Advanced Econometrics: Demonstration of SEM, IV, 2SLS, and 3SLS using Klein's Model I

Prepared by

Aditya Korekallu Srinivasa

Scientist (Senior Scale)

ICAR-Indian Agricultural Research Institute
New Delhi

April 25, 2025

Contents

1 Klein's Model I: Structure and Economic Meaning

1.1 Background

Klein's Model I is a classic simultaneous equations model (SEM) of the U.S. economy for the period 1920–1941. It is used to illustrate the estimation of SEMs, IV, 2SLS, and 3SLS in many econometrics textbooks, including Gujarati & Porter.

1.2 The Equations

Klein's Model I consists of three behavioral equations:

$$
\begin{aligned}
C_t &= a_0 + a_1 P_t + a_2 Wg_t + a_3 P_{t-1} + u_{1t} \quad \text{(Consumption)} \\
I_t &= b_0 + b_1 P_t + b_2 P_{t-1} + b_3 K_{t-1} + u_{2t} \quad \text{(Investment)} \\
W_t &= c_0 + c_1 X_t + c_2 X_{t-1} + c_3 t + u_{3t} \quad \text{(Wage Bill)}
\end{aligned}
$$

where:

- C_t: Consumption expenditure
- P_t: Profits (endogenous)
- Wg_t: Wage bill of government (exogenous)
- P_{t-1}: Lagged profits (predetermined)
- I_t: Investment
- K_{t-1}: Lagged capital stock (predetermined)
- W_t: Wage bill (endogenous)
- X_t: Output (endogenous)
- X_{t-1}: Lagged output (predetermined)
- t: Time trend (exogenous)
- u_{1t}, u_{2t}, u_{3t}: Structural disturbances

1.3 Endogeneity in the Model

- **Endogenous variables**: C_t, I_t, W_t, P_t, X_t
- **Predetermined/exogenous variables**: Wg_t, P_{t-1}, K_{t-1}, X_{t-1}, t
- **Why endogenous?** For example, P_t (profits) is determined within the system and appears as a regressor in both the consumption and investment equations. OLS estimation of these equations would be inconsistent because P_t is correlated with the error terms via the simultaneous system.

2 Step-by-Step Estimation in R: Explanation and Code

2.1 Step 1: Load Data and Prepare Variables

The KleinI dataset is available in the AER package in R. Lagged variables are constructed to match the model.

```
library(AER)
data("KleinI")
# Create lagged variables
KleinI$P_1 = c(NA, KleinI$P[-nrow(KleinI)])      # P_{t-1}
KleinI$K_1 = c(NA, KleinI$K[-nrow(KleinI)])      # K_{t-1}
KleinI$X_1 = c(NA, KleinI$X[-nrow(KleinI)])      # X_{t-1}
KleinI$t = 1:nrow(KleinI)
klein <- na.omit(KleinI)
```

Explanation: - Lagged variables are created by shifting the column by one time period. - The first row is removed because lagged values are undefined.

2.2 Step 2: OLS Estimation (for Comparison)

Estimate each equation by OLS, ignoring the simultaneity.

```
ols_C <- lm(C ~ P + Wg + P_1, data=klein)
ols_I <- lm(I ~ P + P_1 + K_1, data=klein)
ols_W <- lm(W ~ X + X_1 + t, data=klein)
summary(ols_C)
summary(ols_I)
summary(ols_W)
```

Explanation: - OLS is easy to implement, but inconsistent for equations where regressors are endogenous (e.g., P in C and I equations). - OLS is consistent for the wage equation if all regressors are exogenous.

2.3 Step 3: IV/2SLS Estimation (Single Equation)

Estimate the consumption equation using IV/2SLS, treating P as endogenous and using all exogenous variables in the system as instruments.

```
library(AER)
iv_C <- ivreg(C ~ P + Wg + P_1 | Wg + P_1 + K_1 + X + X_1 + t + G, data=klein)
summary(iv_C)
```

Explanation: - The first stage regresses P on all exogenous variables (instruments). - The second stage regresses C on the fitted values of P, Wg, and P_1. - This corrects for endogeneity of P.

