The Multiple Regression Model in Time Series

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + u_t$$

where

$$E(u_t|\mathcal{I}_t) = 0,$$
$$E(u_t^2|\mathcal{I}_t) = \sigma^2.$$

- $\mathcal{I}_t$ denotes the set of "valid conditioning variables", and $(\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots) \in \mathcal{I}_t$.
- Also, $u_{t-j} = y_{t-j} - \boldsymbol{\beta}'\mathbf{x}_{t-j} \in \mathcal{I}_t$ for $j > 0$.
- u_t and $u_t^2 - \sigma^2$ are m.d.s with respect to the information set $\mathcal{I}_t^* = (\mathcal{I}_t, y_t)$.

We need conditions on $\mathbf{x}_t$ to be satisfied so that

$$\text{plim}\, T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' = \mathbf{M}_{XX} \text{ (finite, nonsingular)}$$

Then, same type of arguments as before show

$$\text{plim}\, T^{-1}\sum_{t=1}^{T}\mathbf{x}_t u_t = \mathbf{0}$$

and

$$\begin{aligned}
\text{plim}\,\hat{\boldsymbol{\beta}} &= \text{plim}\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\sum_{t=1}^{T}\mathbf{x}_t y_t \\
&= \boldsymbol{\beta} + \left(\text{plim}\, T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\text{plim}\, T^{-1}\sum_{t=1}^{T}\mathbf{x}_t u_t \\
&= \boldsymbol{\beta} + \mathbf{M}_{XX}^{-1}.\mathbf{0} \\
&= \boldsymbol{\beta}.
\end{aligned}$$

Also, we show by similar arguments that

$$T^{-1/2} \sum_{t=1}^{T} \mathbf{x}_t u_t \underset{asy}{\sim} \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{M}_{XX}).$$

Hence, by Cramér Theorem,

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(T^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} T^{-1/2} \sum_{t=1}^{T} \mathbf{x}_t u_t$$

$$\underset{asy}{\sim} \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{M}_{XX}^{-1})$$

Failure of the Assumptions

1. Why isn't it enough to have $E(u_t|x_t) = 0$?

   Answer: OK in Model B, but in time series data, this assumption does not rule out *autocorrelation*.

$$E(u_t u_{t-j}) \neq 0 \text{ for } j > 0$$

   This problem causes inefficiency, and possible bias.

2. Why do we need $E(u_t^2|\mathcal{I}_t) = \sigma^2$? (Conditional homoscedasticity)

   Answer: We need $\mathrm{Var}(T^{-1/2}\sum_{t=1}^{T}\mathbf{x}_t u_t) \to \sigma^2\mathbf{M}_{XX}$ to get the stated formula, implying

$$E(u_t^2\mathbf{x}_t\mathbf{x}_t') = E(u_t^2)E(\mathbf{x}_t\mathbf{x}_t') = \sigma^2\mathbf{M}_{XX}$$

   If this does not hold, then the argument becomes

$$T^{-1/2}\sum_{t=1}^{T}\mathbf{x}_t u_t \underset{asy}{\sim} \mathrm{N}(\mathbf{0},\mathbf{A})$$

   where $\mathbf{A} = E(u_t^2\mathbf{x}_t\mathbf{x}_t')$. In this case,

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \underset{asy}{\sim} \mathrm{N}(\mathbf{0},\mathbf{M}_{XX}^{-1}\mathbf{A}\mathbf{M}_{XX}^{-1}).$$

   The standard OLS variance formula is wrong. This problem can be remedied by using a "robust" variance formula.

