time t, and v_t and e_t are mean-zero, independent white-noise processes, such that $\mathsf{E}[v_t e_s] = 0\ \forall t, s$. Under the assumptions of rational expectations and that the process generating inflation is stationary, by estimating the regression of i_t on a constant, π_{t-1} and π_{t-2}, is it possible to test the Fisher hypothesis that the long-run relationship between the nominal interest rate and inflation is H_0: $i^* = \rho + \pi^*$, where * denotes the unconditional expectation (steady-state value)?

(3) Show that a test of the sum of the inflation coefficients being unity in that regression is biased against finding evidence in favour of the Fisher hypothesis even though $\mathsf{E}[\pi_t v_s] = 0\ \forall t, s$.

(4) Given your answer to §4.5.3, and the stationarity of (4.21), will a relaxation of stationarity remove that bias? Comment on your findings, especially the causes and testability of the bias.

(5) Suggest an estimation procedure that allows a valid test.

(Oxford M.Phil., 1991)

4.5.1 Stationarity conditions

To find conditions under which (4.20) is stationary, write it in terms of the lag operator L as:

$$y_t = \left(1 - \phi_1 L - \phi_2 L^2\right)^{-1} \epsilon_t.$$

Then (4.20) is stationary if the roots of the following polynomial in Z:

$$Z^2 - \phi_1 Z - \phi_2 = 0,$$

are inside the unit circle; namely, if:

$$|\xi| = \frac{1}{2}\left|\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}\right| < 1. \tag{4.22}$$

Denoting those roots by ξ_i $(i = 1, 2)$, when $\xi_1 \neq \xi_2$ (4.20) is:

$$\begin{aligned}
y_t &= (1 - \xi_1 L)^{-1} (1 - \xi_2 L)^{-1} \epsilon_t \\
&= \frac{1}{\xi_1 - \xi_2} \left(\frac{1}{1 - \xi_1 L} - \frac{1}{1 - \xi_2 L} \right) \epsilon_{t+1}.
\end{aligned}$$

so that the mean is:

$$\mathsf{E}[y_t] = \frac{1}{\xi_1 - \xi_2} \sum_{j=0}^{\infty} \left(\xi_1^j - \xi_2^j \right) \mathsf{E}[\epsilon_{t-j+1}] = 0,$$

and the covariance $\mathsf{C}[y_t, y_{t-\tau}]$ of order τ is:

$$\begin{aligned}
& \left(\frac{1}{\xi_1 - \xi_2} \right)^2 \mathsf{E}\left[\sum_{j,k=0}^{\infty} \left(\xi_1^j - \xi_2^j \right) \left(\xi_1^k - \xi_2^k \right) \epsilon_{t-j+1} \epsilon_{t-k+1-\tau} \right] \\
= & \left(\frac{1}{\xi_1 - \xi_2} \right)^2 \sum_{j,k=0}^{\infty} \left(\xi_1^j - \xi_2^j \right) \left(\xi_1^k - \xi_2^k \right) \mathsf{E}[\epsilon_{t-j+1} \epsilon_{t-k+1-\tau}] \\
= & \left(\frac{1}{\xi_1 - \xi_2} \right)^2 \sigma_\epsilon^2 \sum_{k=0}^{\infty} \left(\xi_1^{k+\tau} - \xi_2^{k+\tau} \right) \left(\xi_1^k - \xi_2^k \right) \\
= & \left(\frac{1}{\xi_1 - \xi_2} \right)^2 \sigma_\epsilon^2 \sum_{k=0}^{\infty} \left(\xi_1^{2k+\tau} - \xi_1^{k+\tau} \xi_2^k - \xi_2^{k+\tau} \xi_1^k + \xi_2^{2k+\tau} \right) \\
= & \left(\frac{1}{\xi_1 - \xi_2} \right)^2 \sigma_\epsilon^2 \left(\frac{\xi_1^\tau}{1 - \xi_1^2} - \frac{\xi_1^\tau}{1 - \xi_1 \xi_2} - \frac{\xi_2^\tau}{1 - \xi_1 \xi_2} + \frac{\xi_2^\tau}{1 - \xi_2^2} \right) \\
= & \frac{\sigma_\epsilon^2}{(\xi_1 - \xi_2)(1 - \xi_1 \xi_2)} \left(\frac{\xi_1^\tau}{1 - \xi_1^2} - \frac{\xi_2^\tau}{1 - \xi_2^2} \right),
\end{aligned}$$

which are both independent of time. Let us now obtain the stationarity conditions in terms of the ϕs. From (4.22):

$$\xi_1 + \xi_2 = \frac{1}{2}\left(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2} \right) + \frac{1}{2}\left(\phi_1 - \sqrt{\phi_1^2 + 4\phi_2} \right) = \phi_1,$$

and:

$$\xi_1 \xi_2 = \frac{1}{4}\left(\phi_1 + \sqrt{\phi_1^2 + 4\phi_2} \right)\left(\phi_1 - \sqrt{\phi_1^2 + 4\phi_2} \right) = -\phi_2,$$

providing us with the following stationarity conditions:

$$1 > |\xi_1| \, |\xi_2| = |\xi_1 \xi_2| = |-\phi_2| = |\phi_2|,$$

and:

$$0 < (1 - \xi_1)(1 - \xi_2) = 1 - (\xi_1 + \xi_2) + \xi_1\xi_2 = 1 - \phi_1 - \phi_2,$$

so that the stationarity conditions in terms of the original parameters are $|\phi_2| < 1$ and $\phi_1 + \phi_2 < 1$.

4.5.2 Testing Fisher's hypothesis

Let us first find out what testing Fisher's hypothesis means under stationarity. To do so, we derive the steady-state solution of the interest-rate equation. From the DGP:

$$\mathsf{E}_t[\pi_{t+1}] = \mu_0 + \mu_1\pi_t + \mu_2\pi_{t-1} \tag{4.23}$$

so that substituting in the interest-rate equation:

$$i_t = \rho + \mu_0 + \mu_1\pi_t + \mu_2\pi_{t-1} + v_t \tag{4.24}$$

and because the steady-state solution for a stationary process is defined as its unconditional expected value (see Hendry, 1995a, p. 213), its long-run solution is:

$$\mathsf{E}[i_t] = \rho + \mu_0 + \mu_1\mathsf{E}[\pi_t] + \mu_2\mathsf{E}[\pi_{t-1}] := i^*. \tag{4.25}$$

But, (4.21) implies that:

$$\mathsf{E}[\pi_t] = \mu_0 + \mu_1\mathsf{E}[\pi_{t-1}] + \mu_2\mathsf{E}[\pi_{t-2}] + \mathsf{E}[e_t] := \pi^*,$$

so that under stationarity:

$$\pi^* = \frac{\mu_0}{1 - \mu_1 - \mu_2},$$

for all t. Thus, (4.25) is:

$$i^* = \rho + \mu_0 + (\mu_1 + \mu_2)\pi^* = \rho + \mu_0 + \frac{(\mu_1 + \mu_2)\mu_0}{1 - \mu_1 - \mu_2} = \rho + \pi^*,$$

so that Fisher's hypothesis is just the long run of (4.21) under stationarity.

