Wooldridge, Introductory Econometrics

Chapter 12: Serial correlation and heteroskedasticity in time series regressions

What will happen if we violate the assumption that the errors are not **serially correlated**, or autocorrelated? We demonstrated that the OLS estimators are unbiased, even in the presence of autocorrelated errors, as long as the explanatory variables are strictly exogenous. This is analogous to our results in the case of heteroskedasticity, where the presence of heteroskedasticity alone does not cause bias nor inconsistency in the OLS point estimates. However, following that parallel argument, we will be concerned with the properties of our interval estimates and hypothesis tests in the presence of autocorrelation.

OLS is no longer BLUE in the presence of serial correlation, and the OLS standard errors and test statistics are no longer valid, even asymptotically. Consider a first-order Markov error process:

$$u_t = \rho u_{t-1} + e_t, \ |\rho| < 1 \tag{1}$$

where the e_t are uncorrelated random variables with mean zero and constant variance. What will be the variance of the OLS slope estimator in a simple (y on x) regression model? For simplicity let us center the x series so that $\bar{x} = 0$. Then the OLS estimator will be:

$$b_1 = \beta_1 + \frac{\sum_{i=1}^{n} x_t u_t}{SST_x} \tag{2}$$

where SST_x is the sum of squares of the x series. In computing the variance of b_1, conditional on x, we must account for the serial correlation in the u process:

$$Var(b_1) = \frac{1}{SST_x^2} Var\left(\sum_{t=1}^{n} x_t u_t\right)$$

$$= \frac{1}{SST_x^2} \left(\begin{array}{l} \sum_{i=1}^{n} x_t^2 Var(u_t) + \\ 2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-1} x_t x_{t-j} E\left(u_t u_{t-j}\right) \end{array} \right)$$

$$= \frac{\sigma^2}{SST_x} + 2\left(\frac{\sigma^2}{SST_x^2}\right) \sum_{t=1}^{n-1} \sum_{j=1}^{n-1} \rho^j x_t x_{t-j}$$

where $\sigma^2 = Var(u_t)$ and we have used the fact that $E\left(u_t u_{t-j}\right) = Cov\left(u_t u_{t-j}\right) = \rho^j \sigma^2$ in the derivation. Notice that the first term in this expression is merely the OLS variance of b_1 in the absence of serial correlation. When will the second term be nonzero? When ρ is nonzero, and the x process itself is auto-correlated, this double summation will have a nonzero value. But since nothing prevents the explanatory variables from exhibiting autocorrelation (and in fact many explanatory variables take on similar values through time) the only way in which this second term will vanish is if ρ is zero, and u is not serially correlated. In the presence of serial correlation, the second term will cause the standard OLS variances of our regression parameters to be biased and inconsistent. In most applications, when serial correlation arises, ρ is positive, so that successive errors are positively correlated. In that case, the second term will be positive as well. Recall that this expression is the true variance of the regression parameter; OLS will only consider the first term. In that case OLS will seriously underestimate the variance of the parameter, and the $t-$statistic will be much too high. If on the other hand ρ is negative–so that successive errors result from an "overshooting" process–then we may not be able to determine the sign of the second term, since odd terms will be negative and even terms will be positive. Surely, though, it will not be zero. Thus the consequence of serial correlation in the errors–particularly if the autocorrelation is positive–will render the standard $t-$ and $F-$statistics useless.

Serial correlation in the presence of lagged dependent variables

A case of particular interest, even in the context of simple y on x regression, is that where the "explanatory variable" is a lagged dependent variable. Suppose that the conditional expectation of y_t is linear in its past value: $E(y_t|y_{t-1}) = \beta_0 + \beta_1 y_{t-1}$. We can always add an error term to this relation, and write it as

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t \tag{3}$$

Let us first assume that the error is "well behaved," i.e. $E(u_t|y_{t-1}) = 0$, so that there is no correlation between the current error and the lagged value of the dependent variable. In this setup the explanatory variable cannot be strictly exogenous, since there is a contemporaneous correlation between y_t and u_t by construction; but in evaluating the consistency of

OLS in this context we are concerned with the correlation between the error and y_{t-1}, not the correlation with y_t, y_{t-2}, and so on. In this case, OLS would still yield unbiased and consistent point estimates, with biased standard errors, as we derived above, even if the u process was serially correlated..

