CHAPTER 2

Regression with Stationary Time Series

2.1 Spurious Regressions: Why Stationarity Is Important

For many decades, economists (particularly macroeconomists) ran time-series regressions based on the Gauss-Markov methodology that we studied earlier. The results appeared to be remarkable! *R*-squared values were commonly over 0.95 and often in the neighborhood of 0.999. The *t* statistics testing the null hypothesis that a coefficient was zero were nearly always significant and often in excess of 10. The regression methodology seemed to find great application in time-series analysis using aggregate data.

However, most of the time series that macroeconomists of the time used—GDP, the money supply, consumption spending, etc.—contained strong trends. Thinking about it intuitively, it is not surprising that two trended series will tend to be highly and positively correlated: both series are smaller at the beginning of the sample than at the end. This correlation will be strong even if the non-trend movements in the two series are independent.

Consider the time plot in Figure 0-1, which depicts two series: total attendance at American League baseball games and real per-capita GDP from 1960 to 2007. The series seem to be highly correlated, largely due to the strong trends in both. Regressing AL attendance on per-capita GDP yields the regression in the first column of Table 2-1 with an *R*-square value of nearly 0.94 and an overwhelming *t* statistic of 26. An economist seeing column (1) would seem justified in concluding that increases in income per person have very strong and statistically reliable effects on baseball attendance, which seems entirely plausible—except that the regressor in (1) is per-capita GDP in Botswana, not in the United States!

Is American baseball attendance really strongly affected by income in a small, sparsely populated African nation? It seems unlikely, yet the conventional statistical measures of *R*-square and *t* are overwhelming: the *p* value associated with the *t*-statistic is 3×10^{-29}.

This is an example of ***spurious regression***, a term coined by Granger and Newbold (1974) in their seminal article on regressions with nonstationary variables. When the variables in a regression are nonstationary, *R*-square values and *t*-statistics no longer follow the usual distributions and can be wildly inflated. Consider the regression in column (2) of Table 2-1, which relates the first difference of attendance to the first difference of Botswana GDP. In

contrast to the levels equation (1), there is no evidence of a relationship in the differenced regression of column (2), with *R*-square of 0.005 and a *t*-statistic less than 1.

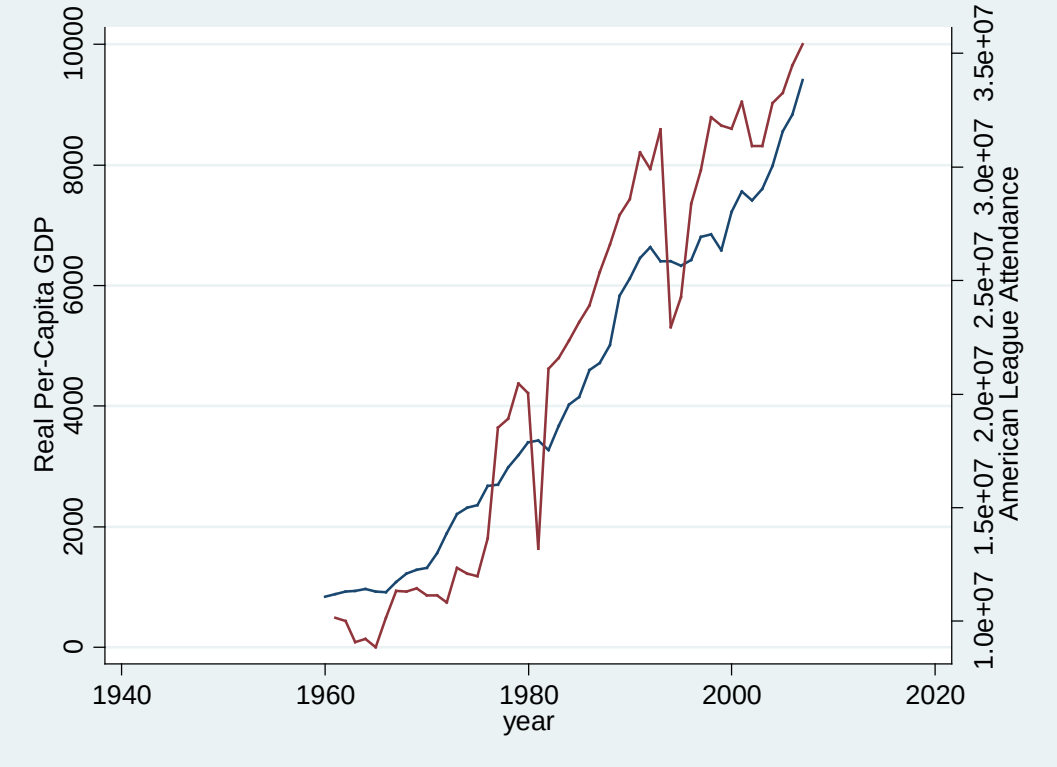

Figure 2-1. Baseball attendance and GDP

	(1) AL Attendance	(2) Δ AL Att.
GDP	3,323***	
	(126.8)	
ΔGDP		765.5
		(1,599)
Constant	7.073e+06***	406,547
	(640,361)	(464,611)
Observations	47	46
R-squared	0.939	0.005