   Eicker-White variance estimator is $\mathbf{M}_{XX}^{-1}\hat{\mathbf{A}}\mathbf{M}_{XX}^{-1}$ where

$$\hat{\mathbf{A}} = T^{-1}\sum_{t=1}^{T}u_t^2\mathbf{x}_t\mathbf{x}_t'.$$

The Partitioned Linear Model

In $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, let $\boldsymbol{\beta} = \begin{array}{c} \left[\begin{array}{c} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{array}\right] \end{array} \begin{array}{c} k_1 \\ k_2 \end{array}$,

$\mathbf{X} = \left[\mathbf{X}_1 \;\; \mathbf{X}_2\right]$, ($\mathbf{X}_1$ $T \times k_1$, $\mathbf{X}_2$ is $T \times k_2$, $k_1 + k_2 = k$)

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}.$$

Normal equations, $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$:

$$\left[\begin{array}{cc} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{array}\right]\left[\begin{array}{c} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{array}\right] = \left[\begin{array}{c} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{array}\right]$$

or

$$\mathbf{X}_1'\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_1'\mathbf{y} \tag{1}$$

$$\mathbf{X}_2'\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{X}_2'\mathbf{y}. \tag{2}$$

Solve for $\hat{\boldsymbol{\beta}}_1$: from (1),

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}(\mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\hat{\boldsymbol{\beta}}_1).$$

From (2)

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1 &= [\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1]^{-1} \\ &\qquad [\mathbf{X}_1'\mathbf{y} - \mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}] \\ &= (\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \end{aligned}$$

where $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$.

(Or, use partitioned inverse formula)

Frisch–Waugh theorem.

Because of the idempotency, symmetry of $\mathbf{M}_2$, write

$$\begin{aligned}\hat{\boldsymbol{\beta}}_1 &= (\mathbf{X}_1'\mathbf{M}_2'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2'\mathbf{M}_2\mathbf{y} \\ &= (\mathbf{X}_1^{*\prime}\mathbf{X}_1^*)^{-1}\mathbf{X}_1^{*\prime}\mathbf{y}^*\end{aligned}$$

where $\mathbf{X}_1^* = \mathbf{M}_2\mathbf{X}_1,$ and $\mathbf{y}^* = \mathbf{M}_2\mathbf{y}$ are residuals from regression on $\mathbf{X}_2$.

F-W says: two-stage regression equivalent to multiple regression.

Residuals

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 \\ &= \mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}(\mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\hat{\boldsymbol{\beta}}_1) \\ &= \mathbf{M}_2\mathbf{y} - \mathbf{M}_2\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 \\ &= [\mathbf{M}_2 - \mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2]\mathbf{y}\end{aligned}$$

Since also

$$\hat{\mathbf{u}} = \mathbf{M}\mathbf{y}$$

for arbitrary $\mathbf{y}$, it follows that

$$\mathbf{M} = \mathbf{M}_2 - \mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{M}_2.$$

Variable Addition Tests

Consider test of $H_0 : \boldsymbol{\delta} = \mathbf{0}$ in the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}$$

where $\mathbf{X}$ is $T \times l$ and $\mathbf{Z}$ is $T \times r$.

Let $\dot{\mathbf{u}} = \mathbf{y} - \mathbf{X}\dot{\boldsymbol{\beta}} = \mathbf{M}_X\mathbf{y}$ and

$$\hat{\mathbf{u}} = \mathbf{M}\mathbf{y} = (\mathbf{M}_X - \mathbf{M}_X\mathbf{Z}(\mathbf{Z}'\mathbf{M}_X\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{M}_X)\mathbf{y}$$

The standard F test of *exclusion restrictions* is

$$\mathrm{W} = \frac{\hat{\boldsymbol{\delta}}'\mathbf{Z}'\mathbf{M}_X\mathbf{Z}\hat{\boldsymbol{\delta}}}{\hat{\sigma}^2} = (T - l - r)\frac{\dot{\mathbf{u}}'\dot{\mathbf{u}} - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \underset{asy}{\sim} \chi^2(r) \text{ on } H_0.$$

Alternative Form

Consider

$$\dot{\mathbf{u}} = \mathbf{X}(\boldsymbol{\beta} - \dot{\boldsymbol{\beta}}) + \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}$$

Note, $R^2 = 1 - \dfrac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\dot{\mathbf{u}}'\dot{\mathbf{u}}}$.

