CHAPTER 4

Regression with Nonstationary Variables

The distributed-lag models discussed above are appropriate when y, x, and u are stationary time series. What happens if one or more of these series is nonstationary? The answer to this question depends on the nature of the nonstationarity.

If there are discrete breakpoints at which the structure of the individual series or the relationship between them changes, then we must adapt our model to accommodate these changes. Splitting the sample into two or more sub-samples is the most obvious and common way of doing this.

Little econometric attention has been devoted to the case where the series are explosive, such as an AR(1) process with a parameter $\phi_1 > 1$. Such series are probably uncommon in economics, since they have the property that a small shock leads to an ever-increasing effect on the series that becomes infinite in the limit.

The cases that have dominated modern time-series analysis are borderline-nonstationary cases. The most common one is regression involving ***unit-root*** or ***integrated*** time-series processes. Another case that has received some attention is the ***trend-stationary*** process for which deviations from a deterministic trend are stationary.

4.1 Nonstationarity Due to Structural Change

4.1.1 A general model of structural change

Economic relationships among variables may change over time. New laws or other aspects of the institutional environment can change discretely at a particular point in time, leading to changes in economic agents' behavior. Or behavior may evolve gradually over time. In either case, the parameters of econometric model are likely to change—suddenly or gradually—through the time-series sample.

Ideally, the econometrician can measure the variables that have caused changes in behavior and incorporate them in the model. Suppose that the model of interest is

$$y_t = \alpha + \beta x_t + \varepsilon_t,$$

where we assume for simplicity that there is only one explanatory variable, there is no lag structure, and the error term is white noise. (All of these simplifications can be readily generalized in a straightforward way.) We observe a variable z_t that is likely to change the relationship between y and x, perhaps a tax rate or other policy variable.

If a change in z affects $E(y_t \mid x_t)$, but not $\partial E(y_t \mid x_t)/\partial x_t$, then only the level of the relationship between y and x is affected by z and we can insert z as a variable in the regression:

$$y_t = \delta + \gamma z_t + \beta x_t + \varepsilon_t, \tag{4.1}$$

with the constant intercept term α being replaced by a function of z: $\delta + \gamma z_t$. If changes in z affect the slope $\beta = \partial E(y_t \mid x_t)/\partial x_t$ as well as the level of the relationship, then an interaction term is required:

$$y_t = \delta + \gamma z_t + (\lambda_0 + \lambda_1 z_t) x_t + \varepsilon_t = \delta + \gamma z_t + \lambda_0 x_t + \lambda_1 x_t z_t + \varepsilon_t. \tag{4.2}$$

Here the original intercept term has been replaced with $\delta + \gamma z_t$ as above, plus the β coefficient measuring the effect of x on y is replaced with $\lambda_0 + \lambda_1 z_t$.

Modeling changes in intercept terms or slope coefficients over time as a function of another variable is the simple extension to a time-series context of the general procedure for adding a variable to the model, including the possibility of an interaction. We can test the hypothesis that $\gamma = 0$ in equation (4.1) or that $\gamma = \lambda_1 = 0$ in equation (4.2) to determine if the effect is statistically significant.

4.1.2 Modeling known structural breakpoints

The variable z may be a dummy variable reflecting a discrete change in the environment or a continuous variable that leads to gradual changes. A common special case occurs when z is a dummy variable that switches from zero to one at a fixed date and remains at one through the rest of the sample:

$$z_t = \begin{cases} 0, & t = 1, 2, \ldots, T_1 \\ 1, & t = T_1 + 1, T_1 + 2, \ldots, T. \end{cases} \tag{4.3}$$

For the case of a dummy variable such as this, we can think of its effect as a "breakpoint" in the sample. The relationship between y and x is different (in intercept, slope, or both) after T_1 than before.

