Семинар 6.

Тестирование гипотез. Теорема Фриша-Во-Ловелла.

Сложение блочных матриц. Две матрицы, разбитые на блоки одинаковым образом, можно складывать:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] + \left[\begin{array}{c|c} E & F \\ \hline G & H \end{array}\right] = \left[\begin{array}{c|c} A+E & B+F \\ \hline C+G & D+H \end{array}\right].$$

Умножение блочных матриц. Если размеры блоков допускают операцию умножения, то:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] \cdot \left[\begin{array}{c|c} E & F \\ \hline G & H \end{array}\right] = \left[\begin{array}{c|c} AE+BG & AF+BH \\ \hline CE+DG & CF+DH \end{array}\right].$$

Транспонирование блочных матриц. Пусть матрица

$$M = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right].$$

Тогда

$$M^T = \left[\begin{array}{c|c} A^T & C^T \\ \hline B^T & D^T \end{array}\right].$$

Формула Фробениуса (блочное обращение).

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]^{-1} = \left[\begin{array}{c|c} A^{-1}+A^{-1}BH^{-1}CA^{-1} & -A^{-1}BH^{-1} \\ \hline -H^{-1}CA^{-1} & H^{-1} \end{array}\right],$$

где A — невырожденная квадратная матрица размерности $n \times n$, D — квадратная матрица размерности $k \times k$, $H = D - CA^{-1}B$.

1. Рассмотрим модель множественной регрессии:

$$y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i.$$

Мы хотим проверить гипотезу следующего вида:

$$H_0 : \beta_{k-q+1} = \beta_{k-q+2} = \ldots = \beta_k = 0.$$

Покажите, что для проверки гипотезы H_0 может быть использована статстистика

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)} \sim F(q,n-k).$$

Решение: см. учебник Магнус, Катышев, Пересецкий. Эконометрика (начальный курс), стр. 82-84. Обратите внимание, что в учебнике ESS — это сумма квадратов остатков, а RSS — объясненная сумма.

2. Множественную регрессию с любым количеством регрессоров можно разбить на несколько шагов со вспомогательными регрессиями с меньшим числом регрессоров.

   Вместо непосредственного включения переменной x в качестве регрессора в модель можно сначала «очистить» от переменной x зависимую переменную y и остальные регрессоры, а затем оценить регрессию для «очищенных» переменных. Эту идею последовательных регрессий формализует теорема Фриша — Во — Ловелла.

Теорема Фриша — Во — Ловелла (англ. Frisch–Waugh–Lovell theorem, FWL theorem)

Рассмотрим два алгоритма.

Алгоритм A: оцениваем регрессию y на полный набор регрессоров X_1 и X_2 с помощью МНК:

$$\hat{y}_A = X_1\hat{\beta}_1^A + X_2\hat{\beta}_2^A.$$

Алгоритм B:

B1. Оцениваем регрессию y на часть регрессоров X_1 с помощью МНК:

$$\hat{y}_B = X_1\hat{\beta}_1^B.$$

B2. Оцениваем регрессию каждого столбца из матрицы X_2 с помощью МНК:

$$\hat{X}_2^B = X_1\hat{\beta}_1^B.$$

Уточним, что здесь $\hat{\beta}_1^B$ — это не вектор, а целая матрица, в которой содержатся оценки регрессии каждого столбца из матрицы X_2 на все регрессоры из матрицы X_1.

B3. Определяем «очищенные» переменные как остатки регрессий первых двух шагов,

$$y^* = y - \hat{y}_B, \quad X_2^* = X_2 - \hat{X}_2^B.$$

B4. Оцениваем регрессию для «очищенных переменных»

$$\hat{y}^* = X_2^*\hat{\beta}_2^B.$$

Алгоритмы A и B дают одинаковые оценки коэффициентов $\hat{\beta}_2^A = \hat{\beta}_2^B$ и финальные векторы остатков $e_A = y - \hat{y}_A = y^* - \hat{y}^* = e_B^*$.

