Econometrics II: Assignment 4

Due: Sunday, March 16th

1 This question concerns the system of wage equations

$$\begin{aligned} LW69 &= \alpha + \beta \text{schooling69} + \gamma \text{IQ} + \eta_1 \\ LW80 &= \alpha + \beta \text{schooling80} + \gamma \text{IQ} + \eta_2. \end{aligned}$$

Assume that the data set also features the variable mother's.education and that the following moment conditions apply:

$$\mathbb{E}(\eta_1) = \mathbb{E}(\eta_2) = \mathbb{E}\left(\text{mother's.education} \times \eta_1\right) = \mathbb{E}\left(\text{mother's.education} \times \eta_2\right) = 0.$$

1.1 Is the first equation identified on its own? Is the second?

1.2 Write out the exact matrix to which the rank condition for identification applies and state the rank that this matrix must have.

1.3 Assume that $\text{cov}\left(\text{IQ}, \text{mother's.education}\right) = 0$. Which parameters are identified?

1.4 Do you find the moment conditions compelling?

2 (from 2012 midterm exam) Consider the system of equations

$$\begin{aligned} Y_1 &= X_1\beta_1 + \varepsilon_1 \\ Y_2 &= X_2\beta_2 + \varepsilon_2 \\ Y_3 &= X_3\beta_3 + \varepsilon_3 \end{aligned}$$

and the moment conditions

$$\mathbb{E}(Z(Y_m - X_m\beta_m)) = 0, m = 1, 2, 3.$$

Assume that the vectors (X_1, X_2, X_3) are all of length 3 and do not have any components in common, whereas Z is of length 5. Assume also homoskedasticity:

$$\mathbb{E}([\varepsilon_1, \varepsilon_2, \varepsilon_3]'[\varepsilon_1, \varepsilon_2, \varepsilon_3]|X_1, X_2, X_3, Z) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}.$$

Finally, assume that further regularity conditions hold as needed.

We will compare two researchers. Researcher 1 estimates this model by an efficient two-step GMM procedure as discussed in the lecture. Researcher 2 does the same thing but optimizes her choice of weighting matrix $\hat{\boldsymbol{W}}$ subject to the constraint that

$$\hat{\boldsymbol{W}} = \begin{bmatrix} \underset{[5\times5]}{\hat{\boldsymbol{W}}_1} & \underset{[5\times5]}{\mathbf{0}} & \underset{[5\times5]}{\mathbf{0}} \\ \underset{[5\times5]}{\mathbf{0}} & \underset{[5\times5]}{\hat{\boldsymbol{W}}_2} & \underset{[5\times5]}{\mathbf{0}} \\ \underset{[5\times5]}{\mathbf{0}} & \underset{[5\times5]}{\mathbf{0}} & \underset{[5\times5]}{\hat{\boldsymbol{W}}_3} \end{bmatrix}.$$

2.1 I claim that researcher 2's approach is really a one-step GMM procedure. Give a concise explanation of what I mean, including a brief algebraic demonstration.

2.2 Compare the procedures. Are both estimators consistent? Are they asymptotically normal? Assuming all conditions hold as stated, is one of them preferred and why? Explain.

2.3 Assume now that all the above assumptions continue to hold except that

$$\mathbb{E}(Z(Y_1 - X_1\beta_1)) = (0, 0, 0, a, b)'$$

with $a, b \neq 0$. How does your answer to 3.2 change? How – if at all – does the gist of your answer change if you learn that researchers are only really interested in β_3?

3. Empirical Exercise The data for the following problem are uploaded. It might be helpful to read Griliches (1976), which is also uploaded. The exercise replicates part of Griliches with the data from Blackburn/Neumark (1992). For a description of the data, check that paper, which is uploaded as well.

The variables are
RNS = southern resident,
MRT = marital status (1=married),
SMSA = residency in metropolitan area,
MED = mother's education,
KWW = knowledge of the world score,
IQ = IQ score,
AGE = age,
S = schooling (years of),
EXPR = work experience (years),
TENURE = tenure (years),
LW = log wage,
YEAR = year of first data point.

The variables are provided for the earliest year in which the individual was observed (coded in YEAR) and for 1980. After deleting black individuals and cases with missing data on MED, the sample size is 758.

3.1 Calculate means and standard deviations for all variables and prepare a table similar to table 1 in Griliches. Also, calculate the correlation between IQ and S.

3.2 Generate year dummies for 1966-1973, noting that no data exist for 1972. Consider the equation

$$\begin{aligned} LW &= \beta S + \gamma IQ + \delta' h + \varepsilon \\ h &= (EXPR, TENURE, RNS, SMSA, yeardummies)'. \end{aligned}$$

Prepare a table similar to 3.2 in Hayashi's textbook (reproduced in this homework), which in turn replicates some results in Griliches. In the 2SLS estimation, the instruments are the exogenous regressors and the excluded variables, i.e. $(S, \boldsymbol{h}, MED, KWW, MRT, AGE)$. Does the comparison of coefficients conform to what one might have expected in this case?

3.3 For the 2SLS estimation, calculate Sargan's (J-)statistic. What are the degrees of freedom? What is the p-value? (The statistic should be 87.655.)