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2-1. Regressions of baseball attendance on GDP

The case for spurious correlation between two strongly trended series as in Figure 2-1 is intuitive. But beyond this, Granger and Newbold demonstrated that nonstationary regression is also unreliable in a less obvious case: random walks with no trend or "drift" that moves the series in the same direction over time. The only thing that these two series have in common is that the (independent) shocks to both series are highly persistent, yet Granger and Newbold's Monte Carlo regressions rejected the null hypothesis of a zero coefficient 76 percent of the time (rather than the appropriate 5 percent) in bivariate regressions. As expected, differenced regressions (which would involve two unrelated white-noise series because the first difference of a random walk is white noise) yielded conventional frequencies of rejection. Moreover, as the sample size gets larger, the problem of spurious regressions with nonstationary variables gets worse, not better. The *t* statistic between unrelated random walks goes to infinity rather than zero as $T \to \infty$.

Much recent research in time-series econometrics has focused on appropriate regression models when the variables are non-stationary. We examine these models in subsequent chapters, but first we adapt our regression model to time-series data assuming that the variables in the regression are all stationary.

2.2 Gauss-Markov Assumptions in Time-Series Regressions

2.2.1 Exogeneity in a time-series context

For cross-section samples, we defined a variable x to be exogenous if for all observations i in the sample, $E(\varepsilon_i \mid x_1, x_2, ..., x_N) = 0$. This means that *each* observation i in the sample must be independent of *every* observation of x, not just the ith observation of x. With independent cross-section observations, the only value of x that is likely to be related to ε_i is x_i corresponding to the same observation.

In a time-series setting, observations that are near each other in time are likely to be related. The ***strict exogeneity*** condition $E(\varepsilon_t \mid ..., x_{t+2}, x_{t+1}, x_t, x_{t-1}, x_{t-2}, ...) = 0$ requires that the regressor's value in period t be unrelated to the disturbance term *in every period*. This means that we cannot have dynamic feedback effects in which the past or future value of the regressor might depend on the current disturbance.

For some properties of time-series regression a weaker form of exogeneity is sufficient. We say that a variable x is ***weakly exogenous*** if $E(\varepsilon_t \mid x_t, x_{t-1}, x_{t-2}, ...) = 0$, which means that the period t disturbance is independent of current and past values of x, but not necessarily independent of future x.

2.2.2 Applying the Gauss-Markov Theorem to time series

For us to apply the Gauss-Markov Theorem to a time-series context, we require the following assumptions:

- **TS-1. Linear model**: The joint data-generating process of y, x_2, …, x_K is

$$y_t = \beta_1 + \beta_2 x_{t,2} + \ldots + \beta_K x_{t,K} + u_t,\ t = 1, 2, \ldots, T,$$

with u_t a sequence of disturbances.

- **TS-2. Strict exogeneity**: The explanatory variables $x_{.j}$ are strictly exogenous with respect to the disturbance term. Mathematically,

$$E(u_t \mid \mathbf{X}) = 0,\ t = 1, 2, \ldots, T,$$

where $\mathbf{X}$ includes all $K-1$ regressors and all T time periods:

$$\mathbf{X} = \begin{bmatrix} x_{1,2} & x_{1,3} & \cdots & x_{1,K-1} & x_{1,K} \\ x_{2,2} & x_{2,3} & \cdots & x_{2,K-1} & x_{2,K} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{T,2} & x_{T,3} & \cdots & x_{T,K-1} & x_{T,K} \end{bmatrix}.$$

- **TS-3. No perfect collinearity**: No regressor $x_{.j}$ is constant or can be expressed as a linear function of other regressors. In other words, there is no set of constants a_j that are not all zero and for which

$$a_1 + a_2 x_{t,2} + \ldots + a_K x_{t,K} = 0,\ \forall t = 1, 2, \ldots, T.$$

This condition means that $\mathbf{X}$ as defined in TS-2 has full rank $K-1$.

- **TS-4. Homoskedasticity**: The conditional variance of u_t is constant:

$$\operatorname{var}(u_t \mid \mathbf{X}) = \sigma^2,\ t = 1, 2, \ldots, T.$$

- **TS-5. No serial correlation**: The disturbance terms are uncorrelated (conditional on $\mathbf{X}$):

$$\operatorname{cov}(u_t, u_{t-s} \mid \mathbf{X}) = 0,\ s = 1, 2, \ldots, T-1.$$

- **TS-6. Normality**: The disturbance terms are normally distributed,

$$u_t \sim N(0, \sigma^2).$$

When conditions TS-1 through TS-3 hold, the OLS coefficient estimator is unbiased. When we add TS-4 and TS-5 we obtain the unbiasedness of the standard OLS estimator of the variance of the OLS coefficient estimator, so we can use the standard tools of OLS inference. Under TS-1 through TS-5, the Gauss-Markov Theorem assures that OLS is BLUE. If we add TS-6, then the OLS coefficient vector has a normal distribution and the ratio of each coefficient to its standard error has a t distribution.

