Models, Testing, and Correction of Serial Correlation

James L. Powell
Department of Economics
University of California, Berkeley

First-Order Serial Correlation

The general problem of serially correlated disturbances in the linear regression model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

is to find methods to accommodate lack of independence in the disturbances across observations, i.e., $C(\varepsilon_t, \varepsilon_s) \neq 0$ if $t \neq s$. Unlike the case of heteroskedasticity, we usually assume the error terms are stationary, so that $Var(y_t) = \sigma_y^2$ is a constant across observations. The usual model for serial correlation assumes that the errors are first-order autoregressive (or $AR(1)$),

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t,$$

where the "primitive" error terms u_t are assumed to be mutually uncorrelated (or, stronger still, i.i.d.) with $E(u_t) = 0$ and $V(u_t) = \sigma_u^2$. By the usual $AR(1)$ calculations, assuming $|\rho| < 1$, we get $E(\varepsilon_t) = 0$, $C(\varepsilon_t, \varepsilon_s) = \sigma_u^2 \cdot \rho^{|t-s|}/(1-\rho^2)$, so the covariance matrix of the vector of error terms $\boldsymbol{\varepsilon}$ is of the form $\sigma_u^2 \cdot \boldsymbol{\Omega} \equiv \sigma_u^2 \cdot [\omega_{ts}]$, with

$$w_{ts} = \rho^{|t-s|}/(1-\rho^2).$$

Since one matrix square root $\mathbf{H}$ of $\boldsymbol{\Omega}^{-1}$ (defined so that $\mathbf{H}'\mathbf{H} = \boldsymbol{\Omega}^{-1}$) has the form

$$\mathbf{H} \equiv \begin{bmatrix} \sqrt{1-\rho^2} & 0 & \ldots & 0 \\ -\rho & 1 & 0 & \ldots \\ 0 & -\rho & 1 & 0 \\ 0 & 0 & -\rho & 1 \end{bmatrix},$$

the Feasible GLS estimator $\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$ for this model can be computed by a regression of transformed dependent variables y_t^* on corresponding regressors $\mathbf{x}_t^*$, where

$$y_1^* = \sqrt{1-\hat{\rho}^2} \cdot y_1, \qquad \mathbf{x}_1^* = \sqrt{1-\hat{\rho}^2} \cdot \mathbf{x}_1,$$

and for $t = 2, ..., T$,

$$y_t^* = y_t - \hat{\rho} y_{t-1},$$
$$\mathbf{x}_t^* = x_t - \hat{\rho}\mathbf{x}_{t-1}.$$

Ignoring the serial dependence in the error terms, we could use the classical least squares estimator $\hat{\beta}_{LS} = (X'X)^{-1}X'y$ to estimate β; this will be inefficient, but still consistent and asymptotically normal with

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}) \to^d \mathcal{N}(0, \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}),$$

where

$$\mathbf{D} = \text{plim}\frac{1}{T}\mathbf{X}'\mathbf{X},$$
$$\mathbf{V} = \text{plim}\frac{1}{T}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}.$$

However, the usual LS standard errors will be inconsistent, and might generally be expected to be too low, making estimated coefficients appear more "significant." To be more precise, suppose the regressors are stationary, and let $z_t = \mathbf{x}_t'\boldsymbol{\beta} - \bar{\mathbf{x}}'\boldsymbol{\beta}$ be the deviations of the true regression functions from their mean values, with $r = cov(\mathbf{x}_t'\boldsymbol{\beta}, \mathbf{x}_{t-1}'\boldsymbol{\beta})$ being their first-order autocorrelation coefficient. Then if $sgn(r) = sgn(\rho)$ (which is typical for economic applications, with both r and ρ positive), the R^2 from the classical least squares fit will be biased toward one – i.e.,

$$\text{plim } \frac{\mathbf{z}'\mathbf{z}}{(\mathbf{y} - \bar{\mathbf{y}})'(\mathbf{y} - \bar{\mathbf{y}})} >> \text{plim } \frac{\mathbf{z}'\boldsymbol{\Omega}^{-1}\mathbf{z}}{(\mathbf{y} - \bar{\mathbf{y}})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \bar{\mathbf{y}})},$$

where the left-hand side is the probability limit of the usual R^2 measure and the right-hand side corresponds to the correct GLS version. Indeed, unusually high values of R^2 (around 0.9) in time-series regressions without lagged dependent variables are generally associated with low values of the Durbin-Watson (DW) statistic discussed below, indicating that the serial correlation in the dependent variable is wrongly being attributed to the regression function, not the error terms.

