Lectures for STAT280/380 (continued)

Nalini Ravishanker

Fall 2005
(ii) Regression with AutoCorrelated Errors

When data is observed over time, the assumption of independent errors is often suspect, as indicated by a plot of residuals versus order (time). The linear regression model with autocorrelated errors has the form

\[X_t = \beta_0 + \beta_1 Z_{t1} + \cdots + \beta_k Z_{tk} + \varepsilon_t, \ t = 1, \cdots, n,\]

(1)

where the subscript \( t \) is used to indicate time, and error terms from different time periods are correlated.

Assume that \( \{\varepsilon_t\} \) is a weakly stationary stochastic process. This assumption implies that the first two moments of the distribution of \( \varepsilon \) does not depend on \( t \), and we can write the variance-covariance matrix of \( \{\varepsilon_1, \cdots, \varepsilon_n\} \) as
$$V = \begin{pmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\
\rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\
\rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1
\end{pmatrix}$$

where $\rho_j = E(\varepsilon_t \varepsilon_{t-j})/\sigma^2 = E(\varepsilon_t \varepsilon_{t+j})/\sigma^2$, $j = 1, 2, \cdots$, is the autocorrelation between two random errors $j$ time periods apart.

This specification is still rather general, and the most commonly assumed specification for the stationary stochastic process is the first-order autoregressive ($AR(1)$) process:

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

(2)

where $\{u_t\}$ are iid errors. This is called serially correlated errors in regression.

For instance, consider the model

$$X_t = \beta Z_t + \varepsilon_t, \quad t = 1, \cdots, n,$$
where \( \varepsilon_t \) are normal random variables with \( E(\varepsilon_t) = 0 \), and \( \text{Cov}(\varepsilon_t, \varepsilon_s) = \sigma^2 \rho^{|t-s|}, \ t, s = 1, \ldots, n, \ |\rho| < 1 \). We can write this in the form of the general linear model with \( x = (X_1, \ldots, X_n)' \), \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)' \), \( z = (Z_1, \ldots, Z_n)' \), \( \beta \) is a scalar parameter, and

\[
V = \begin{pmatrix}
1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\
\rho & 1 & \rho & \cdots & \rho^{n-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1
\end{pmatrix}.
\]

This is a simple linear regression model with autoregressive order 1 (AR(1)) errors, i.e., a serially correlated simple regression model. We can verify that

\[
\{V^{-1}\}_{i,j} =
\begin{cases}
1/(1 - \rho^2), & \text{if } i = j = 1, n \\
(1 + \rho^2)/(1 - \rho^2), & \text{if } i = j = 2, \ldots, n - 1 \\
-\rho/(1 - \rho^2), & \text{if } |j - i| = 1 \\
0, & \text{otherwise}
\end{cases}
\]
Also,

\[(1 - \rho^2)(z'V^{-1}z) = \sum_{i=1}^{n} Z_i^2 + \rho^2 \sum_{i=2}^{n-1} Z_i^2 - 2\rho \sum_{i=2}^{n} Z_i Z_{i-1}, \text{ and} \]

\[(1 - \rho^2)z'V^{-1}x = \sum_{i=1}^{n} Z_i X_i + \rho^2 \sum_{i=2}^{n-1} Z_i X_i - 2\rho \sum_{i=2}^{n} (Z_i X_{i-1} + X_i Z_{i-1}). \]

Then,

\[\hat{\beta}_{GLS} = (z'V^{-1}z)^{-1}z'V^{-1}x\]

is the b.l.u.e. of \(\beta\) with variance given by \(\sigma^2(z'V^{-1}z)^{-1}\).
By substituting an estimate of $\sigma^2$, viz.,

$$\hat{\sigma}_{GLS}^2 = (x - z\hat{\beta}_{GLS})'V^{-1}(x - z\hat{\beta}_{GLS}),$$

we obtain the estimated variance of the GLS estimator of $\beta$. The variance of the OLS estimator of $\beta$ is

$$\text{Var}(\hat{\beta}) = \sigma^2(z'z)^{-1}z'Vz(z'z)^{-1}.$$ 

Therefore,

$$\text{Var}(\hat{\beta}) - \text{Var}(\hat{\beta}_{GLS}) = \sigma^2(z'z)^{-1}z'Vz(z'z)^{-1}$$