But it is often claimed that the joint presence of a lagged dependent variable and autocorrelated errors, OLS will be inconsistent. This arises, as it happens, from the assumption that the u process in (3) follows a particular autoregressive process, such as the first-order Markov process in (1). If this is the case, then we do have a problem of inconsistency, but it is arising from a different source: the misspecification of the dynamics of the model. If we combine (3) with (1), we really have an $AR(2)$ model for y_t, since we can lag (3) one period and substitute it into (1) to rewrite the model as:

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 y_{t-1} + \rho\left(y_{t-1} - \beta_0 - \beta_1 y_{t-2}\right) + e_t \\
&= \beta_0\left(1-\rho\right) + \left(\beta_1 + \rho\right) y_{t-1} - \rho\beta_1 y_{t-2} + e_t \\
&= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + e_t \qquad (4)
\end{aligned}
$$

so that the conditional expectation of y_t properly depends on two lags of y, not merely one. Thus the estimation of (3) via OLS is indeed inconsistent, but the reason for that inconsistency is that y is correctly modelled as $AR(2)$. The $AR(1)$ model is seen to be a dynamic misspecification of (4); as is always the case, the omission of relevant explanatory variables will cause bias and inconsistency in OLS estimates, especially if the excluded variables are correlated with the included variables. In this case, that correlation will almost surely be meaningful. To arrive at consistent point estimates of this model, we merely need add y_{t-2} to the estimated equation. That does not deal with the inconsistent interval estimates, which will require a different strategy.

Testing for first-order serial correlation

Since the presence of serial correlation invalidates our standard hypothesis tests and interval estimates, we should be concerned about testing for it. First let us consider testing for serial correlation in the $k-$variable regression model with strictly exogenous regressors–which rules out, among other things, lagged dependent variables.

The simplest structure which we might posit for serially correlated errors is $AR(1)$, the first order Markov process, as given in (1). Let us assume that e_t is uncorrelated with the entire past history of the u process, and that e_t is homoskedastic. The null hypothesis is $H_0 : \rho = 0$ in the context of (1). If we could observe the

u process, we could test this hypothesis by estimating (1) directly. Under the maintained assumptions, we can replace the unobservable u_t with the OLS residual v_t. Thus a regression of the OLS residuals on their own lagged values,

$$v_t = \kappa + \rho v_{t-1} + \epsilon_t, t = 2, ...n \quad (5)$$

will yield a $t-$ test. That regression can be run with or without an intercept, and the `robust` option may be used to guard against violations of the homoskedasticity assumption. It is only an asymptotic test, though, and may not have much power in small samples.

A very common strategy in considering the possibility of $AR(1)$ errors is the **Durbin-Watson** test, which is also based on the OLS residuals:

$$DW = \frac{\sum_{t=2}^{n} (v_t - v_{t-1})^2}{\sum_{t=1}^{n} v_t^2} \quad (6)$$

Simple algebra shows that the DW statistic is closely linked to the estimate of ρ from the large-sample test:

$$\begin{aligned} DW &\simeq 2\,(1-\widehat{\rho}) \\ \widehat{\rho} &\simeq 1-\frac{DW}{2} \end{aligned} \tag{7}$$

The relationship is not exact because of the difference between $(n-1)$ terms in the numerator and n terms in the denominator of the DW test. The difficulty with the DW test is that the critical values must be evaluated from a table, since they depend on both the number of regressors (k) and the sample size (n), and are not unique: for a given level of confidence, the table contains two values, d_L and d_U. If the computed value falls below d_L, the null is clearly rejected. If it falls above d_U, there is no cause for rejection. But in the intervening region, the test is inconclusive. The test cannot be used on a model without a constant term, and it is not appropriate if there are any lagged dependent variables. In the presence of one or more lagged dependent variables, an alternative statistic may be used: **Durbin's h** statistic, which merely amounts to augmenting (5) with the explanatory variables from the original regression. This test statistic may readily be calculated in Stata with the `estat durbinalt` command.

Testing for higher-order serial correlation

One of the disadvantages of tests for $AR(1)$ errors is that they consider precisely that alternative hypothesis. In many cases, if there is serial correlation in the error structure, it may manifest itself in a more complex relationship, involving higher-order autocorrelations; e.g. $AR(p)$. A logical extension to the test described in 5) and the Durbin "h" test is the **Breusch-Godfrey** test, which considers the null of nonautocorrelated errors against an alternative that they are $AR(p)$. This can readily be performed by regressing the OLS residuals on p lagged values, as well as the regressors from the original model. The test is the joint null hypothesis that those p coefficients are all zero, which can be considered as another nR^2 Lagrange multiplier (LM) statistic, analogous to White's test for heteroskedasticity. The test may easily be performed in Stata using the `estat bgodfrey` command You must specify the lag order p to indicate the degree of autocorrelation to be considered. If $p = 1$, the test is essentially Durbin's "h" statistic.