If we know the breakpoint T_1, then the test of $\gamma = \lambda_1 = 0$ becomes a simple Chow test of the null hypothesis that the coefficients are the same before and after the break. We can often hypothesize about possible breakpoints even if we cannot always measure the underlying variables that cause the relationship to change. For example, there is considerable econometric evidence that macroeconomic relationships in the United States (and other advanced economies) changed around 1973. Many factors may explain why the macroeconomy

changed in the early 1970s: a discrete rise in oil prices followed by OPEC's waxing influence, the move to floating exchange rates, the emergence of Japan as an industrial power and the increasing effects of globalization, and the entry of the baby-boom generation into the work force are a few. Even though we may have variables that measure some of these changes, disentangling the effects of all of these simultaneously changing factors may be beyond the power of our data. Therefore it is common to use a dummy variable to capture the overall effect of multiple changes in the economic environment around 1973 and consider whether the sample should be split there.

4.1.3 Estimating unknown breakpoints

We often suspect that the structure of the model has changed at some point in the sample, but we do not know the date of the breakpoint. In other words, we believe that an equation such as (4.2) with z_t defined in (4.3) is a good model for the data, but the date of the breakpoint T_1 is unknown.

With a known breakpoint, the Chow F statistic provides us with a statistical measure of the magnitude of the break. So comparing the Chow F statistic associated with different possible breakpoints could give us an indicator of which breakpoint seems to be most strongly supported by the data. This is the intuition of the ***Quandt likelihood ratio (QLR) test***. [This section is based on Stock & Watson, 3/e, pp. 558-561. See references there to Quandt (1960) and Andrews (2003).]

To implement the QLR test, we must deal with two issues. First, in order to test whether two sub-samples have the same coefficients, we must have enough observations in each subsample to get reliable coefficient estimates. This means that we cannot detect or test potential breakpoints that are close to either end of the sample. The reliability of the sub-sample estimates depends on the number of degrees of freedom, the difference between the number of observations in the sub-sample and the number of coefficients we want to estimate. Thus, the degree of "trimming" of possible breakpoint dates that is necessary will depend on the length of the sample and the number of parameters in the model. A conventional choice is to trim 15% of the observations from each end of the sample, looking for breakpoints only within the central 70% of observations. Figure 4-1 shows a schematic representation of trimming. The breakpoints must obviously be rounded to integer values.

If τ_1 and τ_2 are the minimum and maximum observations we consider as possible breakpoints, the QLR statistic is defined as

$$\text{QLR} = \max\left\{F_{\tau_1}, F_{\tau_1+1}, \ldots, F_{\tau_2-1}, F_{\tau_2}\right\},$$

where F_τ is the Chow F statistic for a breakpoint at observation τ. Because QLR is the *maximum* of a set of F statistics, it does not follow the F distribution.

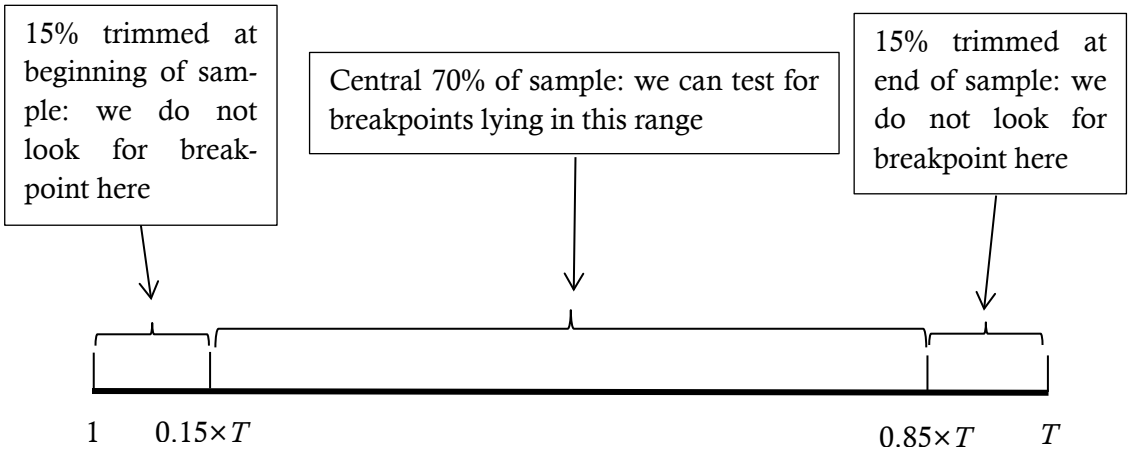

Figure 4-1. Trimming in QLR test

The distribution of the QLR statistic obviously depends on the magnitude of trimming. In an extreme case, think what would happen if we trimmed all but one observation from the ends of the sample, so we tested for only a single possible breakpoint. This is the simple Chow test for a single, fixed breakpoint; the test statistic follows the F distribution.