The question is whether we can test for this hypothesis from the following model of interest rates:

$$i_t = \alpha_0 + \alpha_1\pi_{t-1} + \alpha_2\pi_{t-2} + \eta_t. \tag{4.26}$$

To answer this question, notice that the long-run solution of (4.26) is:

$$i^* = \alpha_0 + (\alpha_1 + \alpha_2)\pi^*,$$

so that we may think of testing Fisher's hypothesis by testing H_0: $\alpha_1 + \alpha_2 = 1$ in (4.26). However, from the DGP in (4.21b) and (4.23):

$$\begin{aligned} i_t &= \rho + \mu_0 + \mu_1(\mu_0 + \mu_1\pi_{t-1} + \mu_2\pi_{t-2} + e_t) + \mu_2\pi_{t-1} + v_t \\ &= \rho + \mu_0(1 + \mu_1) + \left(\mu_1^2 + \mu_2\right)\pi_{t-1} + \mu_1\mu_2\pi_{t-2} + (\mu_1 e_t + v_t), \end{aligned}$$

so that $\alpha_0 = \rho + \mu_0(1+\mu_1)$, $\alpha_1 = \mu_1^2 + \mu_2$, $\alpha_2 = \mu_1\mu_2$ and $\eta_t = \mu_1 e_t + v_t$, and hence, although OLS is valid in equation (4.26):

$$\alpha_1 + \alpha_2 = \mu_1^2 + \mu_2 + \mu_1\mu_2 = \mu_1(\mu_1 + \mu_2) + \mu_2,$$

and because stationarity implies $\mu_1 + \mu_2 < 1$ then:

$$\alpha_1 + \alpha_2 < \mu_1 + \mu_2 < 1,$$

so that, except for sampling errors, the hypothesis H_0: $\alpha_1 + \alpha_2 = 1$ will be rejected.

4.5.3 Biased results against Fisher's hypothesis

This is answered in §4.5.2.

4.5.4 Fisher's hypothesis relaxing stationarity assumptions

If $\mu_1 + \mu_2 = 1$, then (4.21) is non-stationary, and i_t and π_t cointegrate with cointegrating vector $\boldsymbol{\beta}' = (1 : -1)$ because we can write (4.21) as:

$$\begin{aligned}\begin{pmatrix} \Delta i_t \\ \Delta \pi_t \end{pmatrix} &= \begin{pmatrix} \rho + \mu_0(1+\mu_1) \\ \mu_0 \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} i_{t-1} \\ \pi_{t-1} \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & -\mu_1\mu_2 \\ 0 & -\mu_2 \end{pmatrix} \begin{pmatrix} \Delta i_{t-1} \\ \Delta \pi_{t-1} \end{pmatrix} + \begin{pmatrix} \mu_1 e_t + v_t \\ v_t \end{pmatrix} \end{aligned} \quad (4.27)$$

so that the long-run term coefficient is now unity, and appropriate estimation will remove the bias. Hence, for (4.21), stationarity and the Fisher hypothesis are incompatible.

4.5.5 A valid estimator

When the data are non-stationary, as seems likely in practice, OLS applied to the first equation in (4.27) with cointegration imposed (i.e., using $\pi_{t-1} - i_{t-1}$ as a regressor) is consistent, so that a Wald test on the coefficient of that long-run term has the expected properties. When the data are stationary, from (4.21):

$$i_t = \rho + \lambda\pi_{t+1} + v_t - e_{t+1}$$

where $\lambda = 1$, which can be estimated using π_t and π_{t-1} as instruments from:

$$\pi_{t+1} = \mu_0 + \mu_1\pi_t + \mu_2\pi_{t-1} + e_{t+1},$$

and a Wald test of $\lambda = 1$ constructed.

4.6 Asymptotic properties of OLS in nonsense regressions

(1) Consider the data generation process:

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t \\ z_t &= z_{t-1} + \nu_t \end{aligned} \tag{4.28}$$

where $\mathsf{E}[\epsilon_t \nu_s] = 0 \; \forall t, s$. Let $W(r)$ and $V(r)$ be independent Wiener processes for $r \in [0, 1]$, and let $\Rightarrow$ denote weak convergence, then you may use without proof the following results:

$$\begin{aligned} T^{-2} \sum_{t=1}^{T} y_t^2 &\Rightarrow \sigma_\epsilon^2 \int_0^1 W(r)^2 \mathrm{d}r \\ T^{-2} \sum_{t=1}^{T} z_t^2 &\Rightarrow \sigma_\nu^2 \int_0^1 V(r)^2 \mathrm{d}r \\ T^{-2} \sum_{t=1}^{T} y_t z_t &\Rightarrow \sigma_\epsilon \sigma_\nu \int_0^1 W(r) V(r) \mathrm{d}r \end{aligned} \tag{4.29}$$

The model conjectured by an investigator is:

$$y_t = \beta z_t + \zeta_t \;\text{ with }\; \zeta_t \underset{c}{\sim} \mathsf{IN}\left[0, \sigma_\zeta^2\right], \tag{4.30}$$

where $\underset{c}{\sim}$ denotes 'is claimed to be distributed as'. A sample of size T is available on $\{y_t; z_t\}$. Derive the limiting distribution of the least-squares estimator of β:

$$\widehat{\beta} = \left(\sum_{t=1}^{T} z_t^2 \right)^{-1} \sum_{t=1}^{T} y_t z_t. \tag{4.31}$$

Explain any theorems used and discuss the implications of the result.

(2) Derive the limiting distribution of the t-test of H_0: $\beta = 0$ when (4.28) is true and:

$$\mathsf{t}_{\beta=0} = \widehat{\beta} \left[\frac{\widehat{\sigma}_\zeta^2}{\sum_{t=1}^{T} z_t^2} \right]^{-\frac{1}{2}},$$

where $\widehat{\sigma}_\zeta^2$ is the residual variance from (4.30). Describe the implications of this result. What happens to the sample correlation coefficient of y with z as $T \to \infty$?

