Regression Methods

Overview

1. Idea: effects of dependence
2. Examples of estimation (in R)
3. Review of regression
4. Comparisons and relative efficiencies

Idea

Decomposition Well-known way to approach time series analysis is to decompose the observed series into several components,

$$X_t = \text{Trend}_t + \text{Seasonal}_t + \text{Irregular}_t$$

and then model the additive components separately. Here "trend" usually means linear or quadratic patterns.

Stationarity The methods we have developed require (or at least assume to be plausible) the assumption of stationarity to permit averaging over time. That does not make sense for the trend component. A simpler view of the problem is to think of the data as

$$X_t = \text{deterministic patterns}_t + \text{stationary variation}_t$$

and then model the deterministic component using tools like regression, leaving the rest to be handled as a stationary process, eventually reducing the data to

$$X_t = \text{predictable variation}_t + \text{white noise}_t$$

Examples

Data series from the text are:

- Global temperature (since 1900, comparing annual and monthly) A univariate model, with no exogenous variables.
- Mortality data, with several exogenous variables.
- Fish harvest (SOI), with many plausible variables.

Question In these examples, the regression residuals show correlations at several lags. Are the usual summaries of the fitted models (standard errors and t-statistics) reliable?

Quick Review of Regression

Data Response is a column vector $y = y_{1:n} = (y_1, \ldots, y_n)'$ that is a time series in our examples. The explanatory variables (including a leading column of 1s for the intercept) are collected as columns in the $n \times q$ matrix X. Denote a row of X as the column vector x_t.

Model Linear equation with *stationary errors* in scalar form is

$$y_t = x_t'\beta + e_t, \quad E(e_t) = 0, \, \mathrm{Cov}(e_t, e_s) = \gamma(t,s) = \gamma(|t-s|)$$

and in matrix form with the errors in the vector $e = e_{1:n}$

$$y = X\,\beta + e, \quad E(e) = 0, \, \mathrm{Var}(e) = \Gamma = \sigma^2\Phi \tag{1}$$

where Φ is normalized to have 1's along its diagonal (a correlation matrix) by dividing each element in Γ by $\gamma(0)$. Hence $\mathrm{Var}(e_t) = \gamma(0) = \sigma^2$. (Notation with $\gamma(0) = \sigma^2$ gives expressions that look more familiar.)

OLS Estimator Assuming the explanatory variables are known and observed, the OLS estimator for β is

$$\hat\beta = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'e \tag{2}$$

with the residuals $\hat e_t = y_t - x_t'\hat\beta$. The usual unbiased estimator of σ^2 (given uncorrelated errors) is

$$\hat\sigma^2 = \frac{\sum \hat e_t^2}{n-q}$$

Tests The text reviews the basic tests (such as the F test for added variables) in Section 2.2.

Properties of the OLS estimator The second form of (2) makes it easy to see that although it ignores dependence among the observations, the OLS estimator is unbiased. The variance of the estimator is

$$\begin{aligned}\mathrm{Var}(\hat\beta) &= (X'X)^{-1}X' \ \mathrm{Var}(e)\ X(X'X)^{-1} \\ &= (X'X)^{-1}X' \ \Gamma \ X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'\Phi X(X'X)^{-1}\end{aligned}$$

This contrasts with the usual expression in the case of uncorrelated errors,

$$\mathrm{Var}(\hat\beta) = \sigma^2(X'X)^{-1}$$

Sandwich estimator The form of the variance of the OLS estimator suggests a simple, robust estimate of its variance:

$$\mathrm{var}(\hat\beta) = (X'X)^{-1}X'\hat\Gamma X(X'X)^{-1} \tag{3}$$

In the heteroscedastic case, this leads to the *White estimator* which uses

$$\text{White estimator: } \hat\Gamma = \mathrm{diag}(e_t^2)\ .$$

For dependence, one can use blocked estimates of the covariances.

GLS Estimator The GLS estimator (which is the MLE under a normal distribution) is the solution to the following minimization:

$$\min_\alpha (y - X\alpha)'\Gamma^{-1}(y - X\alpha) \tag{4}$$

To solve (4), assume that we know Γ and factor this matrix as $\Gamma = \Gamma^{1/2}\Gamma^{1/2\prime}$ and express the sum of squares to be minimized as

$$(\Gamma^{-1/2}y - (\Gamma^{-1/2}X)\alpha)'(\Gamma^{-1/2}y - (\Gamma^{-1/2}X)\alpha) = (\tilde y - \tilde X\alpha)'(\tilde y - \tilde X\alpha)$$

where $\tilde y = \Gamma^{-1/2}y$ and $\tilde X = \Gamma^{-1/2}X$. This is now formulated as a regular least squares problem, for which the solution is

$$\tilde\beta = (\tilde X'\tilde X)^{-1}\tilde X'\tilde y = (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}y = \beta + (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}e$$

