Two-stage regression without exclusion restrictions

KVREG: Implementation of Klein and Vella, 2010

Michael Barker

Department of Economics
Georgetown University

Stata Conference, 2013

Outline

Outline

Introduction

- Klein, R. and Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, 154(2).
- Stata command: **kvreg**
- Identification uses heteroscedasticity
- Focus on intuition of the estimator and required assumptions

Outline

Related Work

- Lewbel, A. (2012). Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. *Journal of Business & Economic Statistics*, 30(1).
- Stata command: **ivreg2h** by Christopher Baum and Mark Schaffer
- Requires stricter restrictions on the error terms
- Computationally less expensive

Outline

Triangular Simultaneous Equations

The Klein and Vella 2010 estimator is designed to estimate a model of the following form:

$$Y_1 = Y_2\theta + X\beta + u_1 \quad (1)$$

$$Y_2 = X\pi + v_2 \quad (2)$$

- Y_1 and Y_2 are continuous
- $E(u_1 \mid X) = 0$ and $E(v_2 \mid X) = 0$
- $corr(u_1, v_2) \neq 0$

Identification Strategy

- Most common approach is instrumental variables
- IV requires exclusion restriction, which are frequently difficult to justify or non-existent
- Instead, KV use a control function approach
- A non-linear control function is constructed by modeling heteroscedasticity in the error terms

Outline

Control Function Theory

$$Y_1 = Y_2\theta + X\beta + u_1 \tag{1}$$
$$Y_2 = X\pi + v_2 \tag{2}$$

- Problem: $E[u_1 \mid X, Y_2] \neq 0$ because $corr(u_1, v_2) \neq 0$
- Write $u_1 = E[u_1 \mid X, Y_2] + e$, where $E[e|X, Y_2] = 0$
- Include an estimate of $E[u_1 \mid X, Y_2]$ in the Y_1 equation
- Leaving:

$$Y_1 = Y_2\theta + X\beta + E[\widehat{u_1 \mid X,} Y_2] + e$$

Control Function Implementation: 2SLS

2SLS estimation of a linear control function:

$$\hat{v}_2 = Y_2 - X\hat{\pi} \tag{3}$$
$$Y_1 = Y_2\theta + X\beta + \rho\hat{v}_2 + e \tag{4}$$

Giving us the control function, $\rho\hat{v}_2$, where:

- $u_1 = \rho\hat{v}_2 + e$
- $E(e \mid X, Y_2) = 0$

Without an exclusion restriction, eq. 4 cannot be estimated, as $\hat{v}_2$ is a linear function of the other regressors, (Y_2, X).

Non-linear Control Function

Introduce a control function that is a non-linear function of X:

$$Y_1 = Y_2\theta + X\beta + cf(X) + e \qquad (5)$$

- Identification follows from non-linearity
- Exclusion restrictions are not required
- Non-linear control functions are used in sample selection models
  - e.g. Inverse Mills Ratio: $\lambda = \frac{\phi(x\beta)}{\Phi(x\beta)}$
  - requires strong distributional assumptions
- KV construct a non-linear control function that can be estimated with minimal distributional assumptions

Characterize the Error Structure

The error terms should exhibit multiplicative heteroscedasticity:

$$u_1 = S_{u1}(X)\, u_1^* \quad (6)$$

$$v_2 = S_{v2}(X)\, v_2^* \quad (7)$$

- u_1^* is the homoscedastic component
- $S_{u1}(X)$ is a scaling function
- The conditional variance functions follow:
  - $Var(u_1|X) = S_{u1}^2$
  - $Var(v_2|X) = S_{v2}^2$

Restrictions

For $u_1 = S_{u1} u_1^*$ and $v_2 = S_{v2} v_2^*$,

Consistent estimation of $cf(X)$ requires:

$$E(u_1^*|X) = 0 \tag{8}$$

$$E(v_2^*|X) = 0 \tag{9}$$

$$\frac{S_{u1}}{S_{v2}} \neq C \tag{10}$$

$$E(u_1^* v_2^*|X) = E(u_1^* v_2^*) = \rho \tag{11}$$

Interpretation of Error Restrictions

$$E(u_1^* v_2^* | X) = E(u_1^* v_2^*) \equiv \rho \tag{11}$$

- Homoscedastic components must be linearly related
- Cannot be tested - it must be justified with contextual argument
- Linearity assumptions are prevalent in regression analysis
- Specific interpretation of the restriction and coefficient is application specific