Thus it appears straightforward to extend our previous analysis to a time-series setting. However, the assumptions that are often reasonable when we draw plausibly independent observations from a cross-sectional sample frequently fail to hold for sequential, time-series observations. In particular, strict exogeneity is rare and serial correlation is common, so assumptions TS-2 and TS-5 often fail. Strict exogeneity requires that every regressor x be independent of the entire history of disturbances to y—past, present, and future. No serial correlation requires that positive shocks to y in period t neither tend to persist into period $t + 1$ (leading to positive correlation) nor tend to be followed by negative "corrections" in period $t + 1$ (leading to negative correlation).

The focus in time-series regression analysis is mainly addressed to coping with violations of TS-2 and TS-5. If the variables in our model are stationary and ergodic, we can loosen TS-2 to require only weak exogeneity and our OLS estimator will still have desirable asymptotic properties. Coping with serial correlation is discussed in the next section.

2.2.3 Asymptotic properties of OLS

In cross-section samples, the law of large numbers assures us that random variations will tend to "even out" with a large enough sample of independent observations. This happens because the sample means, variances, and covariances among the variables converge to fixed, finite population moments such as

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}x_i=\mu_x,\\
&\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2=\sigma_x^2,\\
&\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(x_{i,j}-\bar{x}_j\right)\left(x_{i,k}-\bar{x}_k\right)=\sigma_{x_jx_k}.
\end{aligned}\tag{2.1}$$

To see why asymptotic properties require careful attention in time-series models, consider what happens to the equations in (2.1) when the x variable is a time trend so that $x_t = t$. The mean of a sample of T observations is

$$\begin{aligned}
\bar{x}_T &\equiv \frac{1}{T}\sum_{t=1}^{T}x_t\\
&=\frac{1}{T}\sum_{t=1}^{T}t=\frac{1}{T}\frac{T\left(T+1\right)}{2}=\frac{T+1}{2},
\end{aligned}$$

and

$$\lim_{T\to\infty}\bar{x}_T=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}x_t=\lim_{T\to\infty}\frac{T+1}{2}=\infty.$$

Therefore the first moment-convergence condition in (2.1) fails when the regressor is a time trend. It is straightforward to show that the second condition in (2.1) also fails for the time trend—the sample variance also diverges as T gets large.

This problem is not restricted to time trends. It occurs with any nonstationary variable because the mean and/or variance do not converge in large samples. This is the basis for the spurious-regression problem with which we began this chapter.

A more subtle difficulty occurs with non-ergodic variables. Recall that these variables have "long memory," so that observations that are far apart in time are still strongly correlated. What this means for regression analysis is that even as we accumulate a large number of observations, the amount of new information in those observations is limited by their correlation with the earlier ones. Intuitively, this means that the information in the sample does not grow fast enough as the sample size increases to lead to asymptotic convergence of estimators to the true parameter values.

If—and it's a big if—we are working with stationary and ergodic time series, then we can weaken the strict exogeneity assumption TS-2 to weak exogeneity and the OLS estimators still have desirable asymptotic properties. Corresponding to the set of assumptions TS-1 through TS-6 that support the small-sample properties of the OLS estimators, we have ATS-1 through ATS-6 that foster consistency, asymptotic efficiency, and asymptotic normality.

- **ATS-1. Linear model**: The joint data-generating process of y, x_2, …, x_K is

  $$y_t = \beta_1 + \beta_2 x_{t,2} + \ldots + \beta_K x_{t,K} + u_t,\ t = 1, 2, \ldots, T,$$

  with u_t a sequence of disturbances. (Identical to TS-1.)

- **ATS-2. Weak exogeneity, stationarity, and ergodicity**: The variables of the model are stationary and ergodic, and the explanatory variables $x_{\cdot,j}$ are weakly exogenous with respect to the disturbance term. Mathematically,

  $$E\left(u_t \mid x_{t,2},\, x_{t,3},\, \ldots,\, x_{t,K}\right) = 0.$$

  Note that this requires u_t to be independent only of the *current* values of the regressors, not of all past, current, and future values.

- **ATS-3. No perfect collinearity**: No regressor $x_{\cdot,j}$ is constant or can be expressed as a linear function of other regressors. In other words, there is no set of constants a_j that are not all zero and for which

  $$a_1 + a_2 x_{t,2} + \ldots + a_K x_{t,K} = 0,\ \forall t = 1, 2, \ldots, T.$$

  (Identical to TS-3.)

- **ATS-4. Homoskedasticity**: The conditional variance of ε_t is constant:

$$\text{var}\left(u_t \mid x_{t,2}, x_{t,3}, ..., x_{t,K}\right) = \sigma^2, \; t = 1, 2, ..., T.$$

(Nearly identical to TS-4 except conditioning is only on current *x*.)

