

Bahan Kuliah Ekonometrika I

Regresi Variabel Instrumen

Imam Awaluddin

JURUSAN EKONOMI PEMBANGUNAN
FAKULTAS EKONOMI DAN BISNIS
UNIVERSITAS LAMPUNG
2024

Pendahuluan

- **Asumsi Regresi Klasik:** Peubah bebas (eksogen) dan galat saling bebas, atau tidak ada korelasi antara variabel bebas X_i dengan variabel gangguan e_i dan dapat dituliskan $E(e_i | X_i) = 0$.
- Ketika variabel bebas X_i berkorelasi dengan variabel gangguan e_i , maka variabel bebas X_i merupakan variabel endogen (endogenous), yaitu variabel yang ditentukan di dalam model.
- Jika asumsi ini tidak terpenuhi, maka estimator OLS adalah estimator yang **bias** dan **tidak konsisten**.

Korelasi antara Variabel Bebas dengan Variabel Gangguan

- Asumsi bahwa tidak ada korelasi antara variabel bebas X_i dengan variabel gangguan e_i atau $E(e_i|X_i) = 0$, hanya terpenuhi jika:
 - 1) Kita memasukkan semua variabel yang relevan ke dalam model;
 - 2) Tidak terjadi kesalahan bentuk fungsional;
 - 3) Tidak ada korelasi antara variabel bebas X_i dengan variabel gangguan e_i atau $cov(X_i, e_i) = 0$.
- Jika variabel bebas X_i berkorelasi dengan variabel gangguan e_i atau $cov(X_i, e_i) \neq 0$.

Variabel Endogen

- Ketika variabel bebas X_i berkorelasi dengan variabel gangguan e_i , maka variabel bebas X_i merupakan variabel endogen (*endogenous*), yaitu variabel yang ditentukan di dalam model.
- Istilah variabel endogen berasal dari model persamaan simultan, yang pembahasannya ada di materi kuliah Ekonometrika II.
- Ada beberapa kondisi yang menghasilkan berkorelasinya variabel bebas dengan variabel gangguan:
 - 1) Kesalahan pengukuran variabel bebas;
 - 2) Tidak memasukkan variabel yang penting (*omitted variables bias*);
 - 3) Model autoregresif;
 - 4) Persamaan simultan.

Variabel Instrumen

- Jika variabel bebas X_i berkorelasi dengan variabel gangguan e_i , maka variabel bebas X_i merupakan variabel endogen (endogenous), maka estimator OLS adalah estimator yang **bias dan tidak konsisten**.
- Apapun sumber korelasi antara variabel bebas X_i dengan variabel gangguan e_i , selama kita mendapatkan variabel lain sebagai proksi X_i , kita dapat mengatasi masalah korelasi antara X_i dengan e_i sehingga dapat menghasilkan estimator OLS yang tidak bias dan konsisten.
- Variabel proksi tersebut dinamakan variabel instrumen (*instrumental variable (s)*), yang mana variabel ini tidak mempunyai pengaruh langsung terhadap variabel terikat Y .

Variabel Instrumen

- Variabel instrumental (terkadang disebut variabel "instrumen") adalah variabel ketiga, Z, digunakan untuk memperhitungkan perilaku tak terduga antar variabel. Menggunakan variabel instrumental untuk mengidentifikasi korelasi tersembunyi (tidak teramati) memungkinkan Anda untuk melihat korelasi sebenarnya antara variabel penjelas dan variabel terikat, Y.
- Variabel instrumen Z mempunyai dua karakteristik:
 - 1) Relevan, yaitu variabel Z berkorelasi kuat terhadap variabel endogen;
 - 2) Eksogen, yaitu variabel Z tidak berkorelasi dengan variabel gangguan e .