$$- \sigma^2(z'V^{-1}z)^{-1}z'V^{-1}VV^{-1}z(z'V^{-1}z)^{-1}$$

$$= \sigma^2sVs', \text{ say},$$

where $s = (z'z)^{-1}z' - (z'V^{-1}z)^{-1}z'V^{-1}$. This expression is p.s.d. since $V$ is p.d., verifying that $\text{Var}(\hat{\beta}) \geq \text{Var}(\hat{\beta}_{GLS})$; equality holds when $V = I_n$. 
Recall that for a regression model with serially correlated errors, we assume $E(\varepsilon_t) = 0$, and $Cov(\varepsilon_t, \varepsilon_s) = \sigma^2 \rho^{|t-s|}$, $s, t = 1, \ldots, n$, with $|\rho| < 1$. Usually $\rho$ is unknown and must be estimated from the data. Let $\hat{\beta}$ denote the OLS estimate of the vector of regression coefficients, and let $\hat{\varepsilon}$ denote the vector of OLS residuals.

**Sample Correlation.** The usual (and simplest) estimate the serial correlation $\rho$ by

$$\hat{\rho} = \frac{\sum_{t=2}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^{n} \hat{\varepsilon}_t^2}. \quad (3)$$

In addition to the residual versus time plot to detect serial correlation, we may use significance tests and enhanced graphical procedures to diagnose correlation in errors.
(iii) Durbin-Watson Test

The most popular test for serial correlation is the Durbin-Watson test (Durbin and Watson, 1950, 1951). The null hypothesis is $H_0 : \rho = 0$ while the alternative hypothesis is either $H_1 : \rho \neq 0$, or $H_1 : \rho > 0$, or $H_1 : \rho < 0$.

The Durbin-Watson test statistic is

$$DW = \left\{ \sum_{t=2}^{n} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2 \right\} \bigg/ \left\{ \sum_{t=1}^{n} \hat{\varepsilon}_t^2 \right\} \quad (4)$$

The exact distribution of $DW$ depends on $Z$. For a given level of significance $\alpha$, the bounds test has the following decision rules based on lower and upper critical values (at level $\alpha$) $d_{L,n,p}$ and $d_{U,n,p}$.
1. If $H_1$ is $\rho > 0$, reject $H_0$ at level of significance $\alpha$ if $DW < d_{L,n,p}$; do not reject $H_0$ if $DW > d_{U,n,p}$; and the test is inconclusive if $d_{L,n,p} \leq DW \leq d_{U,n,p}$.

2. If $H_1$ is $\rho < 0$, the decision rule has the form of the rule for $H_1 : \rho > 0$, replacing $DW$ by $(4 - DW)$.

3. If $H_1$ is $\rho \neq 0$, reject $H_0$ at level of significance $2\alpha$ if $DW < d_{L,n,p}$ or $4 - DW < d_{L,n,p}$; do not reject $H_0$ if $DW > d_{U,n,p}$ or $4 - DW > d_{U,n,p}$; and the test is inconclusive otherwise.

For a proof, see Ravishanker and Dey (2002), sectin 8.4.

Note that the summation in the numerator of the $DW$ statistic runs from $t = 2$ to $n$ since $\hat{e}_0$
is not available; this is referred to as an “end effect”. The statistic $DW$ lies in the range of 0 to 4, with a value of 2 indicating the absence of first-order serial correlation.

When successive values of $\hat{\varepsilon}_t$ are close together, $DW$ is small, indicating the presence of positive serial correlation. Ignoring end effects, we can show an approximate relation between $DW$ and $\hat{\rho}$.

**Theil-Nagar Modification for Estimating $\rho$.** They suggested that

$$\hat{\rho}_{TN} = \frac{n^2(1 - DW/2) + p^2}{n^2 - p^2} \quad (5)$$
It is important to realize that not rejecting the null hypothesis does not necessarily mean that the errors are uncorrelated; it simply means that there is no significant first-order autocorrelation. More complex linear stationary time series models may be employed in order to model autocorrelation in regression errors.