- **ATS-5. No serial correlation**: The disturbance terms are uncorrelated (conditional on **X**):

$$\text{cov}\left(u_t, u_{t-s} \mid x_{t,2}, x_{t,3}, ..., x_{t,K}\right) = 0, \; s = 1, 2, ..., T-1.$$

(Nearly identical to TS-5 except conditioning is only on current *x*.)

Under assumptions ATS-1 through ATS-3, the OLS estimator is consistent. If we add assumptions ATS-4 and ATS-5, it is asymptotically efficient and asymptotically normal. Note that we do not require an asymptotic analog to TS-6, which imposed normality of *u*. That is because assumptions ATS-1 through ATS-5 allow us to use central limit theorems to show that the OLS estimators will converge to a normal distribution in large samples. This allows us to use the normal distribution to assess the asymptotic significance of our *t* statistics and the chi-square distribution to evaluate the asymptotic distribution of *F* statistics.

2.3 Regression with Serially Correlated Errors

2.3.1 Implications of serial correlation

In discussing the small-sample properties of OLS with time-series data, we identified TS-2 (strict exogeneity) and TS-5 (no serial correlation) as assumptions that were often violated in economic data. We then showed that we can relax TS-2 to the more data-friendly ATS-2 (weak exogeneity) if our variables are stationary and ergodic and if our sample is large enough that asymptotic distributions are reasonably close approximations. In the remainder of this chapter we assume that stationarity and ergodicity assumptions in ATS-2 are valid.

We now consider the issue of TS-5 or ATS-5. In economic data, what happens at *t* is often related to what happened at $t - 1$. If that is true of the disturbance terms in our regression, then we have serial correlation. We know from our discussion above that the consistency of the OLS coefficient estimator requires only ATS-1 through ATS-3, so it does not depend on the absence of serial correlation in the disturbance. However, in the presence of serial correlation the variance of the OLS estimator will be larger than some other estimators, so it is not efficient, and the traditional OLS variance estimator will not estimate the true variance accurately so our OLS test statistics will not follow the same distribution that they would if ATS-5 were not violated.

To see the problem, let us suppose that assumptions ATS-1 through ATS-4 are satisfied, but that the disturbance term *u* follows an AR(1) process:

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{2.2}$$

with $|\rho| < 1$ and ε being white noise with variance σ_ε^2. For simplicity, let $K = 2$ so there is only one (non-constant) regressor: $y_t = \beta_1 + \beta_2 x_t + u_t$. [This follows Wooldridge (2009, 409).]

Following the standard derivation, the OLS estimator of β_2 is

$$\hat{\beta}_2 = \frac{\sum_{t=1}^{T}(y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} = \beta_2 + \frac{\sum_{t=1}^{T} u_t (x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}.$$

Just to simplify this expression, assume that we have normalized the regressor so that $\bar{x} = 0$. (This is not a restrictive assumption; it just makes the algebra more transparent.) In this case,

$$\hat{\beta}_2 - \beta_2 = \frac{\sum_{t=1}^{T} u_t x_t}{\sum_{t=1}^{T} x_t^2},$$

and

$$\operatorname{var}(\hat{\beta}_2 \mid \mathbf{X}) = \frac{1}{\left(\sum_{t=1}^{T} x_t^2\right)^2} E\left[\left(\sum_{t=1}^{T} u_t x_t\right)^2 \mid \mathbf{X}\right].$$

In general, the squared-summation term involves all of the cross-products of the u_t terms with u_{t-s} terms. In the special case of no serial correlation, all of the cross-products are zero and the expectation expression reduces to $\sigma^2 \sum_{t=1}^{T} x_t^2$, leading to our usual OLS formula for the variance. But this does not happen if the disturbance terms are serially correlated.

In the general case,

$$\begin{aligned}
E\left[\left(\sum_{t=1}^{T} u_t x_t\right)^2 \mid \mathbf{X}\right] &= E\left[\left(\sum_{t=1}^{T} x_t^2 u_t^2 + 2\sum_{t=1}^{T-1}\sum_{s=1}^{T-t} x_t x_{t+s} u_t u_{t+s}\right) \mid \mathbf{X}\right] \\
&= \sum_{t=1}^{T} x_t^2 E\left(u_t^2 \mid \mathbf{X}\right) + 2\sum_{t=1}^{T-1}\sum_{s=1}^{T-t} x_t x_{t+s} E\left(u_t u_{t+s} \mid \mathbf{X}\right) \\
&= \sum_{t=1}^{T} x_t^2 \operatorname{var}\left(u_t \mid \mathbf{X}\right) + 2\sum_{t=1}^{T-1}\sum_{s=1}^{T-t} x_t x_{t+s} \operatorname{cov}\left(u_t, u_{t+s} \mid \mathbf{X}\right) \\
&= \sigma_u^2 \sum_{t=1}^{T} x_t^2 + 2\sigma_u^2 \sum_{t=1}^{T-1}\sum_{s=1}^{T-t} x_t x_{t+s} \operatorname{corr}\left(u_t, u_{t+s} \mid \mathbf{X}\right) \\
&= \sigma_u^2 \sum_{t=1}^{T} x_t^2 + 2\sigma_u^2 \sum_{t=1}^{T-1}\sum_{s=1}^{T-t} \rho^s x_t x_{t+s}.
\end{aligned}$$