As we widen the middle range of possible values of T_1 that we consider as candidate breakpoints, our test statistic is the maximum of a growing number of Chow statistics, which means that we expect the QLR statistic to be larger than a standard F statistic. In Table 14.6, Stock and Watson (2011, 559) give a tabulation of critical values for the QLR statistic with 15% trimming, based Andrews (2003).

The QLR test can be readily generalized to test for more than one possible breakpoint, and indeed has been demonstrated to be effective at testing the null hypothesis of structural stability even when the change in the coefficients is continuous rather than discrete. There does not seem to be an implementation of the QLR test available for Stata.

4.2 Trend-Stationary Variables

We now suppose that there are no breakpoints in the sample, in other words, that the same coefficients and error properties apply to all of our observations. A second kind of nonstationarity that may apply to such models is that they may include a ***deterministic trend***. The deterministic-trend model has fallen out of favor with time-series econometricians in recent decades, often being replaced by models of ***stochastic trends*** that we will study in the next section. Stock and Watson (2011, 547) argue as follows: "Like many econometricians, we think it is more appropriate to model economic time series as having stochastic rather than deterministic trends, Economics is complicated stuff. It is hard to reconcile the predictability im-

plied by a deterministic trend with the complications and surprises faced year after year by workers, businesses, and governments."

Consider the model

$$y_t = \alpha + \gamma t + u_t, \tag{4.4}$$

where u_t is a stationary disturbance term with constant variance σ_u^2. The variable y_t has constant variance (and covariance) over time, but its mean $E(y_t) = \gamma t$ changes with t, so y_t is nonstationary as, of course, is t itself.

4.3 Integrated Processes: Levels, Differences, and Cointegration

In a regression with one regressor, there are three variables that could be stationary or non-stationary: the dependent variable y, the regressor x, or the disturbance term u. The appropriate econometric treatment of the model depends crucially on the pattern of stationarity and non-stationarity of these three variables.

Recall that in integrated variable (of order one) is a variable whose first difference is stationary. We use the notation $I(1)$ to denote such a variable and $I(0)$ to denote a stationary variable. We use the term "levels" to refer to the actual values of the variable y_t, (or x_t or u_t) and the term "differences" to refer to the first differences $\Delta y_t \equiv (1 - L)y_t = y_t - y_{t-1}$. By definition, if y_t is $I(1)$, then Δy_t is $I(0)$.

To estimate the dynamic relationship between y and x, it is important to get the orders of integration right! If the dependent variable is integrated, then at least some of the regressors must also be integrated, otherwise we are trying to explain something that is nonstationary by a set of explanatory variables that are not. Similarly, if the dependent variable is stationary, then it cannot "follow" an integrated explanatory variable on its nonstationary wanderings, so the model must be misspecified.

With one regressor, the order of integration of y and x must match for the specification to make economic sense. With more than one regressor and an integrated dependent variable, it is possible to have a mixture of integrated and stationary regressors. For example, we could add some (stationary) dummy variables to a regression with integrated y and x. A good rule of thumb is that you can't explain something nonstationary with (only) stationary variables. Any nonstationary regressor will transmit its nonstationarity to the dependent variable, so you cannot explain a stationary variable with a nonstationary one.

4.3.1 Dealing with spurious regressions by differencing

We saw in the baseball example that opened Chapter 2 that regressions in which y and x are nonstationary lead to misleading conclusions: R^2 and t statistics are likely to be large even if the underlying variables are not truly related. The second column of Table 2-1 shows that performing the same regression in terms of first differences yields the (correct) result that we cannot reject the null hypothesis of a zero coefficient.