3.4 Obtain the TSLS estimate from actually running two regressions. The standard errors should be different compared to using a TSLS package. Explain.

3.5 Griliches mentions that schooling may also be endogenous. Why? Estimate the equation by 2SLS, treating both IQ and S as endogenous. What happens to the schooling coefficient? How would you explain the difference? Calculate Sargan's statistic (13.268) and its p-value.

3.6 Estimate the equation by GMM, treating schooling as predetermined as in part 2. Are TSLS standard errors smaller than GMM standard errors? Test whether schooling is predetermined by the C-statistic. For this, you need to do two GMM's with and without schooling as an instrument. For both, use an $\hat{\boldsymbol{W}}$ calculated from the TSLS residual from part 2. (The statistic should be 58.168.)

3.7 Reconsider the estimation strategy from part 5. The large test statistics in 5 and 6 are suspicious. Consider dropping MED and KWW as instruments. What happens to the TSLS estimates? The schooling coefficient should be -529.3%. What do you think is the problem?

4 This exercise is from Hayashi's textbook. For a full understanding of the background, consult Mankiw, Romer, and Weil (1992), which is uploaded.

We use the international panel constructed by Summers and Heston (1991, also uploaded), which has become the standard data set for studying the growth of nations. The Summers-Heston panel (also called the Penn World Table) includes major macro variables measured in a consistent basis for most countries of the world. The file $\texttt{SUM}-\texttt{HES.ASC}$ is an extract for 1960-1985 from version 5.6 of the Penn World Table downloaded from the NBER's home page (www.nber.org). For each country, information on the country is con- tained in multiple records (rows). The first record has the country's identification code and two dummies, one for a communist regime (referred to as COM below) and the other for the Organization of Petroleum Exporting Countries (referred to as OPEC below). The second record has the year, the population (in thousands), real GDP per capita (in 1985 U.S. dollars), and the saving rate (in percent) for 1960. The third record has the same for 1961, and so forth. The twenty-seventh record of the country is for 1985.

The file contains all 125 countries for which the data are available. The mapping between the country ID and the country name is in `country.asc`. Country 1 is Algeria, 2 is Angola, etc. Note that the mapping in Penn World Table version 5.6 is different from that in Summers and Heston (1991). GDP per capita is the country's real GDP converted into U.S. dollars using the purchasing power parity of the country's currency against the dollar for 1985 (the variable "RGDPCH" in the Penn World Table). The saving rate is measured as the ratio of real investment to real GDP ("I" in the Penn World Table). See Sections I1 and I11 (particularly 1II.D) of Summers and Heston (1991) for how the variables are constructed.

The first issue we examine empirically is conditional convergence (cf. Mankiw, Romer, and Weil, 1992). Let S be savings rate and N be the population growth rate of countries (which we take to be the growth rate of labor). In the Solow-Swan growth model with a Cobb-Douglas production function, the steady-state level of output per effective labor, denoted $q*$ in the text, is a linear function of $log(S) - log(N + g + \delta)$, where g is the rate of labor-augmenting technical progress and δ is the depreciation rate of the capital stock. Then, assuming $A(O)$, the initial level of technology, to be the same across countries, one can write

$$Y_t = \phi_t + \rho Y_{t-1} + \gamma_1 \log(S) + \gamma_2 \log(N + g + \delta) + \eta_t.$$

The growth model implies that $\gamma_1 = -\gamma_2$, but we will not impose this restriction. This is the specification estimated in Table IV of Mankiw et al. (1992). As there, assume $g + \delta$ to be 5 percent for all countries, and take S to be the saving rate averaged over the 1960-1985 period. We measure N as the average annual population growth rate over the same period (Mankiw et al. uses the average growth rate of the working-age population). Because our sample of 125 countries includes (former) communist countries and OPEC, we add COM (= 1 if communist, = 0 otherwise) and $OPEC$ (= 1 if OPEC, = 0 otherwise) to the equation:

$$Y_t = \phi_t + \rho Y_{t-1} + \gamma_1 \log(S) + \gamma_2 \log(N + g + \delta) + \gamma_3 COM + \gamma_4 OPEC + \eta_t.$$

4.1 (OLS) By subtracting from both sides, we can rewrite the above as

$$Y_t - Y_{t-1} = \phi_t + (\rho - 1)Y_t - 1 + \gamma_1 \log(S) + \gamma_2 \log(N + g + \delta) + \gamma_3 COM + \gamma_4 OPEC + \eta_t.$$

For this specification, set $T = 1$ (just one equation to estimate) and take $t_0 = 1960, t_2 = 1985$, so the dependent variable, $Y_1 - Y_0$, is the cumulative growth between 1960 and 1985. Plot $Y_1 - Y_0$ against Y_0 for the 125 countries included in the sample. Does this suggest any relation between the initial GDP and subsequent growth (i.e., convergence)? Assuming outright that the error term is orthogonal to the regressors, estimate the above by OLS. Calculate the value of λ (speed of convergence) implied by $\hat{\rho}$. (It should be less than 1 percent per year.) Calculate a standard error for your estimate. (Hint: Use the delta-method to show that the standard error of $\hat{\lambda}$ is the standard error of $\hat{\rho}$ divided by $25\hat{\rho}$).) Can you confirm the finding that "if countries did not vary in their investment and population growth rates, there would be a strong tendency for poor countries to grow faster than rich ones" (Mankiw et al., 1992, p. 428).