The final expression uses the property that corr(u_t, u_{t+s}) = ρ^s for an AR(1) process. Substituting this expression into the variance formula yields

$$\begin{aligned}\operatorname{var}\left(\hat{\beta}_2 \mid \mathbf{X}\right) &= \frac{\sigma_u^2 \sum_{t=1}^{T} x_t^2 + 2\sigma_u^2 \sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \rho^s x_t x_{t+s}}{\left(\sum_{t=1}^{T} x_t^2\right)^2} \\ &= \frac{\sigma_u^2}{\sum_{t=1}^{T} x_t^2} + 2\sigma_u^2 \frac{\sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \rho^s x_t x_{t+s}}{\left(\sum_{t=1}^{T} x_t^2\right)^2}.\end{aligned} \tag{2.3}$$

The standard expression for the OLS variance (assuming no serial correlation) is just the first term of (2.3) and neglects the second. The second term is zero only if either the disturbance term is not serially correlated ($\rho = 0$) or the regressor is not serially correlated in our sample (in which case the cross-product terms would add up to zero). In economic data it is common for $\rho > 0$ and for x to be positively serially correlated as well, which means that the second term in (2.3) is likely to be positive and the true variance of the OLS estimator will be larger than when there is no serial correlation.

Because the usual OLS standard errors neglect the second term, they will generally be inconsistent in the presence of serial correlation, meaning that our t and F statistics based on them will not be valid. We will see below that there are two methods for dealing with a serially correlated disturbance: we can try to transform the model to eliminate the serial correlation or we can use the (inefficient) OLS estimator and correct the standard errors to reflect the second term in (2.3).

2.3.2 Testing for serial correlation

Serial correlation in the error occurs when the condition $\operatorname{corr}\left(u_t, u_{t-s}\right) = 0$ is violated for some s. We can test the null hypothesis of no serial correlation if we have estimators of the error terms u_t that are consistent (when the null hypothesis is true). The OLS residuals $\hat{u}_t$ are the obvious choice, so our tests for serial correlation will involve testing for correlation between $\hat{u}_t$ and $\hat{u}_{t-s}$ for positive values of s up to some chosen limit p. The null hypothesis is therefore $\operatorname{corr}\left(u_t, u_{t-s}\right) = 0,\ s = 1, 2, \ldots, p$.

The oldest test for (first-order) serial correlation is the Durbin-Watson test. This test has fallen into disuse for three reasons. First, the critical values of the test statistic depend on the regressors in the model, so they cannot be tabulated for a general case. Users of the Durbin-Watson test traditionally relied on upper and lower bounds for the critical values, meaning that it was impossible to draw a conclusion for calculated test statistics lying in the interval between the bounds. Second, the Durbin-Watson test has been shown to be invalid if a lagged dependent variable is among the regressors. This rules out its use for any model with

an autoregressive structure of y. Third, the Durbin-Watson statistic only tests for first-order serial correlation and cannot easily be extended to $p > 1$.

The ***Breusch-Godfrey Lagrange multiplier test*** is a regression-based test for order-p autocorrelation of the disturbance. The null hypothesis is that the disturbance is white noise. If the disturbance is white noise, then the current OLS residual $\hat{u}_t$ should be independent of the lagged residuals $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots$. The Breusch-Godfrey test uses the regression

$$\hat{u}_t = \gamma_1 \hat{u}_{t-1} + \ldots + \gamma_p \hat{u}_{t-p} + \mathbf{x}_{t.}\boldsymbol{\beta} + v_t, \tag{2.4}$$

where $\mathbf{x}_{t.}$ is the row vector of all explanatory variables in the model corresponding to period t (including the constant) and $\boldsymbol{\beta}$ is the coefficient vector of the model.

The dependent variable in regression (2.4) is the residual $\hat{u}_t$ from the regression of y_t on $\mathbf{x}_{t.}$, which is by construction uncorrelated with $\mathbf{x}_{t.}$. If there is no autocorrelation in u, then the p lagged residuals should also be uncorrelated with $\hat{u}_t$. Thus, under the null hypothesis of no serial correlation the regressors in (2.4) should have no explanatory power whatsoever for the dependent variable; if they do explain $\hat{u}_t$, then there must be serial correlation.

The Breusch-Godfrey test statistic for autocorrelation of order 1 through p is

$$BG(p) = T_0 \times R^2,$$

where T_0 is the number of observations and R^2 is the coefficient of determination from regression (2.4). $BG(p)$ is asymptotically distributed as a $\chi^2(p)$ variable under the null hypothesis. No autocorrelation and no explanatory power in (2.4) would imply an R^2 near zero and a small test statistic, failing to reject the null hypothesis of white noise.