2.4 Step 4: 2SLS Estimation (System)

Estimate all three equations jointly using 2SLS.

```r
library(systemfit)
eqC <- C ~ P + Wg + P_1
eqI <- I ~ P + P_1 + K_1
eqW <- W ~ X + X_1 + t
eqSystem <- list(consumption = eqC, investment = eqI, wage = eqW)
inst <- ~ Wg + P_1 + K_1 + X + X_1 + t + G
fit_2sls <- systemfit(eqSystem, method="2SLS", inst=inst, data=klein)
summary(fit_2sls)
```

Explanation: - Each equation is estimated by 2SLS, using all exogenous variables as instruments for endogenous regressors. - System estimation allows for joint hypothesis testing and comparison.

2.5 Step 5: 3SLS Estimation (System)

Estimate all three equations jointly using 3SLS, which is more efficient if errors are correlated across equations.

```r
fit_3sls <- systemfit(eqSystem, method="3SLS", inst=inst, data=klein)
summary(fit_3sls)
```

Explanation: - 3SLS combines 2SLS (for endogeneity) and GLS (for cross-equation error correlation). - More efficient than 2SLS if errors are correlated.

2.6 Step 6: Comparison and Interpretation

Compare OLS, IV/2SLS, and 3SLS estimates.

```r
cat("\n--- OLS Estimates (Consumption) ---\n")
print(coef(ols_C))
cat("\n--- IV/2SLS Estimates (Consumption) ---\n")
print(coef(iv_C))
cat("\n--- 2SLS System Estimates ---\n")
print(coef(fit_2sls))
cat("\n--- 3SLS System Estimates ---\n")
print(coef(fit_3sls))
```

Explanation: - OLS estimates are biased for equations with endogenous regressors. - IV/2SLS corrects for endogeneity in a single equation. - 2SLS (systemfit) estimates all equations by 2SLS. - 3SLS is most efficient if errors are correlated.

2.7 Step 7: Residual Correlation Matrix

Check if errors are correlated across equations (justifying 3SLS).

```r
cor(residuals(fit_2sls))
```

Explanation: - If off-diagonal elements are large, 3SLS is preferred.

2.8 Step 8: Joint Hypothesis Testing

Test if certain coefficients (e.g., lagged variables) are jointly zero.

```r
library(car)
Rmat <- rbind(
  c(0,0,0,1,0,0,0,0,0), # P_1 in C
  c(0,1,0,0,0,0,0,0,0), # P in I
  c(0,0,1,0,0,0,0,0,0), # Wg in C
  c(0,0,0,0,0,1,0,0,0)  # X_1 in W
)
rvec <- c(0,0,0,0)
linearHypothesis(fit_3sls, Rmat, rvec)
```

Explanation: - Joint hypothesis testing is easy in system estimation frameworks. - Here, we test if certain lagged variables' coefficients are zero in the system.

3 Summary Table: Methods and Their Properties

Method	Endogeneity	Error Correlation	Efficiency
OLS	Not addressed	Not addressed	Inconsistent if endogeneity
IV/2SLS	Yes (single eqn)	No	Consistent, less efficient
2SLS (system)	Yes (all eqns)	No	Consistent, more efficient than OLS
3SLS	Yes (all eqns)	Yes	Most efficient if errors correlated

4 Practice Questions

1. Explain why P_t is endogenous in Klein's Model I. What would happen if you ignored this?
2. Describe the economic meaning of each equation in Klein's Model I.
3. What is the difference between IV/2SLS and 3SLS? When is 3SLS preferred?
4. Using the R code, estimate the system and interpret the coefficients on P_t and P_{t-1} in the consumption and investment equations.
5. How would you test the validity of instruments in this context?
6. Suppose you have access to additional exogenous variables. How could you improve the identification and efficiency of the system?

