Omitted Variables, Instruments and Fixed Effects

Structural Econometrics Conference

July 2013

Peter Rossi

UCLA | Anderson

Variation

Imagine that our goal is to determine the “pure” or causal effect of changing the variable x_1 on y.

What is the ideal source of variation? Exogenous variation by which we mean experimental variation. As though we conducted an experiment where we randomly changed x_1. This means that all variation is independent of any other variables influencing y.

Then we could simply regress y on x_1

$$y = \beta_1 x_1 + \varepsilon$$

Omitted variables

However, we live in an non-experimental world.

or

Our experiments are imperfect and don’t have full randomization.

There exist other variables which belong in the equation in the sense that they are correlated with x.

$$y = \beta_1 x_1 + \varepsilon_y = \beta_1 x_1 + \left(\beta_2 x_2 + \upsilon\right)$$

Omitted variables

If we regress y on x_1, what does least squares estimate?

$$
\begin{aligned}
E\left[y \mid x_1\right] &= E\left[\beta_1 x_1 + \beta_2 x_2 \delta + \upsilon \mid x_1\right] \\
&= \beta_1 x_1 + \beta_2 E\left[x_2 \mid x_1\right] \\
&= \beta_1 x_1 + \beta_2 \pi x_1
\end{aligned}
$$

$$
\frac{\partial}{\partial x_1} E\left[y \mid x_1\right] = \underbrace{\beta_1}_{\text{Direct Effect of } x_1} + \underbrace{\beta_2 \pi}_{\text{Indirect Effect of } x_1} = \delta
$$

Omitted variables

This means that there can be an asymptotic bias or inconsistency from omission of variables.

$$plim\,\hat{\delta}_{OLS(y\,on\,x_1)} = \delta$$

$$\delta - \beta_1 = \pi\beta_2$$

Implies a bias term which does not go away even in infinite samples.

Is Omitted Variable Bias A Problem?

Not necessarily.

Suppose x_1 is under firm's control but x_2 is not.

If we use our data to estimate the relationship between x_1 and x_2 then this is the same using OLS from y on x_1.

Suppose both variables are under firm's control.

It is still not clear. If x_1 is price, x_2 is promotion (like a display). If my display policy doesn't change (the frequency with which I "support" price changes by display) then I want delta not beta1!

What Can Be Done?

Bring x_2 into the model and use multiple regression.

What does multiple regression do?

Instead of using all of the variability of x_1 to estimate the coefficient, multiple regression uses only that portion uncorrelated to x_2 (regress x_1 on x_2 and use the residuals). Throwing out data!

$$y = b_1 x_1 + b_2 x_2 + e$$

$$y = b_1 e_{1.2} + v$$

$$x_1 = c_0 + c_1 x_2 + e_{1.2}$$

Omitted Variables and Endogeneity

Suppose the omitted variable is not observed by the researcher.

$$y = \beta x + \left(\alpha w + \upsilon_y\right)$$

$$x = w + \varepsilon_x$$

The omitted variable bias is now

$$\begin{aligned} E\left[y \mid x\right] &= \beta x + E\left[\alpha w + \upsilon_y \mid x\right] \\ &= \beta x + \alpha E\left[w \mid x\right] \\ &= \beta x + \alpha \left(\frac{\sigma_w^2}{\sigma_w^2 + \sigma_{\varepsilon_x}^2} \right) x \end{aligned}$$

Omitted Variables and Endogeneity

The omitted variable bias is

$$\alpha\left(\frac{\sigma_w^2}{\sigma_w^2+\sigma_{\varepsilon_x}^2}\right)$$

Another derivation:

$$y=\beta x+\upsilon$$

$$\hat{\beta}=\left(X'X\right)^{-1}X'y$$

$$plim\left(\hat{\beta}\right)=\beta+plim\left(\left(\frac{X'X}{N}\right)^{-1}\frac{X'\upsilon}{N}\right)$$

Omitted Variables and Endogeneity

If $plim\left(\frac{X'\upsilon}{N}\right) \neq 0$ there is an asymptotic bias.