If there are T observations in the original model, then we have values for $\hat{u}_t$ and $\mathbf{x}_{t.}$ for all of these T observations. However, we cannot observe $\hat{u}_{t-1}, \ldots, \hat{u}_{t-p}$ for the first p observations. There are two strategies for handling this difficulty. We can either omit these observations and estimate (2.4) for the $T - p$ observations $t = p + 1, p + 2, \ldots, T$ or we can use all T observations and substitute zero (the expected value) for the missing lagged residuals. In either case, the T_0 used to calculate $BG(p)$ is the number of observations used in (2.4), $T_0 = T - p$ if the observations are dropped and $T_0 = T$ if zero is substituted.

To perform the Breusch-Godfrey test in Stata, we can use the `estat bgodfrey` command. The order of autocorrelation to be tested (p) is specified by the `lags( )` option. By default, Stata substitutes zero for the missing lagged residuals and uses the full sample. To use the shorter sample without substituting, we specify the `nomiss0` option.

An alternative to the Breusch-Godfrey test is the ***Box-Ljung "portmanteau" test*** or ***Q test***. The Q test can be applied to any time series, not just regression residuals, and is based on correlation coefficients. When applied to the residuals $\hat{u}$, the test statistic is

$$Q(p) = T(T+2)\sum_{s=1}^{p}\frac{r_s^2}{T-j},$$

where

$$r_s \equiv \frac{\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}}{\sum_{t=1}^{T}\hat{u}_t^2}$$

is the sample autocorrelation coefficient at lag *s*. Like the Breusch-Godfrey test statistic, *Q*(*p*) converges asymptotically to a χ_p^2 distribution. The *Q* test and the Breusch-Godfrey test are asymptotically equivalent under the null hypothesis if there are no lagged dependent variables among the regressors, but Greene (2012, 923) argues that the Breusch-Godfrey test is more powerful because it controls for the regressors when testing for correlation between the current and lagged residuals.

The Stata command `wntestq` implements the *Q* test with the option `lags( )` specifying *p*. The residuals must be explicitly retrieved and included in the command. For example, to test the first 12 autocorrelations of a residual series stored in `uhat`, we would type `wntestq uhat , lags(12)`. The general command `corrgram` for calculating autocorrelations also shows the successive *Q* statistics at each order *p* up to the limit specified. It is important to remember that *Q* (*p*) (like the Breusch-Godfrey test) tests the joint null hypothesis that the first *p* autocorrelations are zero, *not* the simple null that the individual *p*th-order autocorrelation is zero.

2.3.3 Generalized least squares estimation of models with autoregressive errors

Generalized least squares estimation allows us to transform a model whose error term has a non-classical distribution into one whose error term follows the classical assumptions. In the case of autoregressive error terms, the transformation consists of a "quasi-differencing" filter that purges the error term of autocorrelation. If u_t is an order-*p* autoregressive process

$$u_t = \rho_1 u_{t-1} + \ldots + \rho_p u_{t-p} + \varepsilon_t,$$

where ε_t is white noise, then

$$u_t^* \equiv u_t - \rho_1 u_{t-1} - \ldots - \rho_p u_{t-p} = \varepsilon_t$$

has no autocorrelation.

The most common case is $p = 1$, where the error term follows an AR(1) process. In this case, $u_t^* = u_t - \rho_1 u_{t-1} = (1-\rho_1 L)u_t = \varepsilon_t$. If we begin with the model $y_t = \alpha + \beta x_t + u_t$ and apply

the quasi-differencing transformation $(1-\rho_1 L)$ to each term on both sides of the equation, we get

$$\begin{aligned} y_t - \rho_1 y_{t-1} &= (\alpha + \beta x_t + u_t) - \rho_1 (\alpha + \beta x_{t-1} + u_{t-1}) \\ &= \alpha(1-\rho_1) + \beta(x_t - \rho_1 x_{t-1}) + (u_t - \rho_1 u_{t-1}), \end{aligned}$$

or

$$y_t^* = \alpha c_t^* + \beta x_t^* + u_t^*, \tag{2.5}$$

where

$$\begin{aligned} y_t^* &\equiv y_t - \rho_1 y_{t-1}, \\ c_t^* &\equiv 1 - \rho_1, \\ x_t^* &\equiv x_t - \rho_1 x_{t-1}, \\ u_t^* &\equiv u_t - \rho_1 u_{t-1} = \varepsilon_t. \end{aligned} \tag{2.6}$$

The transformed model (2.5) has an error term $u_t^* = \varepsilon_t$ that is serially uncorrelated, so it can be estimated efficiently by OLS. But in order to apply GLS to this model we must solve two problems: (1) what to do about the first observation and (2) how to obtain an estimate of ρ_1.