Metode Variabel Instrumen

- Metode instrumental variabel (IV) adalah salah satu teknik dalam analisis kausalitas yang digunakan untuk mengatasi masalah variabel endogen yang biasa terjadi dalam penelitian sosial dan ekonomi. Metode ini memungkinkan peneliti untuk mengevaluasi hubungan sebab-akibat antara variabel yang sulit atau tidak mungkin diamati langsung.

Metode Variabel Instrumen

- Metode instrumental variabel menjadi penting dalam analisis kausalitas karena seringkali kita tidak dapat melakukan eksperimen atau mengamati secara langsung variabel yang menjadi fokus penelitian. Misalnya, jika kita ingin mengetahui pengaruh pendidikan terhadap pendapatan seseorang, sulit untuk mengamati langsung pengaruh pendidikan tanpa adanya faktor-faktor lain yang mempengaruhi seperti faktor genetik, latar belakang keluarga, atau faktor ekonomi.

Konsep Dasar Metode Instrumental Variabel

- Sebelum membahas cara menggunakan metode instrumental variabel, kita perlu memahami konsep dasarnya. Dalam analisis kausalitas, terdapat dua jenis variabel: variabel endogen dan eksogen. Variabel endogen adalah variabel yang menjadi fokus penelitian dan dipengaruhi oleh faktor lain dalam sistem yang sedang diamati. Sementara itu, variabel eksogen adalah variabel yang tidak dipengaruhi oleh variabel lain dalam sistem.

Konsep Dasar Metode Instrumental Variabel

- Dalam konteks metode instrumental variabel, variabel instrumental digunakan untuk mengatasi masalah endogenitas. Variabel instrumental harus memenuhi dua syarat: pertama, harus berkorelasi dengan variabel endogen; dan kedua, tidak boleh berkorelasi dengan faktor-faktor lain yang mempengaruhi variabel terikat secara langsung.

Langkah-Langkah Penggunaan Metode Instrumental Variabel

- Untuk menggunakan metode instrumental variabel, kita perlu mengikuti beberapa langkah. Pertama, identifikasi variabel endogen, eksogen, dan instrumental. Variabel endogen adalah variabel yang ingin kita kaji pengaruhnya, variabel eksogen adalah variabel yang mempengaruhi variabel endogen, dan variabel instrumental adalah variabel yang mempengaruhi variabel endogen melalui variabel instrumental.

Langkah-Langkah Penggunaan Metode IV

- Setelah identifikasi, langkah berikutnya adalah memeriksa validitas variabel instrumental. Variabel instrumental harus memenuhi syarat korelasi dengan variabel endogen tanpa berkorelasi dengan faktor-faktor lain yang mempengaruhi variabel terikat.
- Terakhir, kita dapat melakukan estimasi model kausal dengan menggunakan metode variabel instrumental. Metode ini melibatkan regresi variabel endogen terhadap variabel instrumental dan variabel eksogen. Hasil estimasi ini memberikan perkiraan efek sebab-akibat antara variabel endogen dan variabel instrumental.

Keuntungan Metode Instrumental Variabel

- Metode instrumental variabel memiliki beberapa keuntungan.
- Pertama, metode ini memungkinkan kita untuk mengestimasi pengaruh sebab-akibat antara variabel endogen dan variabel instrumental.
- Kedua, metode ini dapat mengatasi masalah endogenitas yang umum terjadi dalam penelitian sosial dan ekonomi.
- Ketiga, metode instrumental variabel dapat memberikan bukti kausalitas yang lebih kuat dibandingkan dengan metode korelasional biasa.

Kerugian Metode Instrumental Variabel

- Metode instrumental variabel juga memiliki kerugian.
- Pertama, validitas variabel instrumental menjadi kunci keberhasilan metode ini. Jika variabel instrumental tidak memenuhi syarat, estimasi yang dihasilkan dapat menjadi bias atau tidak valid.
- Kedua, penggunaan metode instrumental variabel dapat memperkenalkan kompleksitas tambahan dalam analisis data.

