Systems of Equations.

Paul Johnson <pauljohn@ku.edu>

Nov 21, 2006

1 Terminology

exogenous determined "outside" the system under consideration, as we usually consider most "independent variables"

endogenous determined "within" the system, as in "dependent variable"

predetermined in a time series model, a lagged endogenous variable is treated as if it were exogenous

2 Consider two regressions

$$\begin{aligned} Y1_i &= b_0 + b_1 X1_i + e1_i \\ Y2_i &= c_0 + c_1 X1_i + e2_i \end{aligned} \tag{1}$$

You can run OLS on those equations, either
separately
or
after "stacking" them into one data frame and then using "dummy" variables to estimate $b$and c (and possibly adjusting for heteroskedasticity). Let D represent the dichotomous variable, with 1 representing the fact that the observation concerns $Y2_i$(and 0 otherwise). The combined regression would be like this:

$$\begin{aligned} Y1_1 &= b_0 + b_1 X1_1 + e1_1 \\ Y1_2 &= b_0 + b_1 X1_2 + e1_i \\ &\vdots \quad\quad \vdots \\ Y1_N &= b_0 + b_1 X1_N + e1_N \\ Y2_1 &= b_0 + g_0 * D_i + b_1 X1_i + g_1 (X1_1 * D) + e2_1 \\ &\vdots \quad\quad \vdots \\ Y2_N &= b_0 + g_o * D_i + b_1 X1_i + g_1 (X1_N * D) + e2_N \end{aligned} \tag{2}$$

Here, g_0 and g_1 are "intercept shifter" and "slope shifter", respectively. The estimate of the coefficient $c_1$is found by adding $b_1$and g_1. One can easily employ tests for heteroskedasticity to find out if the error term has the same variance in the two dependent variables.

3 Seemingly unrelated regressions

Recall the problem of autocorrelation? That refers to correlation in the error terms between units. If there is autocorrelation, the parameter estimates are biased and the standard errors are wrong. Autocorrelation has usually been thought of as a time series problem, since exogenous shocks persist over time.

However, if there are shocks that cause a correlation between $e1$ and $e2$, then there are correlated errors. This phenomenon is known as "Seemingly unrelated regression" (or SUR) and it was brought to the forefront by U. of Chicago professor Edward Zellner.

4 Endogenous predictors

4.1 Cascade or "blocked" equations

$$\begin{aligned} Y1_i &= b_0 + b_1 X1_i + e1_i \\ Y2_i &= c_0 + c_1 Y1_i + c_2 X1_i + e2_i \end{aligned} \tag{3}$$

$X1$ goes into $Y1$, and then $Y1$goes into $Y2$.
As long as $e1$ and $e2$ are uncorrelated, then OLS can be used to estimate this.

4.2 Simultaneous equations: OLS is not suitable

$$\begin{aligned} Y1_i &= b_0 + b_1 Y2_i + b_2 X1_i + e1_i \\ Y2_i &= c_0 + c_1 Y1 + c_2 X1_i + e2_i \end{aligned} \tag{4}$$

This one strains the brain. $Y1$ is affected by $Y2$ and $e1$ simultaneously and $Y2$ is affected by $Y1$ and $e2$ at the same time.

4.2.1 Correlation between predictor and error

Recall the fundamental assumption in regression that the error term is not correlated with the independent variables. This is violated in equation system 4 because the endogenous predictors are related to the error terms.

Consider this:
$Y2_i$ has to be correlated with $e1_i$.
Because $e1$ is directly "going into" $Y1$, and $Y1$ "goes into" $Y2$, and $Y2$ "goes back into" $Y1$, then some part of $e1$ has to be hidden inside $Y2$.

4.2.2 Consider the linkage between the error terms

Think of $Y1$ and $Y2$ as observed "known values." The theoretical "exogenous shocks" $e1$ and $e2$ are algebraically constrained by the values of $Y1$ and $Y2$ that are observed. Consider the coefficients b and c as "known" (or conjectured) values. Given values for the observed values $Y1_i$ and $Y2_i$ and $X1$, then $e1_i$ and $e2_i$ are mathematically restricted because they have to be added in to produce the correct observed values of $Y1_i$ and $Y2_i$. If $e1_i$ is increased by a certain amount, then it is algebraically necessary that $e2_i$ be adjusted so that the observed values stay at $Y1_i$ and $Y2_i$.

If you accept this argument that $e1$ and $e2$ constrained by each other, then you should easily see that there will be correlation between the endogenous predictors and the error terms.

5 Two Stage Least Squares

This is the simplest "fix" for simultaneous equations. It is a "single equation" or "limited information" approach. As you will see, we create separate estimates of the equations for $Y1$ and $Y2$.

Stage 1: calculate a "correlation purged" estimate for each endogenous predictor. Use any exogenous variables $Z1$, $Z2$ (and so forth) on the right hand side

$$\breve{Y1}_i = \hat{\pi}_0 + \hat{\pi}_1 Z1_i + \hat{\pi}_2 Z2_i$$

Confusion: In the literature, the term "instrumental variable" is used for both Z_i and for $\breve{Y}_i$.
Stage 2: use the new estimate $\breve{Y}1$ in place of $Y1$ in the regression

$$Y2_i = c_0 + c_1 \breve{Y}1_i + c_2 X1_i + e1_i$$

The estimates obtained in this second stage are

- consistent
- efficient

but not unbiased.