The variance of the GLS estimator is

$$\mathrm{Var}(\tilde\beta) = \sigma^2(X'\Phi X)^{-1}$$

Comparison of Estimators: Scalar Model

Scalar model ($q = 1$) simplifies the arithmetic and suggests the matrix expressions.

$$y_t = \beta x_t + e_t \quad \Rightarrow \quad \hat\beta = \frac{x'y}{x'x} = \beta + \frac{x'e}{x'x}$$

Assume that the errors are autoregressive with coefficient α:

$$e_t = \alpha e_{t-1} + w_t, \quad |\alpha| < 1.$$

Impact of dependence The variance of the OLS estimator is (use the fact that $\Phi_{ij} = \alpha^{|i-j|}$)

$$\begin{aligned}
\mathrm{Var}(\hat\beta) &= \frac{\mathrm{Var}(x'e)}{(x'x)^2} = \frac{\sigma^2}{(x'x)^2}\, x'\Phi x \\
&= \frac{\sigma^2}{(x'x)^2}\left(\sum x_t^2 + 2\alpha \sum x_t x_{t-1} + 2\alpha^2 \sum x_t x_{t-2} \cdots\right) \\
&= \frac{\sigma^2}{(x'x)}\left(1 + 2\alpha \frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\alpha^2 \frac{\sum x_t x_{t-2}}{\sum x_t^2} \cdots\right)
\end{aligned}$$

We can obtain nicer expressions by choosing values of x_t so that these sums (which look like estimated correlations) have known values (approximately). The easiest way to do this is to imagine x_t as an autoregression with some coefficient,

$$x_t = \rho\, x_{t-1} + u_t, \quad |\rho| < 1.$$

With this choice, the sums $\frac{\sum x_{t+h} x_t}{\sum x_t^2} \approx \rho^{|h|}$ and the approximation becomes

$$\begin{aligned}
\mathrm{Var}(\hat\beta) &= \frac{\sigma^2}{(x'x)}\left(1 + 2\alpha\rho + 2\alpha^2\rho^2 \cdots\right) \\
&\approx \frac{\sigma^2}{(x'x)} \frac{1+\alpha\rho}{1-\alpha\rho}
\end{aligned}$$

Notice that we are *not* thinking of x_t as a random variable; we are using this method to generate a sequence of numbers with a specific set of autocorrelations.

We can also approximate $x'x$ as $n\,\mathrm{Var}(X)$ and write

$$\mathrm{Var}(\hat\beta) \approx \frac{\sigma^2}{n\,\mathrm{Var}(X)} \frac{1+\alpha\rho}{1-\alpha\rho} \tag{5}$$

Optimistic Compared the expression for the variance with uncorrelated errors, the actual variance differs from the usual expression by the factor

$$\frac{1+\alpha\rho}{1-\alpha\rho}$$

If you set $\rho = 1$, you get the factor from the previous lecture for the variance of $\overline{X}$ in the presence of autocorrelated data. In this case, the presence of positive correlation in the exogenous series compounds the loss of efficiency.

For example, if $\alpha = \rho = 0.8$, then $\frac{1+\alpha\rho}{1-\alpha\rho} = \frac{1+0.64}{0.36} \approx 4.56$. Ignoring the dependence leads to a false sense of accuracy (*i.e.*, confidence intervals based on the OLS expression are too narrow for the stated coverage).

Bias of $\hat\sigma^2$ The estimator is also biased due to the autocorreation. In the scalar model under these same conditions, we can approximate the bias easily: (Note that $\hat\beta - \beta = (\sum x_t e_t)/\sum x_t^2$)

$$\begin{aligned}
\mathbb{E} \sum \hat e_t^2 &= \mathbb{E} \sum \left(e_t - x_t(\hat\beta-\beta)\right)^2 \\
&= \mathbb{E} \sum \left(e_t^2 + x_t^2(\hat\beta-\beta)^2 - 2(\hat\beta-\beta)x_t e_t\right) \\
&\approx \sigma^2\left(n + \frac{1+\alpha\rho}{1-\alpha\rho} - 2\frac{1+\alpha\rho}{1-\alpha\rho}\right) \\
&= \sigma^2(n - \frac{1+\alpha\rho}{1-\alpha\rho})
\end{aligned}$$

which is *also* shrunken toward zero. Hence the usual OLS expression $s^2/(x'x)$ is going to be much too small.