Contoh

- Studi tentang hubungan antara pendidikan dan pendapatan. Kita ingin mengetahui apakah pendidikan memiliki pengaruh positif terhadap pendapatan individu. Namun, terdapat banyak faktor lain yang mempengaruhi pendapatan seperti faktor genetik, latar belakang keluarga, dan faktor ekonomi. Dalam hal ini, kita dapat menggunakan variabel instrumental seperti jumlah tahun sekolah orang tua sebagai pengganti pendidikan individu.

Contoh

- Contoh lainnya adalah studi tentang pengaruh iklan terhadap penjualan produk. Kita ingin mengetahui apakah iklan memiliki dampak positif terhadap penjualan. Namun, terdapat faktor-faktor lain yang mempengaruhi penjualan seperti reputasi merek, harga, dan preferensi konsumen. Dalam hal ini, kita dapat menggunakan variabel instrumental seperti jumlah iklan yang ditayangkan sebagai pengganti iklan individu.

Tahapan dalam Mengestimasi Model Regresi dengan Metode IV

- **Tahap 1: Memilih Variabel Instrumen yang Sesuai**
- **Tahap 2: Spesifikasi Model Regresi**
- **Tahap 3: Estimasi Model Regresi**

Tahap 1: Memilih Variabel Instrumen yang Sesuai

- Langkah pertama dalam mengestimasi model regresi dengan metode IV adalah memilih variabel instrumen yang sesuai. Variabel instrumen harus memenuhi syarat berkorelasi dengan variabel endogen dan tidak berkorelasi dengan residual. Proses pemilihan variabel instrumen dapat dilakukan dengan menggunakan pengetahuan domain, uji statistik, atau metode lainnya seperti uji Hausman.

Tahap 2: Spesifikasi Model Regresi

- Setelah variabel instrumen dipilih, tahap selanjutnya adalah menentukan spesifikasi model regresi. Model regresi harus mencakup variabel endogen, variabel instrumen, serta variabel kontrol jika ada. Penting untuk mempertimbangkan relevansi variabel instrumen dengan variabel endogen dan menghindari overfitting model.

Tahap 3: Estimasi Model Regresi

- Setelah spesifikasi model regresi ditentukan, tahap berikutnya adalah melakukan estimasi model dengan metode IV. Estimasi ini dilakukan dengan menggunakan teknik seperti Two-Stage Least Squares (2SLS) atau Generalized Method of Moments (GMM). Metode ini memperhitungkan korelasi antara variabel instrumen dengan variabel endogen untuk mendapatkan estimasi yang konsisten dan efisien.

Estimasi Variabel Instrumen

- Ketika variabel bebas X_i berkorelasi dengan variabel gangguan e_i , maka variabel bebas X_i merupakan variabel endogen (endogenous), yaitu variabel yang ditentukan di dalam model.
- Jika asumsi ini tidak terpenuhi, maka estimator OLS adalah estimator yang **bias** dan **tidak konsisten**.
- Salah satu metode estimasi untuk mengatasi hal ini adalah menggunakan metode two stage least squares (TSLS).
- Metode TSLS akan menghasilkan estimator yang konsisten

Estimasi Variabel Instrumen

- Misalkan kita punya model: $Y_i = \alpha_0 + \alpha_1 X_i + e_i$
- Kita asumsikan terjadi korelasi antara X_i dan e_i .
- Misalkan kita punya variabel instrument Z , yang mana memenuhi kondisi $cov(Z_i, X_i) \neq 0$, dan $cov(Z_i, e_i) = 0$.
- Langkah estimasi dengan TSLS adalah:

➤ Regresikan antara X dengan Z sehingga diperoleh $\hat{X}_i$

$$X_i = \delta_0 + \delta_1 Z_i + \varepsilon_i$$

➤ Tahap selanjutnya adalah meregresikan antara Y_i dengan $\hat{X}_i$

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + u_i$$

Regresi Variabel Instrumen (IV)