Since $\mathbf{M}_X\dot{\mathbf{u}} = \mathbf{M}_X\mathbf{y}$, test of $H_0 : \boldsymbol{\delta} = \mathbf{0}$ has the same form,

$$\mathrm{W} = (T - l - r)\frac{\dot{\mathbf{u}}'\dot{\mathbf{u}} - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} = (T - l - r)\frac{R^2}{1 - R^2}.$$

NB this test has r degrees of freedom not $l + r - 1$!

Diagnostic tests are often given in the form

$$\mathrm{W}^* = TR^2 = T\frac{\dot{\mathbf{u}}'\dot{\mathbf{u}} - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\dot{\mathbf{u}}'\dot{\mathbf{u}}}.$$

Asymptotically equivalent under H_0 since then

$$\dot{\mathbf{u}}'\dot{\mathbf{u}}/T - \hat{\mathbf{u}}'\hat{\mathbf{u}}/T \xrightarrow{\text{pr}} 0$$

Special cases:

1. $r = 1$: δ a scalar, $\mathbf{z}$ is $T \times 1$:

$$\mathrm{W} = \frac{\hat{\delta}^2}{s^2(\mathbf{z}'\mathbf{M}_X\mathbf{z})^{-1}} \underset{asy}{\sim} \chi^2(1)$$

where $s^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(T-k)$ and note, $\mathrm{W}^{1/2} = |t_\delta|$ where $t_\delta \underset{asy}{\sim} \mathrm{N}(0,1)$ on H_0.

2. $l = 1$: β = intercept, $\mathbf{x} = \boldsymbol{\iota} = (1,1,\dots,1)'$, $\dot{\mathbf{u}} = \mathbf{y} - \boldsymbol{\iota}\bar{y}$.

$$\mathrm{W} = (T-r-1)\frac{\mathbf{y}'\mathbf{y} - T\bar{y}^2 - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}}$$

$$= (T-r-1)\frac{R^2}{1-R^2} \underset{asy}{\sim} \chi^2(r).$$

The "F-test of the regression".

Spurious Regression

Suppose $y_t = y_{t-1} + u_t$ where $u_t \sim \mathrm{NID}(0,1)$, and $x_{jt} = x_{jt-1} + v_{jt},\ v_{jt} \sim \mathrm{NID}(0,1)$ for $j = 1,\dots,k$.

(all errors independent).

Consider regressing $\mathbf{y} = (y_1,\dots y_T)'$ on $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3\ \boldsymbol{\iota}]$.

It can be shown that

$$P(F\text{-test of the regression rejects } H_0) \to 1 \text{ as } T \to \infty.$$

In general, this test is *not valid* for time series applications!

Diagnostic Tests

- H_0 is the model under test, subject to assumption $E(u_t|\mathcal{I}_t) = 0$.
- H_1 usually a "dummy alternative", with some $\mathbf{z}_t \in \mathcal{I}_t$ to capture failures of the assumptions.
- Usual procedure, a variable addition test for significance of $\mathbf{z}_t$.

Examples

Conventionally, let $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

1. Autocorrelation: $\mathbf{z}_t = (\hat{u}_{t-1}, \ldots, \hat{u}_{t-r})$.
2. Mis-specified dynamics:
$$\mathbf{z}_t = (\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-r}, y_{t-1}, \ldots, y_{t-r})^*$$
where * denotes omission of redundant columns.
3. Functional Form (RESET test): $\mathbf{z}_t = (\hat{y}_t^2, \ldots, \hat{y}_t^{1+r})$ where $\hat{y} = \mathbf{x}_t\hat{\boldsymbol{\beta}}$.
4. Other omitted regressors ...
5. Heteroscedasticity: regress $\hat{u}_t^2$ on $\hat{y}_t^2$
6. Conditional heteroscedasticity (ARCH): regress $\hat{u}_t^2$ on $(\hat{u}_{t-1}^2, \ldots, \hat{u}_{t-r}^2)$.