An even more general test often employed on time series regression models is the Box-Pierce or Ljung-Box **Q statistic**, or "portmanteau test," which has the null hypothesis that the error process is "white noise," or nonautocorrelated, versus the alternative that it is not well behaved. The "Q" test evaluates the autocorrelation function of the errors, and in that sense is closely related to the Breusch-Godfrey test. That test evaluates the conditional autocorrelations of the residual series, whereas the "Q" statistic uses the unconditional autocorrelations. The "Q" test can be applied to any time series as a test for "white noise," or randomness. For that reason, it is available in Stata as the command `wntestq`. This test is often reported in empirical papers as an indication that the regression models presented therein are reasonably specified.

Any of these tests may be used to evaluate the hypothesis that the errors exhibit serial correlation, or nonindependence. But caution should be exercised when their null hypotheses are rejected. It is very straightforward to demonstrate that serial correlation may be induced by simple misspecification of the equation–for instance, modeling a relationship as linear when it is curvilinear, or when it represents exponential growth. Many time series models are misspecified in terms of inadequate dynamics: that is, the relationship between y and the regressors may involve many lags of the regressors. If those lags are mistakenly omitted, the equation suffers from misspecification bias, and the regression residuals will reflect the missing terms. In this context, a visual inspection of the residuals is often useful. User-written Stata routines such as `tsgraph`, `sparl` and particularly `ofrtplot` should be employed to better understand the dynamics of the regression function. Each may be located and installed with Stata's `ssc` command, and each is well documented with on–line help.

Correcting for serial correlation with strictly exogenous regressors

Since we recognize that OLS cannot provide consistent interval estimates in the presence of autocorrelated errors, how should we proceed? If we have strictly exogenous regressors (in particular, no lagged dependent variables), we may be able to obtain an appropriate estimator through transformation of the model. If the errors follow the $AR(1)$ process in (1), we determine that $Var(u_t) = \sigma_e^2 / \left(1 - \rho^2\right)$. Consider a simple y on x regression with autocorrelated errors following an $AR(1)$ process. Then simple algebra will show that the **quasi-differenced** equation

$$(y_t - \rho y_{t-1}) = (1 - \rho)\,\beta_0 + \beta_1\,(x_t - \rho x_{t-1}) + (u_t - \rho u_{t-1}) \tag{8}$$

will have nonautocorrelated errors, since the error term in this equation is in fact e_t, by assumption well behaved. This transformation can only be applied to observations $2, ..., n$, but we can write down the first observation in static terms to complete that, plugging in a zero value for the time-zero value of u. This extends to any number of explanatory variables, as long as they are strictly exogenous; we just quasi-difference each, and use the quasi-differenced version in an OLS regression.

But how can we employ this strategy when we do not know the value of ρ? It turns out that the **feasible generalized least squares (GLS)** estimator of this model merely replaces ρ with a consistent estimate, $\hat{\rho}$. The resulting model is asymptotically appropriate, even if it lacks small sample properties. We can derive an estimate of ρ from OLS residuals, or from the calculated value of the Durbin-Watson statistic on those residuals. Most commonly, if this technique is employed, we use an algorithm that implements an iterative scheme, revising the estimate of ρ in a number of steps to derive the final results. One common methodology is the **Prais-Winsten** estimator, which makes use of the first observation, transforming it separately. It may be used in Stata via the `prais` command. That same command may also be used to employ the **Cochrane-Orcutt** estimator, a similar iterative technique that ignores the first observation. (In a large sample, it will not matter if one observation is lost). This estimator can be executed using the `corc` option of the `prais` command.

We do not expect these estimators to provide the same point estimates as OLS, as they are working with a fundamentally different model. If they provide similar point estimates, the FGLS estimator is to be preferred, since its standard errors are consistent. However, in the presence of lagged dependent variables, more complicated estimation techniques are required.

An aside on first differencing. An alternative to employing the feasible GLS estimator, in which a value of ρ inside the unit circle is estimated and used to transform the data, would be to **first difference** the data: that is, transform the left and right hand side variables into differences. This would indeed be the proper procedure to follow if it was suspected that the variables possessed a **unit root** in their time series representation. But if the value of ρ in (1) is strictly less than 1 in absolute value, first differencing approximates that value, since differencing is equivalent to imposing $\rho = 1$ on the error process. If the process's ρ is quite different from 1, first differencing is not as good a solution as applying the FGLS estimator.