Докажите теорему.

Док-во теоремы:

Определим матрицу-шляпницу P_1, проецирующую на линейную оболочку столбцов блока X_1, и матрицу $M_1 = I - P_1$, проецирующую на ортогональное дополнение к линейной оболочке столбцов X_1,

$$P_1 = X_1(X_1^T X_1)^{-1}X_1^T, \quad M_1 = I - P_1.$$

По определению, $H_1 X_1 = X_1$ и $M_1 X_1 = 0$.

Возьмём результат выполнения алгоритма A

$$y = X_1\hat{\beta}_1^A + X_2\hat{\beta}_2^A + e_A$$

и домножим его на матрицу M_1:

$$M_1 y = 0 \cdot \hat{\beta}_1^A + M_1 X_2\hat{\beta}_2^A + M_1 e_A$$

Заметим, что остатки e_A алгоритма A ортогональны и регрессорам из блока X_1, и регрессорам из блока X_2. Сначала воспользуемся тем, что остатки e_A уже лежат в подпространстве, ортогональном регрессорам блока X_1. Дополнительное проецирование в это подпространство никак их не изменяет, $M_1 e_A = e_A$. Следовательно,

$$M_1 y = M_1 X_2\hat{\beta}_2^A + e_A$$

Теперь воспользуемся тем, что остатки e_A уже лежат в подпространстве, ортогональном регрессорам блока X_2, поэтому $X_2^T e_A = 0$.

Ортогональны ли остатки e_A и столбцы матрицы $M_1 X_2$? Проверим!

$$(M_1 X_2)^T e_A = X_2^T M_1^T e_A = X_2^T M_1 e_A = X_2^T e_A = 0.$$

Остаётся лишь сказать, что умножение на матрицу M_1 очищает переменные, $M_1 y = y^*$ и $M_1 X_2 = X_2^*$,

$$y^* = X_2^*\hat{\beta}_2^A + e_A,$$

И мы видим идеальное совпадение с разложением алгоритма B,

$$y^* = X_2^*\hat{\beta}_2^B + e_B^*$$

В силу единственности разложения по ортогональному базису $\hat{\beta}_2^A = \hat{\beta}_2^B$ и $e_A = e_B^*$.

3. Вместо того чтобы оценивать параметры β_1, β_2 в модели

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

(X_1, X_2 — $n \times k_1, n \times k_2$ матрицы соответственно, β_1, β_2 — векторы размерности k_1, k_2 соответственно), строятся МНК оценки этих параметров исходя из модели

$$y = X_1^*\beta_1^* + X_2\beta_2^* + \varepsilon^*,$$

где X_1^* — матрица остатков, полученных в результате регрессии каждого столбца матрицы X_1 на X_2.

(a) Покажите, что полученная таким образом оценка вектора β_2 совпадает с оценкой, полученной в результате регрессии y только на X_2.

(b) Найдите смещение оценки вектора β_2.

(c) Покажите, что оценки вектора параметров β_1 в обеих моделях совпадут.

Решение:

(a) Из определения матрицы X_1^* следует, что каждый ее столбец ортогонален каждому столбцу матрицы X_2, что эквивалентно равенству $X_2^T X_1^* = 0$. Обозначим $\hat{\beta}^* = (\hat{\beta}_1^*, \hat{\beta}_2^*)'$ МНК-оценку вектора $\beta = (\beta_1, \beta_2)^T$ во второй регрессии. По правилам действия с блочными матрицами получаем:

$$\hat{\beta}^* = \left(\begin{bmatrix} X_1^* & X_2 \end{bmatrix}^T \begin{bmatrix} X_1^* & X_2 \end{bmatrix}\right)^{-1} \begin{bmatrix} X_1^* & X_2 \end{bmatrix}^T y$$