This suggests that differencing may be appropriate in nonstationary models, and this is often correct.[1] Granger and Newbold (1974) present strong evidence that regressions involving random walks are spurious when performed on the levels, but not on the differences. Table 4-1 is taken from their subsequent book and reports the results of a Monte Carlo study in which they generated unrelated random walks and performed regressions on them. The top part of the table shows the results for regressions in levels, where all of the variables on both sides of the equation are $I(1)$; the bottom part shows the regressions where all variables are differenced, and thus $I(0)$.

If the variables are unrelated, the true null hypothesis that all β coefficients are zero should be rejected 5% of the time. We see rejection rates of 76% to 96% depending on the number of regressors in the equation when estimating in levels, but correctly see 2% to 10% rejection when the variables are made stationary by differencing.[2] Similarly, the adjusted R^2 coefficients are inflated in levels, but near zero with differencing. With five regressions, they find an adjusted R^2 over 0.7 more than one-third of the time!

Table 4-1. Granger and Newbold's spurious regression result

	Number of regressors	% rejection of F test that all β coefficients are zero (@ 0.05)	Average $\bar{R}^2$	% of $\bar{R}^2 > 0.7$
Levels	1	76	0.26	5
	2	78	0.34	8
	3	93	0.46	25
	4	95	0.55	34
	5	96	0.59	37
Differences	1	8	0.004	0
	2	4	0.001	0
	3	2	–0.007	0
	4	10	0.006	0
	5	6	0.012	0

[1] An important exception is the special case of cointegration, which is discussed below.

[2] Computer time was not cheap in the 1970s, so Granger and Newbold made due with only 100 samples. We could replicate this result for 100,000 samples on today's computers in a matter of minutes.

Formally, suppose that our hypothesized model is $y_t = \alpha + \beta x_t + u_t$. (We could add lags of x or additional $I(1)$ regressors without changing the basic principle.) Both y and x are $I(1)$ variables such as random walks, and we shall assume that the error term u is also $I(1)$. Taking the first difference of the equation yields

$$\begin{aligned} y_t &= \alpha + \beta x_t + u_t \\ - \quad \underline{y_{t-1}} &= \underline{\alpha + \beta x_{t-1} + u_{t-1}} \\ \Delta y_t &= \quad \beta \Delta x_t + \Delta u_t \end{aligned} . \tag{4.5}$$

If y, x, and u are all $I(1)$, then their differences, which appear in (4.5), are all stationary and we can estimate β reliably.

However, notice that the constant term α disappears when we take differences. Because α affects all values of y in the same way, taking the difference eliminates it from the equation. When performing a regression in differences, we generally want to remove the constant term. Including a constant in the differenced equation would be equivalent to having a time trend in the original "levels" equation.

4.3.2 Cointegration

A special case of great interest to econometricians arises when y and x are $I(1)$, but the error term in the relationship between them u is stationary. In this case, we say that y and x are ***cointegrated***. Two series that are cointegrated are nonstationary, but they are nonstationary "together." Think of two variables taking a "random walk together," in which they both move in a nonstationary manner over time, but the difference between them (or some other linear function of them) is stationary, tending to return back to a stable, constant value after being disturbed by a shock.

The econometric concept of cointegration often fits nicely with the economic concept of long-run equilibrium. For example, the median price series for houses in Portland and Beaverton are both nonstationary—they are never going to return to the levels of two decades ago. However, prices in the two cities cannot get too far out of line with each other, so it is plausible that some linear combination of the two price series would be stationary, tending to return to zero after a shock.

In terms of our model, we again have $y_t = \alpha + \beta x_t + u_t$, but we now assume that u is $I(0)$ (with x and y both still assumed to be $I(1)$). We call this equation, with nonstationary variables but a stationary error, a ***cointegrating equation***. The vector of coefficients $(1, -\alpha, -\beta)$ that makes $1y_t - \alpha - \beta x_t = u_t$ stationary is called a ***cointegrating vector***. For two variables, there can be only one cointegrating vector having a coefficient of one on y, although any multiple of the cointegrating vector $(1, -\alpha, -\beta)$ is also a cointegrating vector because, for example, $2y_t - 2\alpha - 2\beta x_t = 2u_t$ is obviously also stationary.