- Including $\mathbf{X}$ in the test regressions 1–4 is important for the test to be unbiased (have rejection probability > 0.05 when H_0 false.)
- However, tests 5 and 6 do *not* have $\mathbf{X}$ as regressors.
- Durbin-Watson test is a common alternative to 1, but NOT recommended.

Modelling Methodology

Problem with econometric work:

- data are non-experimental.
- We don't know the "true" model

Consider the set of all possible models of y_t, containing all possible combinations of variables from $\mathcal{I}_t$, and a data set $y_1, \ldots, y_T$

This set contains two important subsets.

1. *Theory consistent* models (consistent with our prejudices derived from economic theory)
2. *Data-coherent* models (satisfying statistical assumptions for valid regression)

The task of econometric modellers is to find a model in the intersection of these two sets.

- Some models infeasible, since too large for data set (need $k << T$).
- *Principle of Parsimony.* Assume a simple model is adequate - if faced with a choice, prefer the simplest explanation (Occam's razor).

Methods for selecting a model

Use diagnostic tests to verify if a particular model is data coherent.

Inevitably, entails *sequential testing*.
i.e., a sequence of tests where each model tested depends on the outcome of the previous test.
This is not a setting that statistical tests were designed for!

Problems

1. Consider 'nested' hypotheses $H_0 \subset H_1 \subset H_2$.

   Sequence:

   (i) Test H_1 vs H_2 at level α

   (ii) If accepted, test H_0 vs H_1 at level α

   If H_0 true, what is probability of rejecting it?

   Answer (the best one available):

   $P(\text{reject } H_0) = P(\text{reject } H_1 \cup (\text{accept } H_1 \cap \text{reject } H_0 \text{ vs. } H_1))$

   $\leq P(\text{reject } H_1) + P(\text{reject } H_0 \text{ vs. } H_1) = 2\alpha$

   True significance level unknown in multiple tests.

   To bound it, should use $\alpha/2$-level test at each stage.

2. $H_M = H_0 \cup H_A$ = "Maintained Hypothesis" -

   $H_0 = H_M - H_A$ = "H_M but not H_A".

   If H_M is false, then H_0 is false, even if "not H_A" true!

   Test does not have the assumed significance level.

   Example:

   Suppose true model = $(x_t,\ z_{Bt})$.

   t test of $H_0\ (x_t)$ vs. $H_1(x_t,\ z_{At})$ can reject more than 5% of the time.

   e.g., z_{At} may act as proxy for z_{Bt}.

   If next test is of $H_1(x_t,\ z_{At})$ vs. $H_1(x_t,\ z_{At}, z_{Bt})$,

   this test can lack power because of irrelevant z_{At}

   - end up with wrong model.

Three approaches to model selection:

1. "Particular to General" (PTG)
   Start with simple model, modify by adding variables as necessary to make data-coherent.
   - problem of false H_M

2. "General to Particular" (GTP)
   Start with most general model possible, simplify progressively until reaching a model that gets rejected.
   - Usually recommended - but difficult to formulate large H_M.

3. Use an information criterion.

- Choose model on basis of goodness of fit, with penalty to favour parsimony.

Akaike Criterion: $-\left(\log(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}) + \frac{2k}{T}\right)$

Schwarz Criterion: $-\left(\log(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}) + \frac{k\log T}{T}\right)$

Hannan-Quinn Criterion:

$$-\left(\log(\frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}) + \frac{2k\log\log T}{T}\right)$$

$$\bar{R}^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}/T - k}{(\mathbf{y}'\mathbf{y} - T\bar{y}^2)/(T-1)}.$$

In this way, compare *all* models.

(Method 3 recommended! I favour Schwarz crit, encourages parsimony.)