"errors are correlated" with the rhs variable, x

$$plim\left(\frac{X'X}{N}\right) = Var(x) = \sigma_w^2 + \sigma_{\varepsilon_x}^2$$

$$plim\left(\frac{X'\upsilon}{N}\right) = \text{cov}\left(x, \alpha w + \varepsilon_y\right) = \alpha\left(\sigma_w^2\right)$$

$$\Rightarrow plim\hat{\beta} = \alpha\left(\frac{\sigma_w^2}{\sigma_w^2 + \sigma_{\varepsilon_x}^2}\right)$$

“Endogeneity” Problem

In this example, the omitted variable induces a correlation between the rhs variable and the error term. This creates the asymptotic bias. We can’t use estimators that are based on a regression estimator.

$$y = \beta x + \varepsilon_y \qquad E\left[\varepsilon_y \mid x\right] = f\left(x\right) \neq 0$$

This means that this is not a valid regression function and you can’t use the wrong likelihood (the one based on the assumption that it is a regression)!

IV Solution

What is the problem? The variation in x is not independent of y as would be the case if we had experimentally or randomly induced variation.

Solution: assume that a part of the variability of x is "clean" or independent of any other relevant variables. Partition the variability by the use of an instrument.

I.E. "throw out" some portion of the variability of x and hope ("assume") the rest is clean.

A Simple Linear IV Model

$$y = \beta x + \varepsilon_y$$

$$x = z'\delta + \varepsilon_x$$

$$\begin{pmatrix} \varepsilon_y \\ \varepsilon_x \end{pmatrix} \sim N\left(0, \Sigma\right)$$

Here only that portion of the variation of x which can be explained by the instruments (z) can be used to infer about beta.

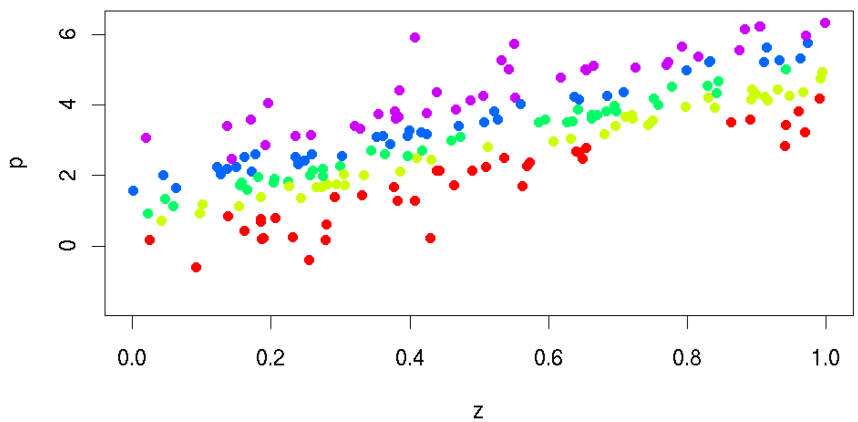
Identification Intuition

A simple example

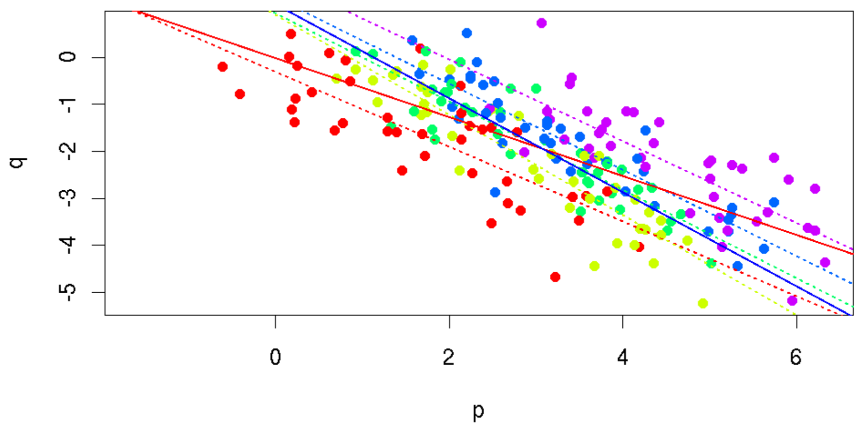
$$p = 4z + \varepsilon_1$$

$$q = -1p + \varepsilon_2$$

$$corr\left(\varepsilon_1, \varepsilon_2\right) = .8$$



Colors
hold
error
variation
fixed



Intuition on IV Estimators

Consider the case of only one instrument. Compute the conditional distribution of $y,x \mid z$ – called the reduced form:

$$y = \beta\delta z + \left(\beta\varepsilon_x + \varepsilon_y\right) = \pi_1 z + v_1$$

$$x = \delta z + \varepsilon_x = \pi_2 z + v_2$$

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \sim N\left(0,\Omega\right)$$