Can we estimate a cointegrated model in differences? Yes, but we probably do not want to. If u is stationary, then its difference Δu is also stationary, so the regression in differences is valid. However, differencing the regression loses information contained in the stable long-run relationship between the series.

To see why this happens consider a simple model. Suppose that P is the price of houses in Portland and B is the price of houses in Beaverton. The long-run relationship between them is

$$B_t = 0 + 0.9P_t + u_t, \tag{4.6}$$

where B and P are $I(1)$ and u is $I(0)$. (The zero constant term is included just to show that there can be one.) The differenced equation is

$$\Delta B_t = 0.9\Delta P_t + v_t, \tag{4.7}$$

where $v = \Delta u$.

Suppose that a shock u_1 in period one causes the Beaverton price to increase relative to the Portland price by 1 unit (perhaps \$1,000). Assuming, for simplicity, that u_0 was zero, then $v_1 = u_1 - u_0 = u_1 = +1$. Because u is stationary, we know that this shock is going to dissipate over time and that Beaverton's house prices will eventually be expected to fall back down to their long-run equilibrium relationship with Portland's. However, if we use equation (4.7), we will predict that future changes in Beaverton's price will be 0.9 times the changes in Portland's because $E(v_2) = E(v_3) = \ldots = 0.$. There is no tendency in (4.7) to restore the long-run equilibrium relationship in (4.6).

4.3.3 Error-correction models

How can we build our knowledge that u is stationary into our prediction? Because $u_1 > 0$, we expect that future values of u will be less than u_1, which means that future values of $v_t = \Delta u_t$ will be negative. Thus, cointegration means that $E(v_2 \mid u_1 > 0) < 0$, which is lost in the differenced equation (4.7) when we ignore past shocks and simply assume $E(v_2) = 0$. We need to modify equation (4.7) to include a term that reflects the tendency of B to return to its long-run equilibrium relationship to P.[3] A model incorporating such a term is called an ***error-correction model***.

From equation (4.6), the degree to which B_t is above or below its long-run equilibrium relationship to P_t at the beginning of period t is measured by $u_{t-1} = B_{t-1} - 0.9P_{t-1}$. An error-

[3] We might expect that an "over-differenced" error term like v would be negatively serially correlated: a positive value in one period would be followed by negative values afterward as u returns to zero.

correction model augments (4.7) with an ***error-correction term*** involving u_{t-1}, providing a mechanism to drive the variables back toward long-run equilibrium:

$$\Delta B_t = \beta \Delta P_t - \lambda u_{t-1} + v_t = \beta \Delta P_t - \lambda \left(B_{t-1} - 0.9 P_{t-1} \right) + v_t . \tag{4.8}$$

We expect that $-\lambda < 0$ so that a positive value of u_{t-1} is associated with reductions in future changes in B below what would be predicted by the corresponding future changes in P as Beaverton price return to their normal relationship to Portland price. Notice that the error-correction equation (4.8) is "balanced" in the sense that all terms on both sides—ΔB, ΔP, u, and v—are $I(0)$. We can also estimate (4.8) successfully by OLS because all terms are stationary.

In order to estimate (4.8), we must know the cointegrating vector so that we can calculate u_{t-1}. In this example, we must know the value 0.9. We can estimate the cointegrating vector in one of two ways. We can estimate (4.8) by nonlinear least-squares with 0.9 replaced by an unknown parameter γ_1 (and a constant γ_0, because we would not know that the constant was zero):

$$\Delta B_t = \beta \Delta P_t - \lambda \left(B_{t-1} - \gamma_0 - \gamma_1 P_{t-1} \right) + v_t .$$

However, we can also follow a two-step procedure in which we first estimate the cointegrating vector by running the "cointegrating regression"

$$B_t = \gamma_0 + \gamma_1 P_t + u_t \tag{4.9}$$

in levels and then using estimated coefficients $\hat{\gamma}_0$ and $\hat{\gamma}_1$ to calculate $\hat{u}_{t-1} \equiv B_{t-1} - \hat{\gamma}_0 - \hat{\gamma}_1 P_{t-1}$ in the error-correction model (4.8).