- Lihat Buku Agus 12.3
- Langkah Regresi Model Variabel Instrumen:
 - 1) Uji Endogenitas Variabel Bebas (contoh 12.1)
 - 2) Estimasi Model awal dengan OLS (contoh 12.2, Tabel 12.2)
 - 3) Estimasi reduce form (Tabel 12.3)
 - 4) Regresi Model Variabel Instrumen dengan TSLS (Tabel 12.4)

Endogenitas Variabel Bebas

- Misalkan kita ingin menganalisis faktor yang mempengaruhi Tingkat upah pekerja dengan model:

$$\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + e_i$$

- Dimana Y adalah upah (\$), X1 adalah Tingkat Pendidikan (lama sekolah dalam tahun), X2 adalah pengalaman kerja (tahun), dan X4 adalah umur pekerja (tahun).
- Dalam model tersebut ada variabel penting yang belum dimasukkan yaitu kemampuan pekerja. Variabel ini sangat berhubungan dengan lamanya sekolah sehingga ada korelasi antara variabel Tingkat Pendidikan dengan variabel gangguan
→ variabel Tingkat Pendidikan adalah variabel endogen.

Uji Endogenitas Variabel Bebas

- Misalkan kita punya variabel Pendidikan ibu (Z1) dan pendidikan ayah (Z2) yang mempengaruhi variabel Tingkat Pendidikan pekerja tetapi tidak berkorelasi dengan variabel gangguan → variabel instrumen
- Untuk menentukan apakah variabel Tingkat Pendidikan (X1) adalah variabel endogen, dengan menggunakan uji Hausman, tahapan pengujiannya adalah:
- Regresikan variabel X1 dengan variabel-variabel instrument dan variabel bebas lainnya

$$X_{1i} = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \alpha_3 X_{2i} + \alpha_4 X_{3i} + \alpha_5 X_{4i} + \varepsilon_i$$

Tabel 12.3

Dependent Variable: X1				
Method: Least Squares				
Date: 05/31/24 Time: 04:40				
Sample: 1 428				
Included observations: 428				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X2	0.045128	0.040292	1.120033	0.2633
X3	-0.001091	0.001222	-0.892805	0.3725
X4	0.006030	0.015149	0.398029	0.6908
Z1	0.160445	0.036635	4.379512	0.0000
Z2	0.188637	0.033868	5.569819	0.0000
C	8.851148	0.762590	11.60670	0.0000
R-squared	0.211767	Mean dependent var	12.65888	
Adjusted R-squared	0.202427	S.D. dependent var	2.285376	
S.E. of regression	2.040999	Akaike info criterion	4.278675	
Sum squared resid	1757.915	Schwarz criterion	4.335579	
Log likelihood	-909.6365	Hannan-Quinn criter.	4.301149	
F-statistic	22.67488	Durbin-Watson stat	1.939878	
Prob(F-statistic)	0.000000			

Uji Endogenitas Variabel Bebas

- Hasil estimasi persamaan tersebut ambil residunya $\hat{\varepsilon}$, dan masukkan ke persamaan awal:

$$\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 \hat{\varepsilon}_i + e_i$$

- Jika residu $\hat{\varepsilon}$ signifikan, maka dapat disimpulkan bahwa ada korelasi antara Tingkat Pendidikan (X1) dengan variabel gangguan (e_i) $\rightarrow$ X1 adalah variabel endogen

Dependent Variable: LOG(Y)

Method: Least Squares

Date: 05/31/24 Time: 07:32

Sample: 1 428

Included observations: 428

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.066718...	0.4524010...	0.1474767...	0.88282...
X1	0.060994...	0.0311182...	1.9600862...	0.05064...
X2	0.044195...	0.0132558...	3.3340802...	0.00093...
X3	-0.00089...	0.0004015...	-2.228289...	0.02638...
X4	-0.00035...	0.0048597...	-0.072891...	0.94192...
RESID01	0.058643...	0.0349356...	1.6786256...	0.09396...
R-squared	0.162419...	Mean dependent	11.190173298849095	
Adjusted R-squared	0.152495...	S.D. dependent	0.7231978173729128	
S.E. of regression	0.665775...	Akaike info crit	2.1038191918723007	
Sum squared resid	187.0546...	Schwarz criterion	2.095095514922751	
Log likelihood	-430.173...	Hannan-Quinn	2.060665706641193	
F-statistic	16.36648...	Durbin-Watson	1.932180392551549	
Prob(F-statistic)	9.055256...			