To account for possible serial correlation, we can adopt either a "structural" or a "nonstructural" approach. In the "structural" approach, we would model the serial correlation process in terms of a few parameters (like assuming the errors are first-order autoregressive, as is traditional), and either test for lack of serial correlation (e.g., test $H_0 : \rho = 0$) and/or use estimates of the serial correlation parameters to do feasible GLS. A "nonstructural" alternative would stick with classical least squares to estimate $\hat{\boldsymbol{\beta}}_{LS}$, but would construct consistent estimates of the true asymptotic covariance matrix $\mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}$of least squares.

The Sample Autocorrelation Coefficient

Assuming that the error terms are $AR(1)$ (the simplest version of the "structural approach), a simple test for the null hypothesis that $\rho = 0$ could be based upon the least squares residuals $e = y - \mathbf{X}\hat{\boldsymbol{\beta}}_{LS}$. Since these residuals are consistent estimators for the true error terms ε, which have a simple linear relation to their lagged values, a natural estimator of ρ would regress e_t on e_{t-1} using least squares, yielding the estimator

$$\hat{\rho} = \frac{\sum_{t=z}^{T} e_t e_{t-1}}{\sum_{t=z}^{T} e_{t-1}^2}.$$

In general, assuming the linear model with $AR(1)$ errors is correctly specified, a suitable central limit theorem can be invoked to show that

$$\sqrt{T}(\hat{\rho} - \rho) \to^d \mathcal{N}(0, 1 - \rho^2),$$

so under the null hypothesis $H_0 : \rho = 0$,

$$\sqrt{T}\hat{\rho} \to^d \mathcal{N}(0, 1),$$

implying that H_o should be rejected (against a one-sided alternative $H_A : \rho > 0$) if $\sqrt{T}\hat{\rho}$ exceeds the upper α critical value $z(\alpha)$ of a standard normal distribution. It is worth noting that the validity of this result assumes that x_t and ε_s are independent (or at least uncorrelated) across *all* t and s, so the regressors cannot include lagged dependent variables (otherwise, the classical LS estimator $\hat{\boldsymbol{\beta}}_{LS}$ of $\boldsymbol{\beta}$ will be *inconsistent* under the alternative hypothesis).

The Durbin Watson Test

The traditional test statistic for (first-order) serially-correlated errors is the *Durbin-Watson statistic*, which is a close relative to the "natural" test statistic $\sqrt{T}\hat{\rho}$. This statistic, based on the least-squares residual vector $\mathbf{e}$, is defined as

$$DW = \sum_{t=2}^{T}(e_t - e_{t-1})^2 / \sum_{t=1}^{T} e_t^2.$$

By expanding the square and collecting terms, it is easy to verify that

$$DW \cong 2(1 - \hat{\rho})$$

for $\hat{\rho}$ defined above, where the approximation requires dropping a term or two in the summation of the squared residuals in the denominator. Thus, values of DW near zero indicate positively-autocorrelated residuals, and values near 4 suggest negative serial correlation. An asymptotically-equivalent version of the one-sided test for autocorrelation based on $\hat{\rho}$ would reject H_o if

$$DW < 2(1 - z(\alpha)/\sqrt{T}),$$

where $z(\alpha)$ is the upper standard normal critical value for a size α test. The traditional interest in the Durbin-Watson test in econometrics comes from a desire for an exact (not asymptotic) test under the assumption of normally-distributed error terms. Durbin and Watson showed that, while an exact critical value for the DW statistic under $H_0 : \rho = 0$ would depend in a complicated way on the matrix X of regressors, some bounds exist for the exact critical values which only depend on the number of regressors, K, and the significance level α. That is, for a one-sided test there are bounds $d_U \equiv d_U(K, \alpha)$ and $d_L = d_L(K, \alpha) < d_U$ such that