How can we get away with estimating (4.6) without encountering spurious regression difficulties given that both of the variables are $I(1)$? It turns out that in the special case of cointegration, the OLS estimator $\hat{\gamma}$ is not only consistent, it is "super-consistent," meaning that its variance converges to zero at a rate proportional to $1/T$ rather than the usual rate, which is proportion to $1/\sqrt{T}$. Intuitively, this happens because it is very easy for OLS to find the "right" values of γ; any other value leads to a non-stationary error term which will tend to have large squared residuals.[4] Moreover, because $\hat{\gamma}$ is super-consistent, we can estimate

$$\Delta B_t = \beta \Delta P_t - \lambda \left(B_{t-1} - \hat{\gamma}_0 - \hat{\gamma}_1 P_{t-1} \right) + v_t \tag{4.10}$$

[4] Although the OLS coefficient estimator in the cointegrating regression is super-consistent, we still cannot use t tests based on its estimated standard error for the same reasons as in the spurious-regression case.

without worrying about the potential inaccuracy of $\hat{\gamma}$—we can treat it as a known constant.

Although the long-run relationship between the levels of B and P can probably be described effectively without worrying too much about lags, the adjustment of B to P over time is probably not immediate. This leads us to think about incorporating lagged differences into the error-correction model. If we allow for p lags of ΔB_t and q lags of ΔP_t, we arrive at a model like

$$\Delta B_t = \phi_1 \Delta B_{t-1} + \ldots + \phi_p \Delta B_{t-p} + \theta_0 \Delta P_t + \theta_1 \Delta P_{t-1} + \ldots + \theta_q \Delta P_{t-q} - \lambda\left(B_{t-1} - \hat{\gamma}_0 - \hat{\gamma}_1 P_{t-1}\right) + \nu_t \quad (4.11)$$

Equation (4.11) is a typical form of an error-correction model, with the lengths p and q of the lags to be determined by the methods discussed in Chapter 3.

4.3.4 Summarizing regression with nonstationary variables

Table 4-2 summarizes the four possible cases of stationarity and nonstationarity ($I(1)$) for regressors and the error term. If $y_t = \alpha + \beta x_t + u_t$, then the time-series behavior of y is governed by the behavior of x and u. The first two columns of the table show the four possible patterns of stationarity and nonstationarity for x and u. The only model that is not plausible is the second line of the table, when x is stationary but nonstationarity in the error makes the dependent variable nonstationary. It is hopeless to attempt to explain a nonstationary variable with regressors that strictly stationary—they cannot capture the wandering over time that will occur in y.

The first case is the one we examined in Chapter 3. It can be estimated with the distributed-lag models discussed there, possibly corrected for (stationary) serial correlation in the error term. The third case is the spurious-regression model, where estimating the model in first-differences is appropriate. The final case is the cointegration case that we have just examined, where the appropriate estimation technique is the error-correction model.

We now know how to deal with nonstationarity if it arises in our regression models, but one key question remains: How do we determine if a variable is stationary or nonstationary? We now turn to this question.

Table 4-2. Regression methods with nonstationary variables

Regressor	Error	Dependent variable	Proper estimation method
x is $I(0)$	u is $I(0)$	y will be $I(0)$	Estimate with standard distributed-lag model.
x is $I(0)$	u is $I(1)$	y will be $I(1)$	Misspecified model. Cannot explain a nonstationary dependent variable with stationary regressors.
x is $I(1)$	u is $I(1)$	y will be $I(1)$	First-difference model to make all variables stationary, then use standard distributed-lag models.
x is $I(1)$	u is $I(0)$	y will be $I(1)$	Variables x and y are cointegrated. Use error-correction model.