5 Assignment: Simultaneous Equation and IV Estimation Using Agricultural Economics Data

Data Description

For this assignment, you will use the **Fertilizer Data** on Indian grain production, available in the `AER` package in R as `Fertilizer`. The data are from a study of input use and grain output in Indian villages.

Variables and Definitions

- `output`: Value of grain output (Rupees)
- `fertilizer`: Value of fertilizer used (Rupees)
- `labour`: Value of hired labor used (Rupees)
- `land`: Value of land used (Rupees)
- `seed`: Value of seed used (Rupees)
- `manure`: Value of manure used (Rupees)
- `area`: Area of the plot (acres)
- `village`: Village identifier

Suggested Model and Endogeneity

Suppose you are interested in estimating the effect of fertilizer use on grain output. However, fertilizer application (`fertilizer`) may be **endogenous**: unobserved factors (such as soil quality or farmer ability) may influence both fertilizer use and output, leading to biased OLS estimates.

A basic simultaneous equations system might be:

$$\begin{aligned} \texttt{output}_i &= \alpha_0 + \alpha_1 \texttt{fertilizer}_i + \alpha_2 \texttt{labour}_i + \alpha_3 \texttt{seed}_i + \alpha_4 \texttt{land}_i + u_{1i} \\ \texttt{fertilizer}_i &= \beta_0 + \beta_1 \texttt{output}_i + \beta_2 \texttt{manure}_i + \beta_3 \texttt{area}_i + v_{2i} \end{aligned}$$

where:

- `output` and `fertilizer` are endogenous in the system,
- `labour`, `seed`, `land`, `manure`, and `area` are assumed exogenous.

Assignment Questions

1. **Data Import and Exploration**
   (a) Load the data in R: `data("Fertilizer", package = "AER")`.
   (b) Provide summary statistics and pairwise scatterplots for the main variables.
2. **OLS Estimation and Endogeneity**
   (a) Estimate the grain output equation (`output` as dependent variable) by OLS.
   (b) Discuss why OLS may be biased in this context.
3. **Instrumental Variables (IV) / 2SLS**
   (a) Propose valid instruments for `fertilizer` (e.g., `manure`, `area`) and justify your choice.
   (b) Estimate the output equation by 2SLS using your chosen instruments.

(c) Compare the OLS and 2SLS results for the effect of fertilizer on output.

4. **Simultaneous Equation Estimation (System Estimation)**

   (a) Estimate the two-equation system by 2SLS (systemfit) and by 3SLS.

   (b) Interpret the estimated coefficients on `fertilizer` and `output` in both equations.

   (c) Test whether the errors in the two equations are correlated (justify use of 3SLS).

5. **Robustness and Policy Implications**

   (a) Discuss the robustness of your findings to the choice of instruments.

   (b) What are the policy implications regarding fertilizer use and grain productivity?

Hints and Resources

- Use the `ivreg()` function from the `AER` package for single-equation IV/2SLS estimation.
- Use the `systemfit()` function from the `systemfit` package for 2SLS and 3SLS system estimation.
- For scatterplots, use `pairs()` or `GGally::ggpairs()`.
- For correlation of errors, use `cor(residuals(...))`.

Deliverables

- R code (well-commented)
- Answers to all assignment questions, including interpretation of results and policy discussion
- Tables and figures as appropriate

References

- Gujarati, D. N., & Porter, D. C. (2010). *Basic Econometrics* (5th ed.).
- R package `AER`: https://cran.r-project.org/web/packages/AER/AER.pdf
- R package `systemfit`: https://cran.r-project.org/web/packages/systemfit/systemfit.pdf
- Fertilizer data: https://cran.r-project.org/web/packages/AER/AER.pdf

References

- Gujarati, D. N., & Porter, D. C. (2010). *Basic Econometrics* (5th ed.). McGraw-Hill.

- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning.

- R package `AER`: https://cran.r-project.org/web/packages/AER/AER.pdf

- R package `systemfit`: https://cran.r-project.org/web/packages/systemfit/systemfit.pdf