$$\Pr\{DW < d_L\} \leq \alpha \qquad \textit{and} \qquad \Pr\{DW > d_U\} \leq 1 - \alpha$$

under the null hypothesis $\rho = 0$, so an exact test would always reject if $DW < d_L$ and accept if $DW > d_U$. If the calculated DW statistic falls between the bounds d_L and d_U the test would be said to be "inconclusive," requiring either more complicated calculations to derive the exact critical value (depending upon the particular X matrix) or an asymptotic approximation. Since the asymptotic critical value $2(1 - z(\alpha))/\sqrt{T}$ always lies between the tabulated bounds, it is probably simpler from the start just to use the asymptotic test, which would yield identical results to the exact "bounds" test whenever the latter was not inconclusive.

The Breusch-Godfrey Test

While the Durbin-Watson test is formulated with the specific alternative hypothesis of $AR(1)$ error terms in mind, it should have some power in detecting other forms of serial correlation, provided $E[\varepsilon_t \varepsilon_{t-1}] \neq 0$ under the alternative hypothesis. Still, more powerful tests for higher-order serial correlation might involve higher-order autocorrelation estimators. Supposing the error terms are $AR(p)$ for $p > 1$, i.e.,

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \ldots + \rho_p \varepsilon_{t-p} + u_t, \qquad u_t \sim i.i.d., E(u_t) = 0, V(u_t) = \sigma^2,$$

a straightforward extension of the preceeding asymptotic testing procedure, proposed by Breusch and Godfrey, would regress the least squares residual e_t on p lagged values $e_{t-1}, \ldots, e_{t-p}$ using least squares (again, without a constant term) and test for joint singificance of these lagged values. Letting $\hat{u}_t$ be defined as the residuals from this regression,

$$\hat{u}_t = e_t - \hat{\rho}_1 e_{t-1} - \ldots - \hat{\rho}_p e_{t-p},$$

the Lagrange Multiplier test statistic for $H_0 : \rho_1 = ... = \rho_p = 0$ (assuming u_t is normally distributed) is

$$LM = TR^2,$$

where

$$R^2 \equiv 1 - \left(\sum_{t=p+1}^{T} \hat{u}_t^2 / \sum_{t=p+1}^{T} {}_t e_t^2 \right)$$

is the "non-constant-adjusted" R^2 for this residual regression. An asymptotic test of H_0 would reject if the LM statistic exceeded the upper α critical value of a chi-squared distribution with p degrees of freedom.

Feasible GLS

Returning to the case of $AR(1)$ errors, once it is determined that $\rho \neq 0$ (through prior reasoning or diagnostic testing), feasible GLS can be used to obtain a more efficient estimator of the original slope coefficients $\boldsymbol{\beta}$ of the linear model for y. There are several variations available for FGLS, depending on particular estimators of ρ and certain computational details. Some of these variants, typically named after the author(s) who proposed them, are as follows:

(1) *Prais-Winsten:* Use $\hat{\rho} = 1 - DW/2$ as an estimator of ρ, with

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$$

the standard FGLS estimator of β.

(2) *Cochrane-Orcutt*: Defining the residuals $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ in terms of the current value of $\hat{\boldsymbol{\beta}}$, iterate between $\hat{\rho} = \sum_{t=2}^{T} e_t e_{t-1} / \sum_{t=2}^{T} e_{t-1}^2$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$, starting with the least squares estimator $\hat{\boldsymbol{\beta}}_{LS}$.