(3) Comment on the implications of the analysis for the problem of 'nonsense regressions' in time-series econometrics. Discuss how to extend these distributional findings to test for the existence of cointegration between y and z.

(Oxford M.Phil., 1992)

4.6.1 Limiting distribution of $\widehat{\beta}$

From (4.29) and because the inverse is a continuous functional, then the OLS estimator of β in (4.31) converges to:

$$\widehat{\beta} \Rightarrow \frac{\sigma_\epsilon}{\sigma_\nu} \left(\int_0^1 V(r)^2 \, \mathrm{d}r \right)^{-1} \left(\int_0^1 W(r) \, V(r) \mathrm{d}r \right). \tag{4.32}$$

This result is obtained under weak conditions on the ϵs and the νs, namely, zero means at all points in time, existence of a moment of a higher order than second, expected values of sample variances converging to a finite quantity, and both $\{\epsilon_t\}$ and $\{\nu_t\}$ being mixing processes. The result itself has implications for modelling. Because of (4.28), y and z are unrelated, yet (4.32) entails that $\widehat{\beta}$ does not converge to its 'true' value of zero. Hence, regression (4.30) will, in general, provide misleading information because its coefficient is likely to appear significantly different from zero, incorrectly suggesting that y and z are related.

4.6.2 Limiting distribution of the t-ratio

The t-ratio is defined by:

$$\mathsf{t}_{\beta=0} = \frac{\widehat{\beta}}{\mathsf{ESE}\left[\widehat{\beta}\right]},$$

where:

$$\mathsf{ESE}\left[\widehat{\beta}\right] = \widehat{\sigma}_\zeta \left(\sum_{t=1}^{T} z_t^2 \right)^{-\frac{1}{2}} \quad \text{with} \quad \widehat{\sigma}_\zeta^2 = \frac{1}{T-1} \sum_{t=1}^{T} \widehat{\zeta}_t^2,$$

and $\widehat{\zeta}_t = y_t - z_t \widehat{\beta}$. To see what the denominator of the t-ratio converges to, write:

$$\widehat{\zeta}_t = y_t - z_t \widehat{\beta} = y_t - z_t \left(\sum_{t=1}^{T} z_t^2 \right)^{-1} \left(\sum_{t=1}^{T} z_t y_t \right),$$

so that:

$$\begin{aligned} T^{-2} \sum_{t=1}^{T} \widehat{\zeta}_t^2 &= T^{-2} \sum_{t=1}^{T} y_t^2 - \left(T^{-2} \sum_{t=1}^{T} z_t^2 \right)^{-1} \left(T^{-2} \sum_{t=1}^{T} z_t y_t \right)^2 \qquad (4.33) \\ &\Rightarrow \sigma_\epsilon^2 \left(\int_0^1 W(r)^2 \, \mathrm{d}r - \left[\int_0^1 V(r)^2 \mathrm{d}r \right]^{-1} \left[\int_0^1 W(r) \, V(r) \mathrm{d}r \right]^2 \right), \end{aligned}$$

and hence $(T-1)^{1/2} \mathsf{ESE}[\widehat{\beta}]$ converges in distribution to:

$$\left(\sigma_\nu^2 \int_0^1 V(r)^2 \mathrm{d}r \right)^{-\frac{1}{2}} \sigma_\epsilon \left(\int_0^1 W(r)^2 \, \mathrm{d}r - \left[\int_0^1 V(r)^2 \mathrm{d}r \right]^{-1} \left[\int_0^1 W(r) \, V(r) \mathrm{d}r \right]^2 \right)^{\frac{1}{2}}$$

$$= \frac{\sigma_\epsilon \left(\int_0^1 W(r)^2 \mathrm{d}r \int_0^1 V(r)^2 \mathrm{d}r - \left[\int_0^1 W(r) V(r) \mathrm{d}r \right]^2 \right)^{\frac{1}{2}}}{\sigma_\nu \int_0^1 V(r)^2 \mathrm{d}r}. \tag{4.34}$$

Thus, from (4.32), (4.33) and (4.29b), the appropriately-scaled t-ratio converges to:

$$(T-1)^{-\frac{1}{2}} \, \mathsf{t}_{\beta=0} \Rightarrow \frac{\mathsf{R}}{(1-\mathsf{R}^2)^{\frac{1}{2}}},$$

where:

$$\mathsf{R} = \frac{\int_0^1 W(r) V(r) \mathrm{d}r}{\left(\int_0^1 W(r)^2 \mathrm{d}r \right)^{\frac{1}{2}} \left(\int_0^1 V(r)^2 \mathrm{d}r \right)^{\frac{1}{2}}}.$$

This result implies that if y and z are generated by unrelated $\mathsf{I}(1)$ processes, the t-ratio for testing the existence of a relationship between them diverges as the sample size increases: even if the variables are unrelated, we can find apparent evidence to the contrary by applying a t-test.

In addition, information provided by the correlation coefficient of the two variables will also be misleading, because this coefficient does not converge to zero (as it would in a stationary case for unrelated variables), but to a random variable. To see that, from (4.29), the sample correlation coefficient between y and z is:

$$\mathsf{R}_T = \frac{T^{-2} \sum_{t=1}^{T} y_t z_t}{\left(T^{-2} \sum_{t=1}^{T} y_t^2 \right)^{\frac{1}{2}} \left(T^{-2} \sum_{t=1}^{T} z_t^2 \right)^{\frac{1}{2}}} \Rightarrow \mathsf{R}.$$

4.6.3 Implications for the analysis of nonsense regressions and cointegration

The implications for analyzing nonsense regressions have already been discussed in §4.6.1 and §4.6.2 above, so we will only consider cointegration. To test for cointegration of y_t and z_t in a single-equation approach Engle and Granger (1987) propose to estimate (4.30) and test whether the residuals from that regression behave as $\mathsf{I}(0)$. To do so, consider the auxiliary regression:

$$\Delta \widehat{\zeta}_t = \alpha \widehat{\zeta}_{t-1} + \omega_t,$$

and test H_0: $\alpha = 0$. The t-ratio for this test is:

$$\mathsf{t}_{\alpha=0} = \frac{\widehat{\alpha}}{\mathsf{ESE}[\widehat{\alpha}]},$$

where:

$$\widehat{\alpha} = \left(\sum_{t=1}^{T} \widehat{\zeta}_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^{T} \widehat{\zeta}_{t-1} \Delta \widehat{\zeta}_t \right),$$