4.2 (Simple Fixed Effects) By setting $M = 25$, $t_0 = 1960$, $t_l = 1961$,..., $t_{25} = 1985$, we can form a system of 25 equations. Estimate the system assuming fixed effects, assuming that error terms are i.i.d. in both dimensions. What is the implied value of λ? (It should be about 6.4 percent per year.) How does the result change under 2-step (efficient GMM) estimation (using the same moment conditions)?

4.3 (IV) As in the previous question, take $M = 25$, $t_0 = 1960$, $t_l = 1961$,..., $t_{25} = 1985$. We can then generate a system of 24 equations by first-differencing:

$$\Delta Y_t = Z_t' \delta + \epsilon_t,$$

where t runs from 2 to 25 and $\Delta Y_t \equiv Y_t - T_{t-1}$. Also,

$$Z_t = (0, \ldots, 0, 1, 0, \ldots, 0, \Delta Y_{t-1})',$$

where the 1 entry is in position t, and

$$\delta = (\mu_2, \ldots, \mu_{25}, \rho)'.$$

If S_t is a country's saving rate (investment/GDP) in year t, use a constant and S_{t-1} as the instrument in the t-th equation. (For example, in the first equation where the nonconstant regressor is $Y_{1961} - Y_{1960}$, the vector of instruments is $(1, S_{1960})')$.) Estimate this by multiple equation GMM assuming conditional homoskedasticity. The implied estimate of λ should be about 36 percent per year.

Answer Key

1.1 For either equation, there are 3 regressors and 2 instruments (counting in the constants), so the equations are not identified.

1.2 The system can be written as

$$\begin{bmatrix} 1 & \mathbb{E}\mathrm{s69}_i & \mathbb{E}\mathrm{IQ}_i \\ 1 & \mathbb{E}\mathrm{s80}_i & \mathbb{E}\mathrm{IQ}_i \\ \mathbb{E}\mathrm{med}_i & \mathbb{E}\left(\mathrm{s69}_i \cdot \mathrm{med}_i\right) & \mathbb{E}\left(\mathrm{IQ}_i \cdot \mathrm{med}_i\right) \\ \mathbb{E}\mathrm{med}_i & \mathbb{E}\left(\mathrm{s80}_i \cdot \mathrm{med}_i\right) & \mathbb{E}\left(\mathrm{IQ}_i \cdot \mathrm{med}_i\right) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} \mathbb{E}LW69_i \\ \mathbb{E}LW80_i \\ \mathbb{E}\left(LW69_i \cdot \mathrm{med}_i\right) \\ \mathbb{E}\left(LW80_i \cdot \mathrm{med}_i\right) \end{bmatrix},$$

which must have full column rank (=3). Thus the system may well be identified even though no individual equation is, illustrating a point that was made in the lecture.

1.3 If IQ and mother's education are uncorrelated, then

$$\underbrace{\begin{bmatrix} \mathbb{E}\mathrm{IQ}_i \\ \mathbb{E}\mathrm{IQ}_i \\ \mathbb{E}\left(\mathrm{IQ}_i \cdot \mathrm{med}_i\right) \\ \mathbb{E}\left(\mathrm{IQ}_i \cdot \mathrm{med}_i\right) \end{bmatrix}}_{3^{rd}\text{ column above}} = \mathbb{E}\mathrm{IQ}_i \cdot \underbrace{\begin{bmatrix} 1 \\ 1 \\ \mathbb{E}\mathrm{med}_i \\ \mathbb{E}\mathrm{med}_i \end{bmatrix}}_{1^{st}\text{ column above}},$$

hence the matrix cannot have full column rank. (If schooling were orthogonal to IQ, its coefficient could still be identified, but this is clearly not going to be the case, so the underidentification is contagious here.)

Interpretation: If this covariance is zero, the instrument does not, in fact, shift the endogenous regressor; in terms of intuitions developed in the very first lecture, using it as an instrument would be like using external coin tosses as instruments.

1.4 The moment conditions reflect the idea that mother's education influences the outcome only through the endogenous regressors. I would argue that this is *not* compelling in the example.

2.1 Researcher 2 performs equation-by-equation 2SLS. The separability of the estimator by equation is easily seen when writing down the closed form, recalling that the inverse of a block diagonal matrix is also block diagonal. Frequent slip: Just because R2 does an equation-by-equation method does not make it one-step. It's one-step because the additional assumption of homoskedasticity turns it into equation-by-equation 2SLS.

2.2 Both estimators are consistent and asymptotically normal but under the assumptions given, separate estimation is generally inefficient, i.e. comes with the larger asymptotic variance.

2.3 This question is about contagion of misspecification. Researcher 1 now has inconsistent estimators in all equations, researcher 2 only in equation 1. This illustrates a robustness-efficiency trade-off.