Also note that if you difference a standard regression equation in y, $x_1, x_2...$ you derive an equation that does not have a constant term. A constant term in an equation in differences corresponds to a linear trend in the levels equation. Unless the levels equation already contains a linear trend, applying differences to that equation should result in a model without a constant term..

Robust inference in the presence of autocorrelation

Just as we utilized the "White" heteroskedasticity-consistent standard errors to deal with heteroskedasticity of unknown form, we may generate estimates of the standard errors that are robust to both heteroskedasticity and autocorrelation. Why would we want to do this rather than explicitly take account of the autocorrelated errors via the feasible generalized least squares estimator described earlier? If we doubt that the explanatory variables may be considered strictly exogenous, then the FGLS estimates will not even be consistent, let alone efficient. Also, FGLS is usually implemented in the context of an AR(1) model, since it is much more complex to apply it to a more complex AR structure. But higher-order autocorrelation in the errors may be quite plausible. Robust methods may take account of that behavior.

The methodology to compute what are often termed heteroskedasticity- and autocorrelation-consistent (**HAC**) standard errors was developed by Newey and West; thus they are often referred to as **Newey-West** standard errors. Unlike the White standard errors, which require no judgment, the Newey-West standard errors must be calculated conditional on a choice of maximum lag. They are calculated from a distributed lag of the OLS residuals, and one must specify the longest lag at which autocovariances are to be computed. Normally a lag length exceeding the periodicity of the data will suffice; e.g. at least 4 for quarterly data, 12 for monthly data, etc. The Newey-West (HAC) standard errors may be readily calculated for any OLS regression using Stata's `newey` command. You must provide the "option" `lag( )`, which specifies the maximum lag order, and your data must be tsset (that is, known to Stata as time series data). Since the

Newey-West formula involves an expression in the squares of the residuals which is identical to White's formula (as well as a second term in the cross-products of the residuals), these robust estimates subsume White's correction. Newey-West standard errors in a time series context are robust to both arbitrary autocorrelation (up to the order of the chosen lag) as well as arbitrary heteroskedasticity.

Heteroskedasticity in the time series context

Heteroskedasticity can also occur in time series regression models; its presence, while not causing bias nor inconsistency in the point estimates, has the usual effect of invalidating the standard errors, $t-$statistics, and $F-$statistics, just as in the cross–sectional case. Since the Newey–West standard error formula subsumes the White (robust) standard error component, if the Newey–West standard errors are computed, they will also be robust to arbitrary departures from homoskedasticity. However, the standard tests for heteroskedasticity assume independence of the errors, so if the errors are serially correlated, those tests will not generally be correct. It thus makes sense to test for serial correlation first (using a heteroskedasticity–robust test if it is suspected), correct for serial correlation, and then apply a test for heteroskedasticity.

In the time series context, it may be quite plausible that if heteroskedasticity—that is, variations in volatility in a time series process—exists, it may itself follow an autoregressive pattern. This can be termed a dynamic form of heteroskedasticity, in which Engle's **ARCH** (autoregressive conditional heteroskedasticity) model applies. The simplest ARCH model may be written as:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 z_t + u_t \\ E\left(u_t^2 | u_{t-1}, u_{t-2}, ...\right) &= E\left(u_t^2 | u_{t-1}\right) = \alpha_0 + \alpha_1 u_{t-1}^2 \end{aligned}$$

The second line is the conditional variance of u_t given that series' past history, assuming that the u process is serially uncorrelated. Since conditional variances must be positive, this only makes sense if $\alpha_0 > 0$ and $\alpha_1 \geq 0$. We can rewrite the second line as:

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + v_t$$

which then appears as an autoregressive model in the squared errors, with stability condition $\alpha_1 < 1$. When $\alpha_1 > 0$, the squared errors contain positive serial correlation, even though the errors themselves do not.

If this sort of process is evident in the regression errors, what are the consequences? First of all, OLS are still BLUE. There are no assumptions on the conditional variance of the error process that would invalidate the use of OLS in this context. But we may want to explicitly model the conditional variance of the error process, since in many financial series the movements of volatility are of key importance (for instance, option pricing via the standard Black–Scholes formula requires an estimate of the volatility of the underlying asset's returns, which may well be time–varying).

Estimation of ARCH models—of which there are now many flavors, with the most common extension being Bollerslev's GARCH (generalised ARCH)—may be performed via Stata's `arch` command. Tests for ARCH, which are based on the squared residuals from an OLS regression, are provided by Stata's `estat archlm` command.