and:

$$\mathsf{ESE}[\widehat{\alpha}] = \widehat{\sigma}_{\omega}\left(\sum_{t=1}^{T}\widehat{\zeta}_{t-1}^{2}\right)^{-\frac{1}{2}} \quad \text{with} \quad \widehat{\sigma}_{\omega}^{2} = \frac{1}{T-1}\sum_{t=1}^{T}\widehat{\omega}_{t}^{2},$$

when $\widehat{\omega}_t = \Delta\widehat{\zeta}_t - \widehat{\zeta}_{t-1}\widehat{\alpha}$ so that:

$$\mathsf{t}_{\alpha=0} = \frac{\left(\sum_{t=1}^{T}\widehat{\zeta}_{t-1}^{2}\right)^{-1}\left(\sum_{t=1}^{T}\widehat{\zeta}_{t-1}\Delta\widehat{\zeta}_{t}\right)}{\widehat{\sigma}_{\omega}\left(\sum_{t=1}^{T}\widehat{\zeta}_{t-1}^{2}\right)^{-\frac{1}{2}}} = \frac{T^{-1}\sum_{t=1}^{T}\widehat{\zeta}_{t-1}\Delta\widehat{\zeta}_{t}}{\widehat{\sigma}_{\omega}\left(T^{-2}\sum_{t=1}^{T}\widehat{\zeta}_{t-1}^{2}\right)^{\frac{1}{2}}}. \tag{4.35}$$

We next find what each term in (4.35) converges to, beginning with $T^{-1}\sum_{t=1}^{T}\widehat{\zeta}_{t-1}\Delta\widehat{\zeta}_{t}$ which equals:

$$\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\left(y_{t-1} - z_{t-1}\widehat{\beta}\right)\left(\Delta y_t - \Delta z_t\widehat{\beta}\right) \\
= \;& T^{-1}\sum_{t=1}^{T}\left(y_{t-1} - z_{t-1}\widehat{\beta}\right)\left(\epsilon_t - \nu_t\widehat{\beta}\right) \\
= \;& T^{-1}\sum_{t=1}^{T}y_{t-1}\epsilon_t - \widehat{\beta}T^{-1}\left(\sum_{t=1}^{T}z_{t-1}\epsilon_t + \sum_{t=1}^{T}y_{t-1}\nu_t\right) + \widehat{\beta}^{2}T^{-1}\sum_{t=1}^{T}z_{t-1}\nu_t \\
\Rightarrow \;& \sigma_{\epsilon}^{2}\Big(\textstyle\int W\mathrm{d}W - \left[\int V^{2}\right]^{-1}\left[\int WV\right]\left[\int V\mathrm{d}W + \int W\mathrm{d}V\right] \qquad (4.36)\\
& + \left[\textstyle\int V^{2}\right]^{-2}\left[\int WV\right]^{2}\int V\mathrm{d}V\Big).
\end{aligned}$$

Next:

$$\begin{aligned}
\widehat{\sigma}_{\omega}^{2} &= \frac{1}{T-1}\sum_{t=1}^{T}\widehat{\omega}_{t}^{2} = \frac{1}{T-1}\sum_{t=1}^{T}\left(\Delta\widehat{\zeta}_t - \widehat{\alpha}\widehat{\zeta}_{t-1}\right)^{2} \\
&= \frac{1}{T-1}\sum_{t=1}^{T}\left[\left(\epsilon_t - \nu_t\widehat{\beta}\right) - \widehat{\alpha}\left(y_{t-1} - z_{t-1}\widehat{\beta}\right)\right]^{2} \\
&= \frac{1}{T-1}\left(\sum_{t=1}^{T}\epsilon_t^{2} - 2\widehat{\beta}\sum_{t=1}^{T}\epsilon_t\nu_t + \widehat{\beta}^{2}\sum_{t=1}^{T}\nu_t^{2} - 2T\widehat{\alpha}\left[T^{-1}\sum_{t=1}^{T}\epsilon_t y_{t-1}\right]\right. \\
&\quad +2T\widehat{\alpha}\widehat{\beta}\left[T^{-1}\sum_{t=1}^{T}\epsilon_t z_{t-1}\right] + 2T\widehat{\alpha}\widehat{\beta}\left[T^{-1}\sum_{t=1}^{T}\nu_t y_{t-1}\right] \\
&\quad -2T\widehat{\alpha}\widehat{\beta}^{2}\left[T^{-1}\sum_{t=1}^{T}\nu_t z_{t-1}\right] + T^{2}\widehat{\alpha}^{2}\left[T^{-2}\sum_{t=1}^{T}y_{t-1}^{2}\right] \\
&\quad \left. -\,2T^{2}\widehat{\alpha}^{2}\widehat{\beta}\left[T^{-2}\sum_{t=1}^{T}y_{t-1}z_{t-1}\right] + T^{2}\widehat{\alpha}^{2}\widehat{\beta}^{2}\left[T^{-2}\sum_{t=1}^{T}z_{t-1}^{2}\right]\right). \qquad (4.37)
\end{aligned}$$

From (4.33) and (4.36):

$$
\begin{aligned}
T\widehat{\alpha} \quad \Rightarrow \quad & \left(\int W(r)^2\,\mathrm{d}r - \left[\int V(r)^2 \mathrm{d}r\right]^{-1}\left[\int W(r)\,V(r)\mathrm{d}r\right]^2\right)^{-1} \\
& \left(\int W\mathrm{d}W - \left[\int V^2\right]^{-1}\left[\int WV\right]\left[\int V\mathrm{d}W + \int W\mathrm{d}V\right]\right. \\
& \left. + \left[\int V^2\right]^{-2}\left[\int WV\right]^2\left[\int V\mathrm{d}V\right]\right),
\end{aligned}
$$

so that from (4.28), (4.29) and (4.32):

$$
\begin{aligned}
\widehat{\sigma}^2_{\omega} \quad = \quad & \frac{1}{T-1}\left(\sum_{t=1}^{T}\epsilon_t^2 - 2\widehat{\beta}\sum_{t=1}^{T}\epsilon_t\nu_t + \widehat{\beta}^2\sum_{t=1}^{T}\nu_t^2\right) + o_{\mathsf{p}}\left(T^{-1}\right) \qquad (4.38) \\
\Rightarrow \quad & \sigma^2_{\epsilon}\left(1 + \left[\frac{\int WV}{\int V^2}\right]^2\right).
\end{aligned}
$$

Hence from (4.33), (4.38) and (4.37), $\mathsf{t}_{\alpha=0}$ converges weakly to:

$$
\frac{\int V^2 \int W\mathrm{d}W + \mathsf{R}^2 \int W^2 \int V\mathrm{d}V - \mathsf{R}\left(\int W^2 \int V^2\right)^{\frac{1}{2}}\left[\int V\mathrm{d}W + \int W\mathrm{d}V\right]}{\left[\int V^2\left(\int V^2 + \mathsf{R}^2\int W^2\right)\left(1-\mathsf{R}^2\right)\int W^2\right]^{\frac{1}{2}}}.
$$

This distribution is well behaved, and has been partially tabulated, allowing an appropriate test to be conducted. Although the asymptotics are supportive of this approach at first sight, Monte Carlo suggests it often performs less well in finite samples (see Banerjee *et al.*, 1986), and can be distorted by nuisance effects such as structural change in the marginal process (see Campos, Ericsson and Hendry, 1996).