4.4 Testing for stationarity and cointegration

4.4.1 Dickey-Fuller test

A stationary variable tends to return to a fixed mean after being disturbed by a shock. We sometimes use the adjective ***mean-reverting*** as a synonym for stationary. This tendency to revert back to the mean is the intuitive basis for the oldest and most basic test for stationarity: the ***Dickey-Fuller test***.

Suppose that we characterize a variable y by the basic AR(1) process

$$y_t = \rho y_{t-1} + u_t \,, \tag{4.12}$$

where u_t is assumed to be stationary. y is $I(1)$ if $\rho = 1$ and $I(0)$ if $\rho < 1$, so we formulate the null hypothesis $H_0 : \rho = 1$ to be non-stationarity and the one-sided alternative hypothesis $H_1 : \rho < 1$ to be stationarity.

It looks like we could just estimate (4.12) by OLS and use the conventional t test to examine the null hypothesis of nonstationarity, but remember the problem with spurious regressions. Under the null hypothesis, y and y_{t-1} are nonstationary, so the t statistic will be inflated and unreliable. Instead, we subtract y_{t-1} from both sides to get

$$\Delta y_t = (\rho - 1) y_{t-1} + u_t = \gamma y_{t-1} + u_t, \tag{4.13}$$

with $\gamma \equiv \rho - 1$. The null hypothesis $\rho = 1$ is now equivalent to $\gamma = 0$ with the alternative $\gamma < 0$. The intuition of equation (4.13) is for a mean-reverting (stationary) process, a high value last period should be associated (on average) with a negative change in the series this period to move it back toward the mean. Thus, if y is stationary, γ should be negative. If y is nonstationary, then there will be no tendency for high values of y in $t - 1$ to be reversed in t, and we should find $\gamma = 0$.

The Dickey-Fuller test statistic is the t statistic of $\hat{\gamma}$ in the OLS regression of (4.13). However, because the regressor is non-stationary under the null, $\hat{\gamma} / \mathrm{s.e.}(\hat{\gamma})$ does not follow the t distribution. Many authors have used Monte Carlo methods to calibrate the distribution of the DF statistic; the critical values for the DF test are more negative than the usual –1.65 that we would use for a one-tailed t test. Hill, Griffiths, and Lim show critical values in Table 12.2 on page 486. If the calculated DF test statistic is less than (i.e., more negative than) the negative critical value, then we reject the null hypothesis and conclude that the variable is stationary. If the test statistic is positive or less negative than the critical value, then we cannot reject the hypothesis that y is nonstationary.

Of course, failing to reject nonstationarity is not the same thing as proving, or even concluding, that y is nonstationary. For series that are very persistent but stationary (such as

AR(1) processes with $\rho > 0.8$), the DF test has very low power, meaning that it often fails to reject false null hypotheses. Thus, deciding that a series is nonstationary based on a marginal failure to reject the DF test can be misleading.

4.4.2 Augmented Dickey-Fuller and Phillips-Perron tests

The DF test is valid if the error term u in equation (4.12) is white noise because then the assumptions of the time-series Gauss-Markov Theorem are satisfied. But we know that error terms in time-series data are usually autocorrelated, and that makes the OLS estimates inefficient and biases the standard errors. We can deal with this problem in either of two ways:

- Try to make the estimates better by eliminating the autocorrelation in u, or
- Correct the standard errors by a process analogous to the Newey-West HAC robust standard errors.

The first correction leads to the ***augmented Dickey-Fuller test*** and is implemented by adding lagged values of Δy to the right-hand side of equation (4.13).[5] Thus, an ADF test with p lags would be a regression

$$\Delta y_t = \gamma y_{t-1} + \beta_1 \Delta y_{t-1} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t, \tag{4.14}$$

where our test statistic is again the t ratio for $\hat{\gamma}$ and we select p to be large enough that the error term ε is white noise. The critical values for the ADF test are different than those for the basic DF test and depend on the number of lags p.

