8: The k-Variable Linear Model 3
ECON 837

Brian Krauth (adapted from notes by Simon Woodcock), Spring 2010

The model so far

At this point I'd like to review what we have so far, and give everyone a shorthand for stating our assumptions clearly. The *classical linear regression model* entails the following assumptions:

- D (data): We have a data set consisting of an n-vector $\mathbf{y}$ and an $n \times k$ matrix $\mathbf{X}$ such that $(X'X)$ is nonsingular.
- B (definition of β): There exists a k-vector β and an n-vector ε such that $\mathbf{y} = \mathbf{X}\beta + \varepsilon$. and $E(\mathbf{X}'\varepsilon) = 0$.
- L (linearity of the CEF): $E(\mathbf{y}|\mathbf{X}) = \mathbf{X}\beta$, or equivalently $E(\varepsilon|\mathbf{X}) = \mathbf{0}$.
- S (spherical errors): $E(\varepsilon'\varepsilon) = \sigma^2 I_n$.
- N (normal errors): $\varepsilon|\mathbf{X} \sim N(\mathbf{0}, \sigma^2 I_n)$.

We define the following statistics:

- The OLS coefficient vector is $\hat{\beta} \equiv (X'X)^{-1}X'y$.
- The vector of OLS fitted values is $\hat{y} \equiv X\hat{\beta}$.
- The vector of OLS residuals is $e \equiv y - \hat{y}$.

Depending on what we are trying to do, we don't need all of these assumptions.

- The various *algebraic* properties of the OLS statistics follow from (D).
- In order to interpret the OLS statistics as estimators we need to add (B).
- Unbiasedness requires that we add (L).
- The Gauss-Markov theorem and the usual estimator for the covariance matrix of $\hat{\beta}$ requires that we add (S).
- Conventional finite-sample inference requires that we add (N).

Constrained Estimators

Last day we derived a test statistic for nested linear hypotheses under normality of $\mathbf{y}$. We stated the conditions under which the test statistic had a known sampling distribution in terms of linear restrictions on the regressors $\mathbf{X}$. However, because we state the null hypothesis in terms of β, it is more natural to think in terms of linear restrictions on β. We will therefore focus today on the following general statement of the null and alternative hypotheses:

$$H_0 \quad : \quad \mathbf{R}\beta = \mathbf{r} \tag{1}$$
$$H_1 \quad : \quad \mathbf{R}\beta \neq \mathbf{r} \tag{2}$$

where $\mathbf{R}$ is $q \times k$, $\mathbf{r}$ is $q \times 1$, and $q < k$. We assume that $rank\,(\mathbf{R}) = q$, so that there are genuinely q independent restrictions under H_0.

Let $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y}$ be the unconstrained (least squares) estimator, and let $\mathbf{b}$ be the constrained estimator satisfying $\mathbf{Rb} = \mathbf{r}$. Typically,[1] $\mathbf{R}\hat{\beta} \neq \mathbf{r}$.

Proposition 1 *Given (D),* $\mathbf{b} = \hat{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{R}'\left[\mathbf{R}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{R}'\right]^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right).$

Proof. The constrained estimator solves

$$\min_{\tilde{\mathbf{b}}} \left(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}}\right)'\left(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}}\right) \text{ subject to } \mathbf{R}\tilde{\mathbf{b}} = \mathbf{r}.$$

Thus we have the Lagrangean for the constrained optimization problem:

$$L = \left(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}}\right)'\left(\mathbf{y} - \mathbf{X}\tilde{\mathbf{b}}\right) - 2\lambda\left(\mathbf{R}\tilde{\mathbf{b}} - \mathbf{r}\right)$$

with first order conditions

$$-\mathbf{X}'\mathbf{y} + \mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{R}'\lambda = \mathbf{0} \tag{3}$$
$$\mathbf{R}\mathbf{b} - \mathbf{r} = \mathbf{0} \tag{4}$$

Premultiplying (3) by $(\mathbf{X}'\mathbf{X})^{-1}$ gives

$$\begin{aligned} \mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{y} + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{R}'\lambda \\ &= \hat{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{R}'\lambda. \end{aligned} \tag{5}$$

We need to eliminate λ. Premultiplying by $\mathbf{R}$ and using the constraint (4) gives

$$\mathbf{r} = \mathbf{R}\hat{\beta} + \mathbf{R}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{R}'\lambda.$$

Premultiplying by $\left[\mathbf{R}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{R}'\right]^{-1}$ gives

$$\left[\mathbf{R}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{R}'\right]^{-1}\mathbf{r} = \left[\mathbf{R}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{R}'\right]^{-1}\mathbf{R}\hat{\beta} + \lambda$$

[1] In fact, since our inference is being conducted under the assumption that ε is normally distributed, we can go further: $\Pr(\mathbf{R}\hat{\beta} = \mathbf{r}) = 0$.

and hence $\lambda = \left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{r}-\mathbf{R}\hat{\beta}\right)$. Substituting this expression for λ into (5) gives

$$\mathbf{b} = \hat{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{r}-\mathbf{R}\hat{\beta}\right).$$

■

Note that Proposition 1 gives us a way to recover the constrained estimator from the unconstrained estimator and knowledge of $\mathbf{X}$. You should also note that this is entirely an algebraic property of the constrained estimator.