4.7 Limiting distribution of OLS under unit roots

Let $\{y_t\}$ be generated for $t = 1, \ldots, T$ by the process:

$$
y_t = \mu t + S_t, \qquad (4.39)
$$

where $S_t = \sum_{j=1}^{t} \upsilon_j$ and $\upsilon_t \sim \mathsf{IN}[0, \sigma^2_{\upsilon}]$ with $S_0 = 0$. Consider estimating by least squares the parameters of the model:

$$
y_t = \mu + \rho y_{t-1} + \omega_t.
$$

(1) Define the scaling matrix:

$$
\mathbf{C}_T = \begin{pmatrix} T^{\frac{1}{2}} & 0 \\ 0 & T^{\frac{3}{2}} \end{pmatrix},
$$

and show that:

$$\mathbf{C}_T \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\rho} - 1 \end{pmatrix} = \begin{pmatrix} 1 & T^{-2} \sum_{t=1}^{T} y_{t-1} \\ T^{-2} \sum_{t=1}^{T} y_{t-1} & T^{-3} \sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-\frac{1}{2}} \sum_{t=1}^{T} \upsilon_t \\ T^{-\frac{3}{2}} \sum_{t=1}^{T} y_{t-1}\upsilon_t \end{pmatrix}$$

$$= \mathbf{B}_T^{-1} \begin{pmatrix} T^{-\frac{1}{2}} \sum_{t=1}^{T} \upsilon_t \\ T^{-\frac{3}{2}} \sum_{t=1}^{T} y_{t-1}\upsilon_t \end{pmatrix}.$$

(2) Given that

$$\begin{aligned} T^{-2} \sum_{t=1}^{T} S_t &\xrightarrow{P} 0, \\ T^{-5/2} \sum_{t=1}^{T} tS_t &\Rightarrow W_1, \\ T^{-2} \sum_{t=1}^{T} S_t^2 &\Rightarrow W_2, \\ T^{-3/2} \sum_{t=1}^{T} S_{t-1}\upsilon_t &\xrightarrow{P} 0, \end{aligned} \tag{4.40}$$

where W_1 and W_2 are non-degenerate distributions, $\xrightarrow{P}$ denotes convergence in probability and $\Rightarrow$ denotes weak convergence, show that:

$$\operatorname*{plim}_{T\to\infty} \mathbf{B}_T = \begin{pmatrix} 1 & \mu/2 \\ \mu/2 & \mu^2/3 \end{pmatrix} := \mathbf{B},$$

and:

$$T^{-\frac{3}{2}} \sum_{t=1}^{T} y_{t-1}\upsilon_t \Rightarrow \mathsf{N}\left[0,\ \tfrac{1}{3}\ \sigma_\upsilon^2\mu^2\right].$$

(3) As the asymptotic covariance between $T^{-1/2} \sum_{t=1}^{T} \upsilon_t$ and $T^{-3/2} \sum_{t=1}^{T} y_{t-1}\upsilon_t$ is $\sigma_\upsilon^2\mu/2$, show that:

$$\begin{pmatrix} T^{-\frac{1}{2}} \sum_{t=1}^{T} \upsilon_t \\ T^{-\frac{3}{2}} \sum_{t=1}^{T} y_{t-1}\upsilon_t \end{pmatrix} \Rightarrow \mathsf{N}_2\left[\mathbf{0}, \sigma_\upsilon^2\mathbf{B}\right],$$

and hence:

$$\mathbf{C}_T \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\rho} - 1 \end{pmatrix} \Rightarrow \mathsf{N}_2\left[\mathbf{0}, \sigma_\upsilon^2\mathbf{B}^{-1}\right],$$

carefully stating any theorems and assumptions which you use.

Hints

Many of the functionals to which sample moments of $\mathsf{I}(1)$ processes converge are expressed in terms of normal densities in table 3.2 in Hendry (1995a) for the process $y_t = \mu t + S_t$ where $S_t = S_{t-1} + \upsilon_t$ and $\upsilon_t \sim \mathsf{IN}[0,1]$. From 1. in his Table 3.2:

$$T^{-1}\sum_{t=1}^{T} S_{t-1}/\sqrt{T} \Rightarrow \int_0^1 W(r)\mathrm{d}r \;\text{ so }\; T^{-1}\sum_{t=1}^{T}\left(S_{t-1}/\sqrt{T}\right)^2 \Rightarrow \int_0^1 W(r)^2\mathrm{d}r,$$

and hence W_1 and W_2 can be made explicit. Since:

$$T^{-(n+1)}\sum_{t=1}^{T} t^n \to (n+1)^{-1} \;\text{ as }\; T\to\infty,$$

given $y_t = \mu t + S_t$, the only parts needed are 2. and 3. from table 3.2, since:

$$T^{-1}\sum_{t=1}^{T} S_{t-1}\upsilon_t \Rightarrow \int_0^1 W(r)\mathrm{d}W(r) \;\text{ so that }\; T^{-\frac{3}{2}}\sum_{t=1}^{T} S_{t-1}\upsilon_t \xrightarrow{P} 0.$$

Thus, the other component converges to:

$$T^{-\frac{3}{2}}\sum_{t=1}^{T} y_{t-1}\upsilon_t = T^{-\frac{3}{2}}\sum_{t=1}^{T}\mu\,(t-1)\,\upsilon_t + T^{-\frac{3}{2}}\sum_{t=1}^{T} S_{t-1}\upsilon_t \Rightarrow \mu\int_0^1 r\mathrm{d}W(r),$$

which is a vector functional of Wiener processes. Alternatively by conventional methods:

$$\begin{pmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T}\upsilon_t \\ T^{-\frac{3}{2}}\sum_{t=1}^{T} y_{t-1}\upsilon_t \end{pmatrix} \xrightarrow{D} \begin{pmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T}\upsilon_t \\ \mu T^{-\frac{3}{2}}\sum_{t=1}^{T} t\upsilon_t \end{pmatrix} \xrightarrow{D} \mathsf{N}_2\left[\mathbf{0}, \sigma_\upsilon^2\mathbf{B}^{-1}\right].$$