The basic DF and ADF tests are tests of whether a series is a random walk, with the alternative being a stationary AR process. There are variants of these tests that can include "drift" (a random walk with a nonzero mean period-to-period change) or a linear trend. To test for a random walk with drift (against stationarity), we add a constant term to (4.14) to get

$$\Delta y_t = \alpha + \gamma y_{t-1} + \beta_1 \Delta y_{t-1} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t. \tag{4.15}$$

The test statistic for (4.15) is again the t statistic on $\hat{\gamma}$, but the critical values are different in the presence of a constant term than in (4.14). To test whether the series is $I(1)$ against the alternative that it is stationarity around a fixed. linear trend we add a trend term along with the constant:

$$\Delta y_t = \alpha + \delta t + \gamma y_{t-1} + \beta_1 \Delta y_{t-1} + \ldots + \beta_p \Delta y_{t-p} + \varepsilon_t.$$

Once again, we must use a different table of critical values when including the trend.

[5] There is a close relationship between lagged dependent variables and serial correlation of the error. Adding lagged dependent variables as regressors can be an effective alternative to using a transformation such as Prais-Winsten to correct the error term.

Rather than trying to figure out how many lags should be included in the ADF specification, the ***Phillips-Perron test*** uses the OLS t test from (4.13) and uses the Newey-West procedure to correct the standard errors for autocorrelation in u.

Stata implements the ADF and Phillips-Perron tests with the commands dfuller and pperron, respectively. In both cases, one can specify the presence or absence of a constant and/or trend (the default is to include a constant but not a trend) and the number of lags (of Δy in the ADF and in the Newey-West approximation in Phillips-Perron). Stata will show 67pprox.imate critical values for the test, tailored to the particular specification used, so you should not need to refer to any textbook tables.

4.4.3 DF-GLS test

We noted above that the ADF and Phillips-Perron tests find it difficult to distinguish between $\rho = 1$ (nonstationarity) and ρ just less than one (stationary but persistent). Because many stationary series are persistent, this low power is problematic. Stock and Watson recommend an alternative test as being more powerful in these cases. The ***DF-GLS test*** is a Dickey-Fuller test in which the variables are "quasi-differenced" in a manner similar to the Prais-Winsten GLS procedure we use for stationary AR(1) error terms.

To implement this test, we first use the following formulas to create two quasi-differenced series, z which is a transformed version of y and x_1, which is analogous to a constant:

$$z_t = \begin{cases} y_t, \text{ for } t = 1, \\ y_t - \left(1 - \frac{7}{T}\right) y_{t-1}, \text{ for } t = 2, 3, \ldots, T. \end{cases}$$

$$x_{1t} = \begin{cases} 1, \text{ for } t = 1, \\ \frac{7}{T}, \text{ for } t = 2, 3, \ldots, T. \end{cases}$$

We then regress $z_t = \delta_0 x_{1t} + u_t$ with no constant term (because x_1 is essentially a constant) and calculate a "detrended" y series as $y_t^d \equiv y_t - \hat{\delta}_0$. We then apply the Dickey-Fuller test to the detrended series y^d using critical values developed for the DF-GLS test. No DF-GLS procedure is included in Stata, but there is a command dfgls that can be downloaded from the online Stata libraries that implements it.

4.4.4 Testing for cointegration

Testing for cointegration is simply testing the stationarity of the error term in the cointegrating equation. When Engle and Granger first explored cointegrated models, their first test, the ***Engle-Granger test***, simply applies the ADF test to the residuals of the cointegrating regression (4.9). Because these are estimated residuals rather than a free-standing time series, yet another set of custom critical values must be used for this test. A more recent test, the

Johansen-Juselius test is more general, and tests for the possibility of multiple cointegrating relationships when there are more than two variables. This test is integrated into the procedure for vector error-correction models in Stata. We study these models in Chapter 5.

References

Granger, C. W. J., and Paul Newbold. 1974. Spurious Regressions in Econometrics. *Journal of Econometrics* 2 (2):111-120.

Stock, James H., and Mark Watson. 2011. *Introduction to Econometrics*. 3rd ed. Boston: Pearson Education/Addison Wesley.