Now we'll derive the sampling distribution of $\mathbf{b}$. To do so, we'll use our decomposition of $\hat{\beta}$ into its expected value plus sampling error to obtain a similar decomposition for $\mathbf{b}$.

Proposition 2 *Given (D,B,L), the constrained estimator is unbiased under the null hypothesis (1). That is $E[\mathbf{b}] = \beta$ under H_0.*

Proof. Substitute $\hat{\beta} = \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon$ into the definition of $\mathbf{b}$:

$$\begin{aligned}
\mathbf{b} &= \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{r}-\mathbf{R}\beta-\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon\right) \\
&= \beta + \left[\mathbf{I}_k - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\mathbf{R}\right]\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon \\
&\quad + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\left[\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{r}-\mathbf{R}\beta\right)
\end{aligned}$$

Imposing $\mathbf{R}\beta = \mathbf{r}$ under H_0 and taking expectations gives the result. ■

Proposition 3 *Given (D,B,L,S), under the null hypothesis (1), the constrained estimator has smaller variance than the least squares estimator. That is $Var[\mathbf{b}] \le Var\left[\hat{\beta}\right]$ under H_0.*

Proof. Let $\mathbf{A} = \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'$. Under H_0,

$$\mathbf{b} - \beta = \left[\mathbf{I}_k - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\right]\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon.$$

Therefore

$$\begin{aligned}
Var[\mathbf{b}] &= E\left[(\mathbf{b}-\beta)(\mathbf{b}-\beta)'\right] \\
&= \left[\mathbf{I}_k - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\right]\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left[\varepsilon\varepsilon'\right]\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left[\mathbf{I}_k - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\right]' \\
&= \sigma^2\left[\mathbf{I}_k - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\right]\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left[\mathbf{I}_k - \mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\
&= \sigma^2\left[\begin{array}{c} \left(\mathbf{X}'\mathbf{X}\right)^{-1} - 2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{array}\right] \\
&= \sigma^2\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} - 2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\
&= \sigma^2\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1} - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\mathbf{A}^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\
&\le \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \qquad \text{(why?)} \qquad (6) \\
&= Var\left[\hat{\beta}\right].
\end{aligned}$$

∎

How does this result relate to the Gauss-Markov Theorem? You can show that $\mathbf{b}$ is BLUE when $\mathbf{R}\beta = \mathbf{r}$. However, it is biased when $\mathbf{R}\beta \neq \mathbf{r}$. The unrestricted least squares estimator is BLUE for a larger class of models, including the case where $\mathbf{R}\beta \neq \mathbf{r}$.

Proposition 4 *Given (D, B, L, S, N) (i.e. under normality of $\mathbf{y}$), $\mathbf{b}$ is normally distributed (whether or not the null holds).*

Some Things for You to Prove

Proposition 5 *Let $\mathbf{e} = \mathbf{y} - \mathbf{X}b$ be the residuals from the constrained model. An unbiased estimator of σ^2 in the constrained model is*

$$\mathbf{s}^2 = \frac{\mathbf{e}'\mathbf{e}}{n - (k - q)}.$$

Proposition 6 *Under normality of $\mathbf{y}$, the maximum likelihood estimator of β in the constrained model is $\mathbf{b}_{ML} = \mathbf{b}$. As usual, the maximum likelihood estimator of σ^2 is*

$$\sigma^2_{ML} = \frac{\mathbf{e}'\mathbf{e}}{n}.$$

Testing the Linear Restrictions

Having reformulated our nested linear hypotheses in terms of linear restrictions on β, we now relate these to the test statistic we derived last day. Recall the test statistic was

$$F = \frac{(\mathbf{e}'\mathbf{e} - \mathbf{e}^{*\prime}\mathbf{e}^*) / (tr(\mathbf{M}) - tr(\mathbf{M}^*))}{\mathbf{e}^{*\prime}\mathbf{e}^* / tr(\mathbf{M}^*)} \sim F_{tr(\mathbf{M}) - tr(\mathbf{M}^*), tr(\mathbf{M}^*)} \tag{7}$$

where $\mathbf{e}$ is the vector of residuals in the restricted model, $\mathbf{e}^*$ is the vector of unrestricted residuals, $tr(\mathbf{M}) - tr(\mathbf{M}^*)$ is the number of restrictions being imposed, and $tr(\mathbf{M}^*)$ is the degrees of freedom in the unrestricted model. Our goal is to relate the loss of fit (i.e., the increase in the sum of squared residuals) from imposing the restrictions to the distance between $\mathbf{R}\hat{\beta}$ and $\mathbf{r}$.