The first-observation problem is that we cannot use the transformation in (2.6) when $t = 1$ because we generally do not have observations for y_0 and x_0. There are two choices for dealing with this problem. We can omit the first observation and estimate (2.5) for $T-1$ observations starting with $t = 2$, but this solution loses the information from the omitted observation, which may be significant in small samples.[1] Alternatively, we can include the first observation, but if we were to add this one untransformed observation to the $T-1$ transformed observations, it would have a different error variance. An untransformed observation has variance $\text{var}(u) = \sigma_u^2$. The transformed observations have a smaller variance equal to $\text{var}(\varepsilon) = (1-\rho_1^2)\sigma_u^2$. Thus, to make the variance of the untransformed initial observation match the transformed observations we must calculate

$$\begin{aligned} y_1^* &= \sqrt{1-\rho_1^2}\, y_1, \\ c_1^* &= \sqrt{1-\rho_1^2}, \\ x_1^* &= \sqrt{1-\rho_1^2}\, x_1, \\ u_1^* &= \sqrt{1-\rho_1^2}\, u_1 \end{aligned} \tag{2.7}$$

[1] If $p > 1$, the transformations in (2.6) have p lags and we would lose p observations from the beginning of the sample.

in order to add the first observation into the transformed estimating sample.

To perform feasible GLS in the AR(1) model, we require an estimate of ρ_1. The most common way of estimating ρ_1 is using the residuals $\hat{u}$ from an OLS regression to approximate u and then calculating ρ_1 either as the correlation coefficient between $\hat{u}_t$ and $\hat{u}_{t-1}$ or as the estimated coefficient in a regression $\hat{u}_t = \rho_1 \hat{u}_{t-1} + \varepsilon_t$. The ***Prais-Winsten estimator*** uses this method to estimate ρ_1 and includes all T observations using (2.7) for $t = 1$. The ***Cochrane-Orcutt estimator*** estimates ρ_1 in the same way but omits the first observation. The two estimators are asymptotically equivalent because the importance of the first observation becomes negligible as T gets large. Because Prais-Winsten is more efficient in small samples, we shall not consider the Cochrane-Orcutt estimator further.

Although it is asymptotically efficient as described above, the Prais-Winsten estimator can be iterated, which may improve efficiency in small samples. Prais-Winsten calculates ρ_1 based on OLS residuals. The OLS coefficient estimators that are used to compute these residuals are inefficient due to the autocorrelation that we are trying to correct, thus the residuals are inefficient estimators of the true error term. Correcting for autocorrelation improves the efficiency of the coefficient estimators, therefore it should produce better estimates of the residuals, which can then produce a more efficient estimate of ρ_1. Iterative Prais-Winsten estimation recalculates ρ_1 based on the FGLS residuals, then uses (2.6) and (2.7) to recalculate the transformed variables and calculate more efficient coefficient estimators from a regression of (2.5) on the new transformed variables. Of course, because these estimated coefficients are better than the original FGLS estimates, they can produce *even better* residuals and estimates of ρ_1, and so on and so on and so on. The process stops when the estimates converge: successive iterations produce estimates that differ by less than a specified convergence criterion.

The efficiency of the Prais-Winsten estimator depends on having a consistent estimator of ρ_1. This, in turn, depends on having residuals that are consistent estimators of the error term, which requires consistent initial estimates of the coefficients to use in computing the residuals. An important case in which the OLS coefficient estimators and residuals are inconsistent occurs when there is a lagged dependent variable y_{t-1} among the regressors and the error is autocorrelated. In this case we cannot use OLS residuals $\hat{u}$ to estimate ρ_1 so we need an alternative to Prais-Winsten. The ***Hildreth-Lu estimator*** searches over the stationary region $-1 \leq \rho_1 \leq 1$ to find the value of ρ_1 that minimizes the sum of squared residuals in the transformed regression, thus bypassing the use of OLS altogether.

To estimate FGLS models with AR(1) error terms in Stata, we use the `prais` command, which has a format similar to `regress`. The default method is iterated Prais-Winsten estimation. We can alter the method used by specifying options: `corc` produces Cochrane-Orcutt estimates; `twostep` disables iteration; `ssesearch` specifies the Hildreth-Lu search procedure.

2.3.4 Newey-West HAC-robust standard errors

As noted above, autocorrelation introduces two problems with the OLS model: inefficient estimators and inconsistent standard errors. GLS methods seek to achieve efficient estimation of parameters in the presence of autocorrelated errors. An alternative is to accept the inefficiency of the OLS estimators but correct the standard errors so that valid inference can be performed.

The ***Newey-West HAC-robust standard errors*** for the OLS estimators are consistent when the error term is heteroskedastic, autocorrelated, or both, as long as the regressors are stationary and ergodic. These robust standard errors are kin to White's heteroskedasticity-robust standard errors, but the formulas are more complex.

We begin with the standard regression model (with one variable, for simplicity):

$$y_t = \alpha + \beta x_t + u_t, \tag{2.8}$$

where $\text{cov}(u_t, u_{t-s}) \neq 0$ for $s \neq 0$. For this derivation, we assume that $\text{var}(u_t) = \sigma_u^2$ for all observations, although the Newey-West standard errors are consistent even when the error term is heteroskedastic.