(3) *Durbin's Method:* Use the GLS transformation

$$y_t - \rho y_{t-1} = (\mathbf{x}_t - \rho \mathbf{x}_{t-1})'\boldsymbol{\beta} + u_t, \qquad t = 2, \ldots, T,$$

to obtain the linear model

$$y_t = \rho y_{t-1} + \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{x}_{t-1}'\boldsymbol{\gamma} + u_t, \qquad t = 2, \ldots, T,$$

ignoring the restriction $\boldsymbol{\gamma} = -\rho\boldsymbol{\beta}$. Regressing y_t on y_{t-1}, $\mathbf{x}_t$, and $\mathbf{x}_{t-1}$ (deleting redundant constants when necessary) yields an estimator $\hat{\rho}$ of ρ as the coefficient on y_{t-1}, which is then used in the FGLS formula for an estimator of $\boldsymbol{\beta}$.

(4) *Hildreth-Liu:* Minimize the transformed sum of squared residuals,

$$SSR(\boldsymbol{\beta}, \rho) \equiv (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})'\boldsymbol{\Omega}^{-1}(\rho)(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}),$$

over $\boldsymbol{\beta}$ and a grid of values of ρ, which is essentially a nonlinear least squares variation on Durbin's method which imposes the restriction $\boldsymbol{\gamma} = -\rho\boldsymbol{\beta}$.

(5) *Maximum Likelihood:* Maximize the average log-likelihood function

$$\mathcal{L}(\boldsymbol{\beta}, \rho, \sigma^2) = \text{constant} - \frac{1}{2T\sigma^2} SSR(\boldsymbol{\beta}, \rho) - \frac{1}{2}\log\sigma^2 - \frac{1}{T}\log(1 - \rho^2)$$

over $\boldsymbol{\beta}$, ρ, and σ^2.

In all of these variations, the resulting estimators of $\boldsymbol{\beta}$ and ρ will have the same asymptotic properties, provided the form of the serial correlation is correctly specified. These approaches can be extended (some more conveniently than others) to deal with $AR(p)$ errors, though the simplest approach would use the estimated autoregression coefficients from the Breusch-Godfrey test to construct estimates of $\boldsymbol{\beta}$ using the generalized differenced model

$$\begin{aligned} y_t^* &= y_t - \rho_1 y_{t-1} - \ldots - \rho_p y_{t-p} \\ &\equiv (\mathbf{x}_t - \rho_1 \mathbf{x}_{t-1} - \ldots - \rho_p \mathbf{x}_{t-p})'\boldsymbol{\beta} + u_t \\ &= (\mathbf{x}_t^*)'\boldsymbol{\beta} + u_t \end{aligned}$$

using observations from $t = p + 1, ..., T$.

Testing with Lagged Dependent Variables

If the regressors x_t include lagged values of the dependent variable y_t, the results discussed above are not applicable. As noted above, the classical least squares estimators of β will be inconsistent if $\rho \neq 0$, and even when $\rho = 0$ the usual test statistic $\sqrt{T}\hat{\rho}$ has a distribution which is more tightly distributed around zero than a standard normal distribution – i.e.,

$$\sqrt{T}\hat{\rho} \rightarrow^d \mathcal{N}(0, \nu),$$

for $\nu << 1$ – so the actual significance level of the usual test will be much lower than the nominal size, making it more difficult to reject the null hypothesis. To obtain consistent estimators of β with serially-correlated errors and lagged dependent variables, other techniques (like instrumental variables estimation) will be needed. To test $H_0 : \rho = 0$ in a model with lagged dependent variables as regressors, Durbin proposed a correction to the usual statistic $\sqrt{T}\hat{\rho}$, now called "Durbin's h," which is of the form

$$h \equiv \frac{\sqrt{T}\hat{\rho}}{\sqrt{1 - T \cdot [SE(\hat{\beta}_1)]^2}},$$

where "$SE(\hat{\beta}_1)$" is the estimated standard error for the least-squares coefficient on the first lagged dependent variable y_{t-1} in the original regression of y_t on $\mathbf{x}_t$. Though this h statistic has a limiting standard normal distribution under the null hypothesis, there is no guarantee that the denominator will be well-defined in finite samples.