Estimasi Model Awal

Dependent Variable: LOG(Y)

Method: Least Squares

Date: 05/31/24 Time: 07:37

Sample: 1 428

Included observations: 428

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.53337...	0.2778430...	-1.919702...	0.05556...
X1	0.107522...	0.0141744...	7.5856540...	2.12300...
X2	0.041562...	0.0131909...	3.1508332...	0.00174...
X3	-0.00081...	0.0003995...	-2.040165...	0.04195...
X4	0.000283...	0.0048552...	0.0584204...	0.95344...
R-squared	0.156827...	Mean dependent	11.190173298849095	
Adjusted R-squared	0.148853...	S.D. dependent	0.7231978173729128	
S.E. of regression	0.667204...	Akaike info crit	2.040174041364418	
Sum squared resid	188.3036...	Schwarz criter	2.087593704864206	
Log likelihood	-431.597...	Hannan-Quinn	2.058902197962906	
F-statistic	19.66913...	Durbin-Watson	1.960484046947316	
Prob(F-statistic)	7.328088...			

Estimasi Reduced Form

Dependent Variable: X1 Method: Least Squares Date: 05/31/24 Time: 04:40 Sample: 1 428 Included observations: 428				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X2	0.045128	0.040292	1.120033	0.2633
X3	-0.001091	0.001222	-0.892805	0.3725
X4	0.006030	0.015149	0.398029	0.6908
Z1	0.160445	0.036635	4.379512	0.0000
Z2	0.188637	0.033868	5.569819	0.0000
C	8.851148	0.762590	11.60670	0.0000
R-squared	0.211767	Mean dependent var	12.65888	
Adjusted R-squared	0.202427	S.D. dependent var	2.285376	
S.E. of regression	2.040999	Akaike info criterion	4.278675	
Sum squared resid	1757.915	Schwarz criterion	4.335579	
Log likelihood	-909.6365	Hannan-Quinn criter.	4.301149	
F-statistic	22.67488	Durbin-Watson stat	1.939878	
Prob(F-statistic)	0.000000			

- Variabel Z1 dan Z2 signifikan berpengaruh terhadap X1 → memenuhi syarat sebagai variabel instrumen

Estimasi TSLS

Dependent Variable: LOG(Y)

Method: Two-Stage Least Squares

Date: 05/31/24 Time: 07:41

Sample: 1 428

Included observations: 428

Instrument specification: X2 X3 X4 Z1 Z2

Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.066718...	0.4591101...	0.1453216...	0.88452...
X1	0.060994...	0.0315797...	1.9314431...	0.05409...
X2	0.044195...	0.0134524...	3.2853586...	0.00110...
X3	-0.00089...	0.0004074...	-2.195726...	0.02865...
X4	-0.00035...	0.0049318...	-0.071826...	0.94277...
R-squared	0.135349...	Mean dependent	11.190173298849095	
Adjusted R-squared	0.127172...	S.D. dependent	0.7231978173729128	
S.E. of regression	0.675649...	Sum squared res	196.1002632318843	
F-statistic	6.084966...	Durbin-Watson	1.946128398714296	
Prob(F-statistic)	9.084780...	Second-Stage	2.522162481535394	
J-statistic	0.397401...	Instrument rank		6
Prob(J-statistic)	0.528434...			

Tugas

- Tugas akan diberikan di kelas