$$= \left(\begin{bmatrix} X_1^{*T} \\ X_2^T \end{bmatrix} \begin{bmatrix} X_1^* & X_2 \end{bmatrix}\right)^{-1} \begin{bmatrix} X_1^{*T} \\ X_2^T \end{bmatrix} y = \begin{bmatrix} X_1^{*T} X_1^* & X_1^{*T} X_2 \\ X_2^T X_1^* & X_2^T X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1^{*T} y \\ X_2^T y \end{bmatrix}$$

$$= \begin{bmatrix} X_1^{*T} X_1^* & 0 \\ 0 & X_2' X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1^{*T} y \\ X_2^T y \end{bmatrix} = \begin{bmatrix} (X_1^{*T} X_1^*)^{-1} X_1^{*T} y \\ (X_2^T X_2)^{-1} X_2^T y \end{bmatrix}.$$

Таким образом, $\hat{\beta}_2^* = (X_2' X_2)^{-1} X_2' y$, что совпадает с МНК-оценкой вектора β_2 в регрессии y только на X_2.

(b) Имеем:

$$\hat{\beta}_2^* = (X_2^T X_2)^{-1} X_2^T y = (X_2^T X_2)^{-1} X_2^T (X_1\beta_1 + X_2\beta_2 + \varepsilon)$$

$$= \beta_2 + (X_2^T X_2)^{-1} X_2^T X_1 \beta_1 + (X_2^T X_2)^{-1} X_2^T \varepsilon.$$

Отсюда

$$\mathbb{E}(\hat{\beta}_2^*) = \beta_2 + (X_2^T X_2)^{-1} X_2^T X_1 \beta_1,$$

т.е. смещение оценки $\hat{\beta}_2^*$ равно $(X_2^T X_2)^{-1} X_2^T X_1 \beta_1$.
Смещение будет равно нулю, если $X_2^T X_1 = 0$, то есть когда регрессоры из набора X_1 некоррелированы с регрессорами из набора X_2.

(c) МНК-оценки компонент вектора $\beta = (\beta_1, \beta_2)^T$ в регрессии (1) получаются как коэффициенты разложения ортогональной проекции $\hat{y}$ вектора y в евклидовом пространстве $\mathbb{R}^n$ на подпространство, порожденное столбцами матрицы $X = [X_1 \, X_2]$:

$$\hat{y} = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2.$$

Подпространства, порожденные столбцами матриц $[X_1 \, X_2]$ и $[X_1^* \, X_2]$, совпадают и, следовательно, совпадают ортогональные проекции на них вектора y. То есть

$$X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 = X_1^* \hat{\beta}_1^* + X_2 \hat{\beta}_2^*.$$

Так как столбцы матрицы X_1^* представляют собой остатки регрессий столбцов матрицы X_1 на матрицу X_2, то

$$X_1^* = X_1 - X_2 (X_2^T X_2)^{-1} X_2^T X_1.$$

Получаем

$$X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 = X_1^* \hat{\beta}_1^* + X_2 (\hat{\beta}_2 - (X_2^T X_2)^{-1} X_2^T X_1 \hat{\beta}_1) = X_1^* \hat{\beta}_1^* + X_2 \hat{\gamma}^*,$$

где $\hat{\gamma}^* = \hat{\beta}_2 - (X_2^T X_2)^{-1} X_2^T X_1 \hat{\beta}_1$.
Так как матрица X имеет по предположению полный ранг, то ее столбцы линейно независимы, и поэтому из последнего соотношения вытекают следующие равенства:

$$\hat{\beta}_1 = \hat{\beta}_1^*,$$
$$\hat{\beta}_2 = \hat{\gamma}^* = \hat{\beta}_2 - (X_2^T X_2)^{-1} X_2^T X_1 \hat{\beta}_1,$$

чтд.

Заметим также, что из этих равенств следует, что

$$\hat{\beta}_2^* = \hat{\beta}_2 + (X_2^T X_2)^{-1} X_2^T X_1 \hat{\beta}_1,$$

что согласуется с (b).