(Oxford M.Phil., 1991)

4.7.1 A convenient expression for OLS

Let $\mathbf{X} = (\boldsymbol{\iota} : \mathbf{y}_1)$, with $\boldsymbol{\iota}$ being a $T\times 1$ vector of ones, $\mathbf{y}_1 = L\mathbf{y}$ and $\boldsymbol{\beta} = (\mu : \rho)'$. The OLS estimator of $\boldsymbol{\beta}$ is:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}.$$

But, from (4.39), $\mathbf{y}$ can also be written as:

$$\mathbf{y} = \mu\boldsymbol{\iota} + \mathbf{y}_1 + \boldsymbol{\upsilon} = \mathbf{X}\begin{pmatrix}\mu \\ 1\end{pmatrix} + \boldsymbol{\upsilon} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\upsilon}$$

so that:

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\theta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\upsilon}$$

and hence, subtracting $\boldsymbol{\theta}$ from both sides of this equation and premultiplying by $\mathbf{C}_T$, then $\mathbf{C}_T(\widehat{\boldsymbol{\beta}}-\boldsymbol{\theta})$ equals:

$$\begin{aligned}
& \left(\mathbf{C}_T^{-1}\mathbf{X}'\mathbf{X}\mathbf{C}_T^{-1}\right)^{-1}\mathbf{C}_T^{-1}\mathbf{X}'\boldsymbol{\upsilon} \\
= & \left[\begin{pmatrix} T^{-\frac{1}{2}} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix}\begin{pmatrix} T & \sum_{t=1}^{T} y_{t-1} \\ \sum_{t=1}^{T} y_{t-1} & \sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}\begin{pmatrix} T^{-\frac{1}{2}} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix}\right]^{-1} \\
& \times\begin{pmatrix} T^{-\frac{1}{2}} & 0 \\ 0 & T^{-\frac{3}{2}} \end{pmatrix}\begin{pmatrix} \sum_{t=1}^{T} \upsilon_t \\ \sum_{t=1}^{T} y_{t-1}\upsilon_t \end{pmatrix} \\
= & \begin{pmatrix} 1 & T^{-2}\sum_{t=1}^{T} y_{t-1} \\ T^{-2}\sum_{t=1}^{T} y_{t-1} & T^{-3}\sum_{t=1}^{T} y_{t-1}^2 \end{pmatrix}^{-1}\begin{pmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T} \upsilon_t \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}\upsilon_t \end{pmatrix} \\
= & \mathbf{B}_T^{-1}\begin{pmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T} \upsilon_t \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}\upsilon_t \end{pmatrix}.
\end{aligned}$$

4.7.2 Convergence of terms in the OLS expression

Let us first find what each element in matrix $\mathbf{B}_T$ converges to. First:

$$T^{-2}\sum_{t=1}^{T} y_{t-1} = T^{-2}\sum_{t=1}^{T}\left[\mu\left(t-1\right)+S_{t-1}\right] = \frac{\left(T-1\right)T\mu}{2T^2} + T^{-2}\sum_{t=1}^{T} S_{t-1},$$

but, from (4.40a), $T^{-2}\sum_{t=1}^{T} S_{t-1} \xrightarrow{P} 0$ and so:

$$T^{-2}\sum_{t=1}^{T} y_{t-1} \xrightarrow{P} \frac{\mu}{2}.$$

Next:

$$\begin{aligned}
T^{-3}\sum_{t=1}^{T} y_{t-1}^2 &= T^{-3}\sum_{t=1}^{T}\left[\mu\left(t-1\right)+S_{t-1}\right]^2 \\
&= T^{-3}\sum_{t=1}^{T}\left[\mu^2\left(t-1\right)^2+2\mu\left(t-1\right)S_{t-1}+S_{t-1}^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(T-1)\,T\,(2T-1)\,\mu^2}{6T^3} + 2\mu T^{-3}\sum_{t=1}^{T}(t-1)\,S_{t-1} \\
&\quad + T^{-3}\sum_{t=1}^{T}S_{t-1}^2 \\
&\Rightarrow \frac{\mu^2}{3},
\end{aligned}
$$

because of (4.40b,c). Hence:

$$
\mathbf{B}_T \xrightarrow{P} \begin{pmatrix} 1 & \mu/2 \\ \mu/2 & \mu^2/3 \end{pmatrix} = \mathbf{B}.
$$

Next, we find the marginal asymptotic distribution of $T^{-3/2}\sum_{t=1}^{T} y_{t-1}\upsilon_t$. Write:

$$
\begin{aligned}
T^{-3/2}\sum_{t=1}^{T} y_{t-1}\upsilon_t &= T^{-3/2}\sum_{t=1}^{T}\left[\mu\,(t-1)+S_{t-1}\right]\upsilon_t \qquad (4.41) \\
&= \mu T^{-3/2}\sum_{t=1}^{T}(t-1)\,\upsilon_t + T^{-3/2}\sum_{t=1}^{T}S_{t-1}\upsilon_t \\
&= \mu T^{-3/2}\sum_{t=1}^{T}(t-1)\,\upsilon_t + o_{\mathsf{p}}(1),
\end{aligned}
$$

because of (4.40d). Hence, by Cramér's theorem the asymptotic distribution of $T^{-3/2}\sum_{t=1}^{T} y_{t-1}\upsilon_t$ equals that of $\mu T^{-3/2}\sum_{t=1}^{T}(t-1)\,\upsilon_t$. Let us find that next. Because $\{\upsilon_t\}$ is independently distributed, so is $\{(t-1)\,\upsilon_t\}$ with mean and variance:

$$
\mathsf{E}\left[(t-1)\,\upsilon_t\right] = (t-1)\,\mathsf{E}\left[\upsilon_t\right] = 0,
$$

and:

$$
\mathsf{V}\left[(t-1)\,\upsilon_t\right] = (t-1)^2\,\sigma_\upsilon^2,
$$

so that $(t-1)\upsilon_t$ is independently, but heterogenously, distributed. Thus, we check whether Liapunov's central limit theorem can be applied. For that, two conditions must be satisfied. First, a central absolute moment of order $2+\delta$, $\delta > 0$, must exist; and second, the ratio of the sum of that moment to the sum of the variances to the power δ must converge to zero as $T\to\infty$. Because $\{\upsilon_t\}$ is normally distributed then $\{(t-1)\,\upsilon_t\}$ is also normally distributed, and so moments of higher order than second exist, and in particular its fourth-order moment is:

$$
\mathsf{E}\left[(t-1)^4\,\upsilon_t^4\right] = (t-1)^4\,\mathsf{E}\left[\upsilon_t^4\right] = 3\,(t-1)^4\,\sigma_\upsilon^4 < \infty,
$$

for all t, so that the first condition is satisfied. Also, the sum of the variances is:

$$
S_T^2 = \sum_{t=1}^{T}\mathsf{V}\left[(t-1)\,\upsilon_t\right] = \sigma_\upsilon^2\sum_{t=1}^{T}(t-1)^2 = \frac{(T-1)\,T\,(2T-1)\,\sigma_\upsilon^2}{6},
$$

so that:

$$\lim_{T\to\infty} \frac{3\sigma_\upsilon^4}{S_T^4} \sum_{t=1}^{T-1} t^4 = \lim_{T\to\infty} \frac{108\,(3\,\{T-1\}\,T-1)}{30\,(T-1)\,T\,(2T-1)} = 0,$$

and hence the second condition is also satisfied. Consequently, Liapunov's theorem can be applied, yielding:

$$\left[\frac{1}{6}\,(T-1)\,T\,(2T-1)\,\sigma_\upsilon^2\right]^{-\frac{1}{2}} \sum_{t=1}^{T} (t-1)\,\upsilon_t \xrightarrow{D} \mathsf{N}\,[0,1]\,.$$

Since:

$$\left(\frac{6}{(T-1)\,T\,(2T-1)}\right)^{\frac{1}{2}} - \left(\frac{3}{T^3}\right)^{\frac{1}{2}} = o_{\mathsf{p}}\,(1)\,,$$

and by Cramér's theorem:

$$\frac{\sqrt{3}}{\sigma_\upsilon} T^{-3/2} \sum_{t=1}^{T} (t-1)\,\upsilon_t \xrightarrow{D} \mathsf{N}\,[0,1] \tag{4.42}$$

then by Cramér's theorem again:

$$T^{-3/2} \sum_{t=1}^{T} y_{t-1}\upsilon_t \xrightarrow{D} \mathsf{N}\left[0, \frac{1}{3}\mu^2\sigma_\upsilon^2\right].$$

4.7.3 Joint asymptotic distribution

To find the asymptotic distribution of $\mathbf{C}_T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta})$, use (4.41) to write its expression as:

$$\begin{aligned}
&\mathbf{B}_T^{-1} \begin{pmatrix} T^{-\frac{1}{2}} \sum\limits_{t=1}^{T} \upsilon_t \\ \mu T^{-3/2} \sum\limits_{t=1}^{T} (t-1)\,\upsilon_t + o_{\mathsf{p}}\,(1) \end{pmatrix} \\
&= \mathbf{B}_T^{-1} T^{-\frac{1}{2}} \sum_{t=1}^{T} \begin{pmatrix} 1 \\ \mu T^{-1}\,(t-1) \end{pmatrix} \upsilon_t + o_{\mathsf{p}}\,(1) \\
&= \mathbf{B}_T^{-1} T^{-\frac{1}{2}} \sum_{t=1}^{T} \mathbf{z}_t \upsilon_t + o_{\mathsf{p}}\,(1) \\
&= \mathbf{B}_T^{-1} T^{-\frac{1}{2}} \mathbf{Z}'\boldsymbol{\upsilon} + o_{\mathsf{p}}\,(1)\,,
\end{aligned} \tag{4.43}$$

where $\mathbf{Z} = \left(\boldsymbol{\iota}, \mu T^{-1}\mathbf{w}\right)$ with $\mathbf{w}' = (0, 1, \ldots, T-1)$, and derive the variance matrix $\boldsymbol{\Omega}_T$ of:

$$\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} \upsilon_t : \mu T^{-3/2} \sum_{t=1}^{T} (t-1)\upsilon_t\right)',$$

as follows. $\mathbf{\Omega}_T$ is defined by:

$$\mathsf{E}\left[\begin{pmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T}\upsilon_t \\ \mu T^{-3/2}\sum_{t=1}^{T}(t-1)\upsilon_t \end{pmatrix}\left(T^{-\frac{1}{2}}\sum_{t=1}^{T}\upsilon_t : \mu T^{-3/2}\sum_{t=1}^{T}(t-1)\upsilon_t\right)'\right]$$

$$= \mathsf{E}\begin{pmatrix} T^{-1}\left(\sum_{t=1}^{T}\upsilon_t\right)^2 & \mu T^{-2}\left(\sum_{t=1}^{T}\upsilon_t\right)\left(\sum_{t=1}^{T}(t-1)\upsilon_t\right) \\ \mu T^{-2}\left(\sum_{t=1}^{T}\upsilon_t\right)\left(\sum_{t=1}^{T}(t-1)\upsilon_t\right) & \mu^2 T^{-3}\left(\sum_{t=1}^{T}(t-1)\upsilon_t\right)^2 \end{pmatrix}.$$

Keeping in mind that $\{\upsilon_t\}$ is IID, consider each element separately. First:

$$\begin{aligned}\mathsf{E}\left[\left(\sum_{t=1}^{T}\upsilon_t\right)^2\right] &= \mathsf{E}\left[\sum_{t=1}^{T}\upsilon_t^2 + \sum_{t,s=1}^{T}\upsilon_t\upsilon_s\right] \\ &= \sum_{t=1}^{T}\mathsf{E}\left[\upsilon_t^2\right] + \sum_{t\neq s}^{T}\mathsf{E}\left[\upsilon_t\upsilon_s\right] = T\sigma_\upsilon^2;\end{aligned}$$

second:

$$\mathsf{E}\left[\left(\sum_{t=1}^{T}\upsilon_t\right)\left(\sum_{t=1}^{T}(t-1)\upsilon_t\right)\right] = \mathsf{E}\left[\sum_{t,s=1}^{T}(s-1)\upsilon_t\upsilon_s\right] = \frac{(T-1)T\sigma_\upsilon^2}{2};$$

and third:

$$\begin{aligned}\mathsf{E}\left[\left(\sum_{t=1}^{T}(t-1)\upsilon_t\right)^2\right] &= \mathsf{E}\left[\sum_{t,s=1}^{T}(t-1)(s-1)\upsilon_t\upsilon_s\right]^2 \\ &= \frac{(T-1)T(2T-1)\sigma_\upsilon^2}{6}.\end{aligned}$$