Another approach to testing for serially-correlated residuals extends the Breusch-Godfrey testing approach when lagged dependent variables appear in $\mathbf{x}_t$. In this case, to test whether ε_t is $AR(p)$, a second-step regression of the least-squares residuals e_t on $e_{t-1}, \ldots, e_{t-p}$, *and* $\mathbf{x}_t$ (including all lagged dependent variables) is used, with $T \cdot R^2$ from this second-stage regression being asymptotically $\chi^2(p)$ under the null hypothesis that all the serial correlation coefficients are zero.

Consistent Asymptotic Covariance Matrix Estimation

Assuming now that the regressor matrix $\mathbf{X}$ does not include lagged endogenous variables, the "non-structural" approach to dealing with serially-dependent errors would just modify the standard errors of the classical least squares estimator

$$\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

to account for possible serial correlation, and even for possible heteroskedasticity or distributional heterogeneity. Assuming only that the serial dependence in the errors declines sufficiently quickly as the time gap between the errors declines, it can be shown under suitable conditions that the asymptotic distribution of the least squares estimator has the form

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_{LS}-\boldsymbol{\beta}) \to^d \mathcal{N}(0, \mathbf{D}^{-1}\mathbf{V}\mathbf{D}^{-1}),$$

where

$$\begin{aligned} \mathbf{D} &= \text{plim}\frac{1}{T}\mathbf{X}'\mathbf{X}, \\ \mathbf{V} &= \text{plim}\frac{1}{T}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}, \end{aligned}$$

and $\boldsymbol{\Sigma} \equiv E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}]$ (as noted above). Obviously a consistent estimator $\hat{\mathbf{D}} = \frac{1}{T}\mathbf{X}'\mathbf{X}$ of D is available; what is needed is a consistent estimator for $\mathbf{V}$, which can be written in the general form

$$\mathbf{V} \equiv \boldsymbol{\Gamma}_0 + \sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_j+\boldsymbol{\Gamma}_j'),$$

where

$$\boldsymbol{\Gamma}_j \equiv \text{plim}\ \frac{1}{T}\sum_{t=j+1}^{T} \varepsilon_t\varepsilon_{t-j}\mathbf{x}_t\mathbf{x}_{t-j}.$$

Note that $\boldsymbol{\Gamma}_j = E[(\varepsilon_t\mathbf{x}_t)(\varepsilon_{t-j}\mathbf{x}_{t-j})']$ if $y_t, \mathbf{x}_t$ are jointly stationary; the leading term, $\boldsymbol{\Gamma}_0$, is the "middle matrix" for the asymptotic covariance of least squares under heteroskedasticity (but no serial correlation). For each j, a consistent estimator of $\boldsymbol{\Gamma}_j$ is

$$\mathbf{C}_j = \frac{1}{T}\sum_{t=j+1}^{T} e_t e_{t-j}\mathbf{x}_t\mathbf{x}_{t-j},$$

where $e_t = y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_{LS}$ is the least squares residual. However, for a fixed sample size T only $T-1$ of these estimators can be calculated, so an estimator of $\mathbf{V}$ must necessarily truncate the infinite sum, defining $\hat{V}$ to be a finite sum of the $\mathbf{C}_j$ terms. Using the fact that $\boldsymbol{\Gamma}_j$ tends to zero as j increases, we can obtain a consistent estimator by using only a relatively small number of $\mathbf{C}_j$ estimates in the estimator $\hat{\mathbf{V}}$. An estimator of $\mathbf{V}$ proposed by Newey and West uses a weighted sum of the $\mathbf{C}_j$ matrices,

$$\hat{\mathbf{V}} = \mathbf{C}_0 + \sum_{j=1}^{M}\left[1 - \frac{j}{M}\right] \cdot (\mathbf{C}_j+\mathbf{C}_j')$$

which they show will be consistent for $\mathbf{V}$ if M, the number of terms in the sum, increases slowly with the sample size, say

$$M = M(T) \to \infty, \qquad \frac{M}{\sqrt[3]{T}} \to 0 \qquad \text{as} \qquad T \to \infty$$

under some regularity conditions. Many other variations on this approach, which involve consistent estimation of the "spectral density matrix at frequency zero" of the variables $e_t \cdot \mathbf{x}_t$ (which is proportional to the matrix $\mathbf{V}$), have also been proposed.