Application: Wages and Education

Modeling the impact of Years of Education on Wages:

$$wage = educ * \theta + X\beta + u_1$$
$$educ = X\pi + v_2$$

With an additive linear error structure:

$$u_1^* = \rho v_2^* + \varepsilon^* \qquad cov\left(v_2^*, \varepsilon^* | X\right) = 0$$

Interpretation

$$u_1^* = \rho v_2^* + \varepsilon^*$$

- v_2^* is unobserved scholastic ability
- ρ measures the impact of unobserved ability on wages
- Returns to unobserved scholastic ability are constant and do not depend on other individual characteristics
- Violated if wages rise exponentially with unobserved ability
- Violated if the relationship grows weaker with age

Definition of $cf(X)$

Given an error structure satisfying the restrictions above, the control function is defined as:

$$cf(X) = \rho \frac{S_{u1}}{S_{v2}} v_2 \tag{12}$$

Adding the control function into the Y_1 equation gives:

$$Y_1 = Y_2\theta + X\beta + \rho \frac{S_{u1}}{S_{v2}} v_2 + e \tag{13}$$

Outline

The Role of SLS in KVREG

- Semiparametric least squares is used to estimate $S_{v2}^2 = Var\left(v_2|X\right) = E\left[v_2^2|X\right]$
- Semiparametric estimation is not required for identification
- Any consistent estimate of $S_{v2}^2 = E\left[v_2^2|X\right]$ will work:

Parametric	$S_{v2}^2 = exp(X\delta)$
Nonparametric	$S_{v2}^2 = g(X)$
Semiparametric	$S_{v2}^2 = g(X\delta)$

- Semiparametric estimation is computationally feasible with minimal distributional assumptions
- **`sls`** will be available through SSC as a stand-alone estimation command

Estimation Procedure

$$Y_1 = Y_2\theta + X\beta + \rho\frac{S_{u1}(X\gamma)}{S_{v2}(X\delta)}v_2 + e$$

. **kvreg** Y_1 Y_2 X

Estimation: Step 1

$$Y_1 = Y_2\theta + X\beta + \rho \frac{S_{u1}(X\gamma)}{S_{v2}(X\delta)} v_2 + e$$

1. Estimate v_2, secondary equation residual

```
. reg Y2 X
. predict v2, residual
```

Estimation: Step 2

$$Y_1 = Y_2\theta + X\beta + \rho \frac{S_{u1}(X\gamma)}{S_{v2}(X\delta)} v_2 + e$$

2. Estimate $S_{v2}(X\delta)$, the sq root of $S_{v2}^2 = Var(v_2|X) = E\left[v_2^2|X\right]$

```
. gen v2^2 = v2^2
. sls v2^2 X
. predict S2v2, yhat
. gen Sv2 = sqrt(S2v2)
```

Estimation: Step 3

$$Y_1 = Y_2\theta + X\beta + \rho\frac{S_{u1}(X\gamma)}{S_{v2}(X\delta)}v_2 + e$$

3. Estimate remaining parameters in a single minimization problem using moptimize

$$\min_{\{\theta,\beta,\rho,\gamma\}} \sum_{i=1}^{N}\left[Y_{1i} - \left(Y_{2i}\theta + X_i\beta + \rho\frac{S_{ui}(X\gamma)}{S_{vi}(X\delta)}v_i\right)\right]^2$$

Estimation: Step 3a

$$Y_1 = Y_2\theta + X\beta + \rho \frac{S_{u1}(X\gamma)}{S_{v2}(X\delta)} v_2 + e$$

3a. Estimate $S_{u1}(X\gamma)$ within each iteration of the minimization procedure.