Relative efficiency The simplified univariate setup makes it easy to compare the variances of the OLS and GLS estimators. To obtain the GLS estimator in the scalar case, we can write the effects of multiplying by $\Gamma^{-1/2}$ directly. Namely, subtract the expression for αy_{t-1} from the expression for y_t:

$$y_t - \alpha y_{t-1} = \beta(x_t - \alpha x_{t-1}) + \underbrace{(e_t - \alpha e_{t-1})}_{w_t}$$

Hence, we can estimate β efficiently via OLS using these so-called "generalized differences" of the observed data. The resulting GLS estimator is

$$\tilde\beta = \frac{\sum(x_t - \alpha x_{t-1})(y_t - \alpha y_{t-1})}{\sum(x_t - \alpha x_{t-1})^2}$$

The variance is then (treating x_t as uncorrelated with w_t, a must for least squares, and then viewing as conditional on x_t)

$$\begin{aligned}\mathrm{Var}(\tilde\beta) &= \mathrm{Var}\left(\frac{\sum(x_t - \alpha x_{t-1})w_t}{\sum(x_t - \alpha x_{t-1})^2}\right)\\ &= \frac{\sigma_w^2}{\sum(x_t - \alpha x_{t-1})^2}\end{aligned}$$

Assuming $\{X_t\}$ is autoregressive with coefficient ρ as before implies that (expand the square and approximate the 3 sums)

$$(1/n)\sum(x_t - \alpha x_{t-1})^2 \approx \sigma_x^2(1+\alpha^2 - 2\alpha\rho)$$

We also know that $\mathrm{Var}(e_t) = \sigma_w^2/(1-\alpha^2)$, so $\sigma_w^2 = (1-\alpha^2)\sigma^2$ (note: $\sigma^2 = \mathrm{Var}(e_t)$). Hence, when all are pulled together we get

$$\mathrm{Var}(\tilde\beta) \approx \frac{\sigma^2}{n\sigma_x^2}\frac{1-\alpha^2}{1+\alpha^2-2\alpha\rho}$$

and the ratio of variances (see (5)) is

$$\frac{\mathrm{Var}(\tilde\beta)}{\mathrm{Var}(\hat\beta)} \approx \frac{1-\alpha\rho}{1+\alpha\rho}\;\frac{1-\alpha^2}{1+\alpha^2-2\alpha\rho}\,.$$

For the example with $\alpha = \rho = 0.8$, the efficiency is 0.21.

Conclusion OLS is not only not efficient, it makes claims as though it were. Under this model of positively correlated data,

- OLS expressions underestimate the variance of $\hat\beta$
- OLS estimates are much less efficient than GLS estimates.

Comparison of Estimators: General Case

Variances The variances of the estimators are

$$\text{OLS:}\quad \mathrm{Var}(\hat\beta) = \sigma^2(X'X)^{-1}X'\Phi X(X'X)^{-1}$$

and

$$\text{GLS:}\quad \mathrm{Var}(\tilde\beta) = \sigma^2(X'\Phi^{-1}X)^{-1}$$

Equality of estimators The columns of X must span the same linear subspace as the columns of ΦX (as if $\Phi X = X$). In general, that does not happen. (We will later see some situations in which it does, at least for n large.) If the columns of X are eigenvectors of Φ, equality also obtains. Asymptotically, that will occur for some very special Xs.

Comparison criterion There's no unique way to order matrices. To find a criterion, consider this one: because the GLS estimator has smaller variance than OLS, we know that for any linear combination of estimates

$$\mathrm{Var}(a'\tilde{\beta}) \leq \mathrm{Var}(a'\hat{\beta})$$

It follows then that we can use "generalized variances" (*i.e.*, determinants) to order the estimators:

$$|Var(\tilde{\beta})| \leq |\,\mathrm{Var}(\hat{\beta})|$$

and the ratio of these generalized variances (efficiency?) is

$$\frac{|\,\mathrm{Var}(\tilde{\beta})|}{|\,\mathrm{Var}(\hat{\beta})|} = \frac{|X'X|^2}{|X'\Phi X|\;|X'\Phi^{-1}X|}$$

Bounds By orthogonalizing X so that $X'X = I$, it can be shown that there is a lower bound to the efficiency given by

$$\frac{\mathrm{Var}(a'\tilde{\beta})}{\mathrm{Var}(a'\hat{\beta})} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n > 0$ are the eigenvalues of Φ (which is assumed to be full rank). These bounds come from the *Kantorovich inequality*: For $0 < \lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$ and nonnegative weights p_i for which $\sum p_i = 1$, then

$$\left(\sum p_i\lambda_i\right)\,\left(\sum p_i\frac{1}{\lambda_i}\right) \leq \frac{A^2}{G^2}$$

where $A = (\lambda_1+\lambda_n)/2$ is the arithmetic mean and $G = \sqrt{\lambda_1\lambda_n}$ is the geometric mean.

Lingering Questions

Computing in practice Suppose that we estimate α from the data. Is the resulting approximate GLS estimator as good as suggested by these approximations?

Model specification What if we don't know that the process is autoregressive, much less its coefficient? Are the gains going to be so large as possible from these indications?