Proposition 7 *Under (D),* $\mathbf{e}'\mathbf{e} - \mathbf{e}^{*\prime}\mathbf{e}^* = \left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)' \left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1} \left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)$

Proof.

$$\begin{aligned} \mathbf{e} &= \mathbf{y} - \mathbf{X}\mathbf{b} = \mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{X}\left(\mathbf{b} - \hat{\beta}\right) = \mathbf{e}^* - \mathbf{X}\left(\mathbf{b} - \hat{\beta}\right) \\ &\Rightarrow \mathbf{e}'\mathbf{e} = \mathbf{e}^{*\prime}\mathbf{e}^* + \left(\mathbf{b} - \hat{\beta}\right)' \mathbf{X}'\mathbf{X}\left(\mathbf{b} - \hat{\beta}\right) \end{aligned}$$

because the cross-product terms involve $\mathbf{X}'\mathbf{e}^* = \mathbf{0}$. We know from Proposition 1 that $\mathbf{b} - \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{A}^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)$ where $\mathbf{A} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ as before. Now notice that

$$\begin{aligned}
\left(\mathbf{b} - \hat{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\mathbf{b} - \hat{\beta}\right) &= \left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)'\mathbf{A}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{A}^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right) \\
&= \left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)'\mathbf{A}^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\mathbf{A}^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right) \\
&= \left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)'\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right) \\
&= \left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)'\mathbf{A}^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)
\end{aligned}$$

and therefore

$$\mathbf{e}'\mathbf{e} - \mathbf{e}^{*\prime}\mathbf{e}^* = \left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right).$$

■

This shows that the numerator of the test statistic F can be computed using only the unrestricted estimates $\hat{\beta}$. We already know that the denominator is computed from the residuals in the unrestricted model. Thus, we can restate the test statistic (7) solely in terms of the unrestricted model:

$$\begin{aligned}
F &= \frac{\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)/(n-k+q-(n-k))}{\mathbf{e}^{*\prime}\mathbf{e}^*/(n-k)} \\
&= \frac{\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)/q}{s^2} \\
&= \frac{\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)'\left[s^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\left(\mathbf{r} - \mathbf{R}\hat{\beta}\right)}{q} \qquad (8)
\end{aligned}$$

where s^2 is our unbiased estimator of the error variance in the unrestricted model. Despite this result, we frequently estimate both the restricted and unrestricted models and compute the test statistic (7) because least squares is so computationally cheap.

Example 8 *Consider the model* $\mathbf{y} = \beta_0 + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \varepsilon$ *with the restriction* $\beta_1 + \beta_2 = 3$. *One approach is to substitute for* β_1, *which yields*

$$\begin{aligned}
\mathbf{y} &= \beta_0 + (3 - \beta_2)\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \varepsilon \\
&\Rightarrow \mathbf{y} - 3\mathbf{x}_1 = \beta_0 + \beta_2(\mathbf{x}_2 - \mathbf{x}_1) + \varepsilon.
\end{aligned}$$

So we can regress $(\mathbf{y} - 3\mathbf{x}_1)$ *on* $(\mathbf{x}_2 - \mathbf{x}_1)$ *and a constant term, and collect the restricted residuals to compute* $\mathbf{e}'\mathbf{e}$. *We then estimate the unrestricted regression of* $\mathbf{y}$ *on* $\mathbf{x}_1, \mathbf{x}_2$ *and a constant term, collect the unrestricted residuals, and compute* $\mathbf{e}^{*\prime}\mathbf{e}^*$. *Then we can test the null* $H_0 : \beta_1 + \beta_2 = 3$ *using the test statistic:*

$$F = \frac{(\mathbf{e}'\mathbf{e} - \mathbf{e}^{*\prime}\mathbf{e}^*)}{\mathbf{e}^{*\prime}\mathbf{e}^*/(n-3)} \sim F_{1,n-3}.$$

Alternately, we could avoid estimating the restricted regression and construct the equivalent test statistic based solely on the unrestricted regression:

$$F = \left(3 - \hat{\beta}_1 - \hat{\beta}_2\right)^2 \left[s^2 \mathbf{R} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{R}'\right]^{-1} \sim F_{1,n-3}$$

where

$$\mathbf{R} = [\ 0\ \ 1\ \ 1\ ],\ \mathbf{r} = 3,\ \hat{\beta}' = \left[\ \hat{\beta}_0\ \ \hat{\beta}_1\ \ \hat{\beta}_2\ \right].$$