An example is the autoregressive moving-average process of order \((p, q)\) defined as

\[
\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \cdots + \phi_p \varepsilon_{t-p} + u_t + \\
\theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}
\]

where \(u_t\)'s are iid \(N(0, \sigma^2)\) variables, and the \((p+q)\) parameters \((\phi_1, \cdots, \phi_p)\), and \((\theta_1, \cdots, \theta_q)\) must satisfy certain restrictions. Plots of the sample autocorrelation function and partial autocorrelation function enable us to identify the model order in these cases.
(iv) Cochrane-Orcutt Procedure

If the errors in a regression model are serially correlated, an iterative estimation procedure called the Cochrane-Orcutt procedure is useful. Let \( \hat{\beta} \) denote the OLS estimate of \( \beta \), and let \( \hat{\varepsilon} \) be the OLS residual vector. In this initial step, the first-order serial correlation is estimated by (3).

Denote these quantities by \( \hat{\beta}^{(0)} \), \( \hat{\varepsilon}^{(0)} \) and \( \hat{\rho}^{(0)} \) respectively. Using \( \hat{\rho}^{(0)} \), we transform the original model and obtain OLS estimates from the transformed regression

\[
X_t^* = \beta_0(1 - \hat{\rho}^{(0)}) + \beta_1 Z_{t1}^* + \cdots + \beta_k Z_{tk}^* + u_t, \quad t = 1, \ldots, n,
\]

where \( X_t^* = X_t - \hat{\rho}^{(0)} X_{t-1} \), and

\[
Z_{tj}^* = Z_{tj} - \hat{\rho}^{(0)} Z_{t-1,j}, \quad j = 1, \ldots, k.
\]

Let \( \hat{\beta}^{(1)} \) and \( \hat{\varepsilon}^{(1)} \) respectively denote the vector of regression estimates, and residual vector from
this first iteration. We repeat this procedure several times until convergence. This procedure works well in practice, although there is always the danger that the procedure may tend to a local rather than a global maximum.
**Numerical Example**  Consider data from the first quarter of 1952 to the fourth quarter of 1956 on consumer expenditure in the U.S. in billions of dollars ($Y$), and money stock in billions of dollars ($X$). Let $t$ denote the time period (Chatterjee and Price, 1991). The regression estimates and the ANOVA table are shown below.

Least squares estimates for Numerical Example

<table>
<thead>
<tr>
<th>Parameter</th>
<th>d.f.</th>
<th>Estimate</th>
<th>s.e.</th>
<th>$t$-value</th>
<th>Pr &gt; $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>$-154.95$</td>
<td>19.88</td>
<td>$-7.79$</td>
<td>$&lt;.0001$</td>
</tr>
<tr>
<td>$X$</td>
<td>1</td>
<td>$2.30$</td>
<td>0.12</td>
<td>$20.06$</td>
<td>$&lt;.0001$</td>
</tr>
</tbody>
</table>
ANOVA table

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>F-value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>6395.17</td>
<td>6395.17</td>
<td>402.34</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>286.11</td>
<td>15.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>19</td>
<td>6681.28</td>
<td>402.34</td>
<td>&lt; .0001</td>
<td></td>
</tr>
</tbody>
</table>

Also, \( \hat{\sigma} = 3.987 \), \( R^2 = 957 \), \( R_{\text{adj.}}^2 = 955 \), and the estimate of the first-order serial correlation \( \hat{\rho} = 0.752 \), with s.e. 0.160. The Durbin-Watson statistic is \( DW = 0.326 \); comparing it to the lower and upper critical values \( d_L = 1.201 \) and \( d_U = 1.411 \), we reject \( H_0 : \rho = 0 \) at the 5% level of significance. One way to fit a regression model with \( AR(1) \) errors to the data is via a two-step least squares procedure. The results from this fit are shown below.

Simultaneous estimates of \( \beta \) and \( \rho \)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>d.f.</th>
<th>Estimate</th>
<th>s.e.</th>
<th>t-value</th>
<th>Pr &gt; t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>−158.28</td>
<td>32.13</td>
<td>−4.93</td>
<td>.0001</td>
</tr>
<tr>
<td>( X )</td>
<td>1</td>
<td>2.33</td>
<td>0.19</td>
<td>12.54</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

The estimate of the \( AR(1) \) parameter is \( \hat{\rho} = 0.752 \), with standard error 0.160.