At the current parameter estimates, $\left(\tilde{\theta}, \tilde{\beta}, \tilde{\gamma}\right)$:

. **gen** u_1 **=** $Y_1 - \left(Y_2 * \tilde{\theta} + X * \tilde{\beta}\right)$

. **gen** u_1^2 **=** u_1^2**^2**

. **gen** S_{u1}^2 **=** $E\left[u_1^2 | X * \tilde{\gamma}\right]$

. **gen** S_{u1} **=** **sqrt(**S_{u1}^2**)**

Outline

Simulation Coefficient Estimates

Table: Coefficient Estimates

Variable	True	OLS	CF
Y_2	1	1.29	.98
		(.050)	(.200)
x_1	1	.72	1.02
		(.070)	(.205)
x_2	1	.72	1.02
		(.091)	(.205)
Cons	1	.72	1.03
		(.071)	(.203)
ρ	0.33		.314
			(.181)
N = 1000	R = 100	Ave. Time =	839 sec

Computational Requirements

- Optimization is computationally expensive
- Average time is 14 min. for N=1,000
  - 64-bit Linux
  - 8 GB ram
  - Intel Xeon CPU @ 2.70GHz
- Objective function is not convex, so direct search is required
  - Nelder-Mead algorithm

Outline

Variance Estimates

- Current variance estimates are roughly 4-times larger than variance of simulated coefficients
- Variance inflation related to semiparametric component
  - derivative of kernel density estimator
  - bandwidth choice for derivative estimates
- Final component before posting **`kvreg`** and **`sls`** to SSC archive

NP System

- Nonparametric estimation of $E[Y|X]$ is central to many semi and nonparametric estimators
- **kvreg** and **sls** share components for nonparametric conditional expectation
- Components should be formalized in a system like Stata's **optimize** and **moptimize**
- see Hayfield and Racine's np package in R

Summary

- Identification is through non-linearity
- Semiparametric estimation is one way to estimate conditional variance functions and requires minimal distributional assumptions
- Watch for **`kvreg`** and **`sls`** in the SSC archive
- For current code, email: mdb96@georgetown.edu

Thank You

Michael Barker

mdb96@georgetown.edu

Lewbel, 2013: Error Restrictions

Application of the Lewbel, 2013 error restrictions for identification through heteroscedasticity to the KV, 2010 model:

Lewbel, 2013		KV, 2010
$cov\left(X, v_2^2\right) \neq 0$	$\Longrightarrow$	$S_{v2} \neq C$
$cov\left(X, u_1 v_2\right) = 0$	$\Longrightarrow$	$\rho = corr(u^*, v^*) = 0$

Lewbel, 2013: Correlated Error Structure

The Lewbel estimator can be applied to correlated error structures if each error term can be written as additive components as follows:

$$u_1 = A_1 + B_1$$
$$v_2 = A_2 + B_2$$

where:

$$A_1,\ A_2 \text{ are correlated and homoscedastic}$$
$$B_1,\ B_2 \text{ are uncorrelated and heteroscedastic}$$

This type of error structure can be applied to unobserved single factor models.

Semiparametric Least Squares

Estimation of S_v^2 via Semiparametric Least Squares

Model:

$$v^2 = g\left(X\delta\right) + \varepsilon$$

Estimation:

$$\min_{\{\delta\}} \sum_{i=1}^{N} \left(v_i^2 - \widehat{E}\left[v_i^2 | X_i \delta\right]\right)^2$$

For any candiate $\tilde{\delta}$:

$$\widehat{E}\left[v_i^2 | X_i \tilde{\delta}\right] = \frac{\sum_{j \neq i} v_j^2 * K\left(\frac{X_i\tilde{\delta} - X_j\tilde{\delta}}{h}\right)}{\sum_{j \neq i} K\left(\frac{X_i\tilde{\delta} - X_j\tilde{\delta}}{h}\right)}$$

Simulation Model

Error Terms:

$$v_2^*,\ \varepsilon^* \sim N(0,1)$$
$$u_1^* = .33 * v_2^* + \varepsilon^*$$
$$u_1 = \exp(.2 * x_1 + .6 * x_2) * u_1^*$$
$$v_2 = \exp(.6 * x_1 + .2 * x_2) * v_2^*$$

Define Model:

$$x_1 \sim N(0,1)$$
$$x_2 \sim \chi^2(1)$$
$$Y_1 = 1 + x_1 + x_2 + Y_2 + u_1$$
$$Y_2 = 1 + x_1 + x_2 + v_2$$

Simulation Variance Estimates

Table: Variance Estimates

Variable	Simulation	Estimated	Trimmed (10-90)
Cons	.041	.185	.046
		(.566)	(.070)
Y_2	.040	.185	.044
		(.576)	(.071)
x_1	.042	.189	.046
		(.588)	(.069)
x_2	.042	.183	.047
		(.551)	(.071)
ρ	.027	.205	.054
		(.646)	(.080)
N = 1000	R = 100	Ave. Time =	839 sec