Hence, as $T\to\infty$, the variance matrix is:

$$\begin{aligned}\mathbf{\Omega}_T &= \sigma_\upsilon^2\begin{pmatrix} 1 & \frac{(T-1)\mu}{2T} \\ \frac{(T-1)\mu}{2T} & \frac{(T-1)(2T-1)\mu^2}{6T^2}\end{pmatrix} \\ &\xrightarrow{P} \sigma_\upsilon^2\begin{pmatrix} 1 & \mu/2 \\ \mu/2 & \mu^2/3\end{pmatrix} \\ &= \sigma_\upsilon^2\mathbf{B}.\end{aligned}$$

Having derived $\mathbf{\Omega}_T$, we next turn to finding the joint asymptotic distribution of $\mathbf{B}_T^{-1}T^{-1/2}\mathbf{Z}'\boldsymbol{\upsilon}$. To do so, we first check whether the conditions for a central-limit theorem are satisfied. $\mathbf{z}_t$ is non-stochastic and, by assumption, $\upsilon_t \sim \mathsf{IN}[0, \sigma_\upsilon^2]$ so that $\{\mathbf{z}_t\upsilon_t\}$ is independently, although not identically, distributed. Hence, we again think of checking whether the conditions for Liapunov's theorem are satisfied. Define:

$$w_t = \boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\upsilon_t,$$

where $\boldsymbol{\lambda}$ is a 2×1 vector of constants such that $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$. Then because $\{\mathbf{z}_t\upsilon_t\}$ is independently distributed, w_t is independently distributed, with mean:

$$\mathsf{E}\left[w_t\right] = \mathsf{E}\left[\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\upsilon_t\right] = \boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\mathsf{E}\left[\upsilon_t\right] = 0,$$

and variance:

$$\begin{aligned}
\mathsf{V}\left[w_t\right] &= \mathsf{E}\left[\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\upsilon_t^2\mathbf{z}_t'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\boldsymbol{\lambda}\right] \\
&= \boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\mathsf{E}\left[\upsilon_t^2\right]\mathbf{z}_t'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\boldsymbol{\lambda} \\
&= \sigma_\upsilon^2\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\mathbf{z}_t'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\boldsymbol{\lambda} \\
&= \sigma_\upsilon^2\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\begin{pmatrix} 1 & \mu\left(t-1\right)/T \\ \mu\left(t-1\right)/T & \mu^2\left(t-1\right)^2/T^2 \end{pmatrix}\boldsymbol{\Omega}_T^{-\frac{1}{2}}\boldsymbol{\lambda}.
\end{aligned}$$

The first condition is:

$$\begin{aligned}
\mathsf{E}\left[w_t^4\right] &= \mathsf{E}\left[\left(\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\upsilon_t\right)^4\right] \\
&= \left(\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\right)^4\mathsf{E}\left[\upsilon_t^4\right] \\
&= 3\sigma_\upsilon^4\left(\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\mathbf{z}_t\right)^4 < \infty
\end{aligned}$$

for all t, because $\boldsymbol{\lambda}$ is a vector of constants, $\boldsymbol{\Omega}_T$ is non-singular, and so $\boldsymbol{\Omega}_T^{-1}$ is finite, and $\mathbf{z}_t$ is finite for all t. Next, for the second condition:

$$\begin{aligned}
S_T^2 &= \sum_{t=1}^{T}\mathsf{V}\left[w_t\right] \\
&= \sigma_\upsilon^2\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\sum_{t=1}^{T}\begin{pmatrix} 1 & \mu\left(t-1\right)/T \\ \mu\left(t-1\right)/T & \mu^2\left(t-1\right)^2/T^2 \end{pmatrix}\boldsymbol{\Omega}_T^{-\frac{1}{2}}\boldsymbol{\lambda} \\
&= \sigma_\upsilon^2T\boldsymbol{\lambda}'\boldsymbol{\Omega}_T^{-\frac{1}{2}}\begin{pmatrix} 1 & \dfrac{(T-1)\mu}{2T} \\ \dfrac{(T-1)\mu}{2T} & \dfrac{(T-1)\left(2T-1\right)\mu^2}{6T^2} \end{pmatrix}\boldsymbol{\Omega}_T^{-\frac{1}{2}}\boldsymbol{\lambda} \\
&= \sigma_\upsilon^2T\boldsymbol{\lambda}'\boldsymbol{\lambda} = \sigma_\upsilon^2T > 0.
\end{aligned}$$

Also, because $\boldsymbol{\Omega}_T$ is $O(1)$ and is non-singular, then $\boldsymbol{\Omega}_T^{-1/2}$ is $O(1)$, so that the term of largest order in T in $\sum_{t=1}^{T} \mathsf{E}[w_t^4]$ is $O(T)$. Hence:

$$S_T^{-4} \sum_{t=1}^{T} \mathsf{E}\left[w_t^4\right] = O\left(T^{-1}\right),$$

implying that the second condition for Liapunov's theorem is also satisfied. So, by Liapunov's central-limit theorem:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{\lambda}' \boldsymbol{\Omega}_T^{-\frac{1}{2}} \mathbf{z}_t \upsilon_t = T^{-\frac{1}{2}} \boldsymbol{\lambda}' \boldsymbol{\Omega}_T^{-\frac{1}{2}} \mathbf{Z}' \boldsymbol{\upsilon} \xrightarrow{D} \mathsf{N}\left[0, 1\right],$$

and by the Cramér–Wold device:

$$T^{-\frac{1}{2}} \boldsymbol{\Omega}_T^{-\frac{1}{2}} \mathbf{Z}' \boldsymbol{\upsilon} \xrightarrow{D} \mathsf{N}_2\left[\mathbf{0}, \mathbf{I}_2\right].$$

We can finally apply Cramér's theorem to find that:

$$\mathbf{C}_T \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\rho} - 1 \end{pmatrix} = \mathbf{B}_T^{-1} T^{-\frac{1}{2}} \mathbf{Z}' \boldsymbol{\upsilon} + o_{\mathsf{p}}(1) \xrightarrow{D} \mathsf{N}_2\left[\mathbf{0}, \sigma_\upsilon^2 \mathbf{B}^{-1}\right].$$

This, of course, is the distribution of the estimated coefficients in a unit-root process with drift, which are asymptotically normal (see e.g., West, 1988).