In stage 1, it is customary to use ALL EXOGENOUS predictors to predict ALL ENDOGENOUS variables.

reduced form equations refers to the system equations in which the endogenous variables are arranged on the left hand side and the exogenous predictors are on the right hand side:

$$\begin{aligned} Y1_i &= \pi_{10} + \pi_{11}X1_i + \pi_{12}X2_i + \zeta 1_i \\ Y2_i &= \pi_{20} + \pi_{21}X1_i + \pi_{22}X2_i + \zeta 1_i \end{aligned}$$

6 The identification problem

If two stage squares worked all the time, our work would be done. But it doesn't. And, somewhat to my surprise, the problem all traces back to multicollinearity.

6.1 Perfect collinearity in stage 2

Recall the multicollinearity problem. One cannot obtain reliable, separate estimates of coefficients for several parameters because the variables are redundant (intercorrelated). The redundancy of data makes estimation unreliable or impossible.

Consider equation system 4. Suppose we have only one exogenous variable, $X1$ and we use it to calculate the stage 1 estimate.

$$\breve{Y1}_i = \hat{\pi}_0 + \hat{\pi}_1 X1_i \tag{5}$$

And then that is inserted into the system in stage 2

$$Y2_i = c_0 + c_1(\hat{\pi}_0 + \hat{\pi}_1 X1_i) + c_2 X1_i + e2_i \tag{6}$$

Don't be a bonehead! Can't you see that's the definition of multicollinearity? The variable $X1$ can be used in a linear combination to reproduce the part in parentheses.

Ouch! There's perfect multicollinearity here! Essentially, $X1_i$is included twice. The coefficients $c_1$and $c_2$are not separately identified.

6.2 How do you fix that? You need more exogenous variables.

Imagine you have some more exogenous variables sitting around. Suppose, for example, you have $X1$and $X2$ in the first stage:

$$\breve{Y1}_i = \hat{\pi}_0 + \hat{\pi}_1 X1_i + \hat{\pi}_2 X2_i \tag{7}$$

$$Y2_i = c_0 + c_1(\hat{\pi}_0 + \hat{\pi}_1 X1_i + \hat{\pi}_2 X2_i) + c_2 X1_i + e2_i \tag{8}$$

Note, now the multicollinearity still exists, but it is not perfect multicollinearity. So the model can be estimated.

The "trick" is to find some new exogenous variable $X2$ that is NOT INCLUDED in the equation for $Y2$.

This works because there is an exogenous variable that is used to create the new predictor $\breve{Y1}$and that exogenous variable is not included in the model for $Y2$.

$$
\begin{aligned}
Y1_i &= b_0 + b_1 Y2_i + b_2 X1_i + b3 X2_i + e1_i \\
Y2_i &= c_0 + c_1 Y1 + c_2 X1_i + e2_i
\end{aligned} \tag{9}
$$

6.3 Order condition

This is the most common way that people use to check to see if an equation within a system can be estimated by 2SLS. Basically, it says one must omit at least as many exogenous variables from an equation as there are included endogenous variables.

In a 2 equation system, identification requires that at least one X must be excluded from each equation.

Please note, the order condition is necessary, but not sufficient for identification. So, it is mathematically wrong to assume the order condition provides all of the needed information. But people usually do, because the disjuncture between sufficiency and necessity is not large in most cases.

6.4 Rank condition

This is a matrix algebra check that can be applied to a system to find out if its parameters are identified. It gives an answer which is both necessary and sufficient.

7 2SLS with generalized linear models

In the mid 1980s, the logistic regression model had become well established. What if a logistic regression model is included in one of these systems?

This has been a "fly by the seat of the pants" problem. One must specify a theoretical model that is logically meaningful. And it is not always easy. For example, suppose $Y1$ is dichotomous, coded 0 and 1. If $Y1$ is included in the equation for $Y2$, do we mean that the input variable is really 0 or 1, or is it the probability of 1, representing the proclivity?

8 Full Information Maximum Likelihood versus Limited Information models

2SLS is a "limited information" approach because the estimation of the equations is done separately.

A "full information" approach is one that tries to estimate the parameters for all equations jointly.

Three Stage Least Squares is the most commonly used full information approach. The process begins with 2SLS, and then uses the 2SLS estimates to estimate the inter-correlation of the error terms from the different equations. Then a sort of weighted least squares approach is used in the third stage.

The danger of full information approaches is that a mistake in specification of one equation will affect the estimates from all equations.

9 Structural Equation Models

The econometric approach to these problems is to act as though the input variables are measured correctly and then are used to predict the output variables.

Psychologists, on the other hand, often suppose that they can't measure the real inputs. Rather, they have multiple indicators. So the goal is to somehow, simultaneously, use the multiple indicators to get an idea about a subject's position on an "underlying" or "latent" variable, and then find out if that score on the latent variable is affecting some other variables. The model known as LISREL was pioneered by Joreskog to dramatically expand the kinds of structures that are investigated in systems modeling.

Structural Equation Modeling is a system in which systems are posited and estimated.

Identification is a major problem in SEM work, and some software programs will provide parameter estimates for technically non-estimable problems.

Bayesian models using Gibbs Sampling offers an alternative estimation technique for these models.