The derivation of equation (2.3) suggests how the OLS standard errors can be corrected to account for the presence of serial correlation . [The derivation here is based on Stock and Watson (2011, 595-600).] We know that the OLS estimator for β in equation (2.8) can be written

$$b = \beta + \frac{\frac{1}{T}\sum_{i=1}^{T}(x_t - \bar{x})u_t}{\frac{1}{T}\sum_{i=1}^{T}(x_t - \bar{x})^2}.$$

Recall that probability limits are very forgiving in that $\text{plim}\left[f(x)\right] = f\left[\text{plim}(x)\right]$ quite generally. We know that $\text{plim}\,\bar{x} = \mu_x$ and, if x is stationary and ergodic, then

$$\text{plim}\left(\frac{1}{T}\sum_{i=1}^{T}(x_t - \bar{x})^2\right) = \sigma_x^2.$$

Therefore,

$$\text{plim}(b - \beta) = \frac{\text{plim}\left(\frac{1}{T}\sum_{i=1}^{T}(x_t - \mu_x)u_t\right)}{\sigma_x^2} = \frac{\text{plim}(\bar{v})}{\sigma_x^2},$$

where $v_t \equiv (x_t - \mu_x)u_t$ and $\bar{v} = \frac{1}{T}\sum_{t=1}^{T} v_t$.

In large samples where we can assume that b is arbitrarily close to plim (b),

$$\text{var}(b) = \text{var}\left(\frac{\bar{v}}{\sigma_x^2}\right) = \frac{\text{var}(\bar{v})}{\sigma_x^4}.$$

If there is no autocorrelation in u, then the variance of the average $\bar{v}$ is just

$$\text{var}(\bar{v}) = \frac{1}{T}\text{var}(v_t) = \frac{\sigma_v^2}{T},$$

which simplifies to the usual formula for the OLS standard error. But with serial correlation, when we take the variance of the sum in $\bar{v}$ the covariance terms are *not* zero.

In the case where there is serial correlation we have to take into account the covariance of the v_t terms:

$$\begin{aligned}
\text{var}(\bar{v}) &= \text{var}\left(\frac{v_1 + v_2 + \ldots + v_T}{T}\right) \\
&= \frac{1}{T^2}\left[\sum_{i=1}^{T}\sum_{j=1}^{T} E(v_i v_j)\right] \\
&= \frac{1}{T^2}\sum_{i=1}^{T}\left(\text{var}(v_i) + \sum_{j\neq i}\text{cov}(v_i, v_j)\right) \\
&= \frac{1}{T^2}\left[T\,\text{var}(v_t) + 2(T-1)\text{cov}(v_t, v_{t-1}) + 2(T-2)\text{cov}(v_t, v_{t-2}) + \ldots + 2\,\text{cov}(v_t, v_{t-(T-1)})\right] \\
&= \frac{\sigma_v^2}{T} f_T,
\end{aligned}$$

where

$$\begin{aligned}
f_T &\equiv 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\text{corr}(v_t, v_{t-j}) \\
&= 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j.
\end{aligned}$$

Thus, $\text{var}(b_2) = \left[\frac{1}{T}\frac{\sigma_v^2}{\sigma_x^4}\right] f_T$, which expresses the variance as the product of the no-autocorrelation variance and the f_T factor that corrects for autocorrelation. In order to implement this, we need to know f_T, which depends on the autocorrelations of v for orders 1 through $T-1$.

In practice, just as in GLS estimation, these are not known and must be estimated. As usual, we use the OLS residuals $\hat{u}_t$ as estimators for u_t, which allows us to compute estimates of v_t for each observation in the sample. For ρ_1 we have lots of information because

there are $T-1$ pairs of values for (v_t, v_{t-1}) in the sample. Similarly, we have $T-2$ pairs to use in estimating ρ_2, $T-3$ pairs for ρ_3, and so on.

As the order of autocorrelation j gets larger, there are fewer and fewer observations from which to estimate ρ_j. When we get to ρ_{T-1}, there is only one pair of observations that are $T-1$ periods apart—namely (u_T, u_1)—on which to base an estimate, so this correlation cannot be calculated at all.

The Newey-West procedure truncates the summation in f_T at some value m, so we estimate the first m autocorrelations of v using the OLS residuals and compute $\hat{f}_T = 1 + 2\sum_{j=1}^{m}\left(\frac{m-j}{m}\right)r_j$. We must choose m to be large enough to provide a reasonable correction but small enough relative to T to allow the ρ values to be estimated well. Stock and Watson suggest choosing $m = 0.75T^{\frac{1}{3}}$ as a reasonable rule of thumb. Greene (2012, 920) suggests $m = T^{1/4}$.

To implement Newey-West HAC-robust standard errors in Stata, we use the command `newey`, which has a format similar to the `regress` command. You must include the option `lags(m)` in order to get the Newey-West estimator; omitting this option causes the procedure to calculate the White heteroskedasticity-robust standard errors without correcting for autocorrelation.

References

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

Greene, William H. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, N.J.: Pearson Education.

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

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*. 4th ed. Mason, O.: South-Western Cengage Learning.