Intuition on IV Estimators

The reduced form is simply a Multivariate Regression Model for which MLE is least squares. We can then form the ratio of least squares estimates.

$$\hat{\beta}_{ILS} = \frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{\left(z'z\right)^{-1} z'y}{\left(z'z\right)^{-1} z'x} = \left(z'x\right)^{-1} z'y$$

The intuition here is that if we move z, then both the x, y outcomes are altered. The ratio is the effect of x on y.

Asymptotic Distribution

$$plim\left(\hat{\beta}_{ILS}\right) = plim\left(\frac{z'x}{N}\right)^{-1}\frac{z'y}{N} = \text{plim}\left(\beta + \left(\frac{z'x}{N}\right)^{-1}\frac{z'\varepsilon_y}{N}\right)$$

$$= \beta + \text{cov}\left(z,x\right)\text{plim}\left(\frac{z'\varepsilon_y}{N}\right) = \beta$$

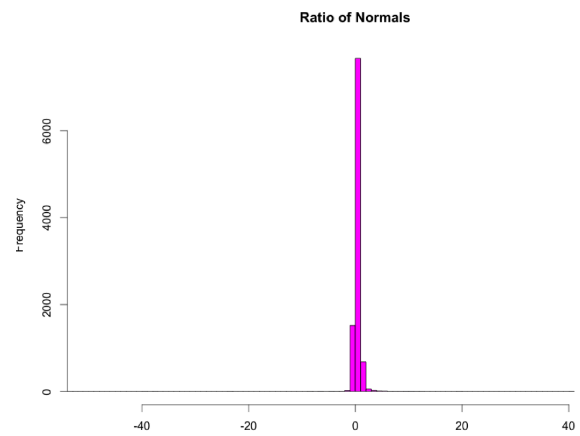
$$\sqrt{N}\left(\hat{\beta} - \beta\right) = \left(\frac{z'x}{N}\right)^{-1}\sqrt{N}\frac{z'\varepsilon_y}{N}$$

$$\sqrt{N}\frac{z'\varepsilon_y}{N} \xrightarrow{d} N\left(0,\sigma^2\right)$$

Intuition on IV Estimators

Asymptotics is normal since the denominator is assumed to converge to a non-zero constant.

$$\hat{\beta}_{ILS} = \frac{z'y}{z'x}$$



But, suppose z'x has a distribution concentrated on "small" values. Then we have problems. The sampling distribution behaves like a ratio of normals.

2SLS

Another way of motivating the IV estimator is two stage least squares: 1. regress x on z; 2. put the fitted value into the "structural" equation.

$$x = \hat{x} + e_x = \hat{\delta} z + e_x$$

fit

$$y = \hat{\beta}_{2SLS} \hat{x} + e_y$$

or

$$y = \hat{\beta}_{2SLS} x + c e_x$$

If you put x and the residual into a multivariate regression, only that portion of x which varies independently of the residual is used.

Sounds like Petrin and Train!

Costs|Dangers of the IV approach

1. Reduces variation in rhs variables
2. Requires a valid instrument. There is no test for validity of instruments. Validity must stem from economic reasoning. Validity means that the IV must not enter the outcome or structural equation. (exclusion or (y, z) indep given x).
3. There is no valid test for endogeneity- Hausman test (compare IV and non-IV procedures) requires a valid instrument.

y,

Costs|Dangers of the IV approach

4. Obviously, we want a strong instrument so we can have maximum precision in estimation. Can there ever be a strong, yet, valid instrument? Or would we ever care about endogeneity in the case of a strong instrument?

5. IV estimators are biased (finite sample) with sampling distributions that are hard to approximate. Standard Asymptotics fails in the case of many and weak instruments. Asymptotic standard errors are much too small and nominal 95 CI can have coverage less than .75.

Estimation Versus Inference

In econometrics, there has been a lot of emphasis on improved inference – getting CI with the correct size.

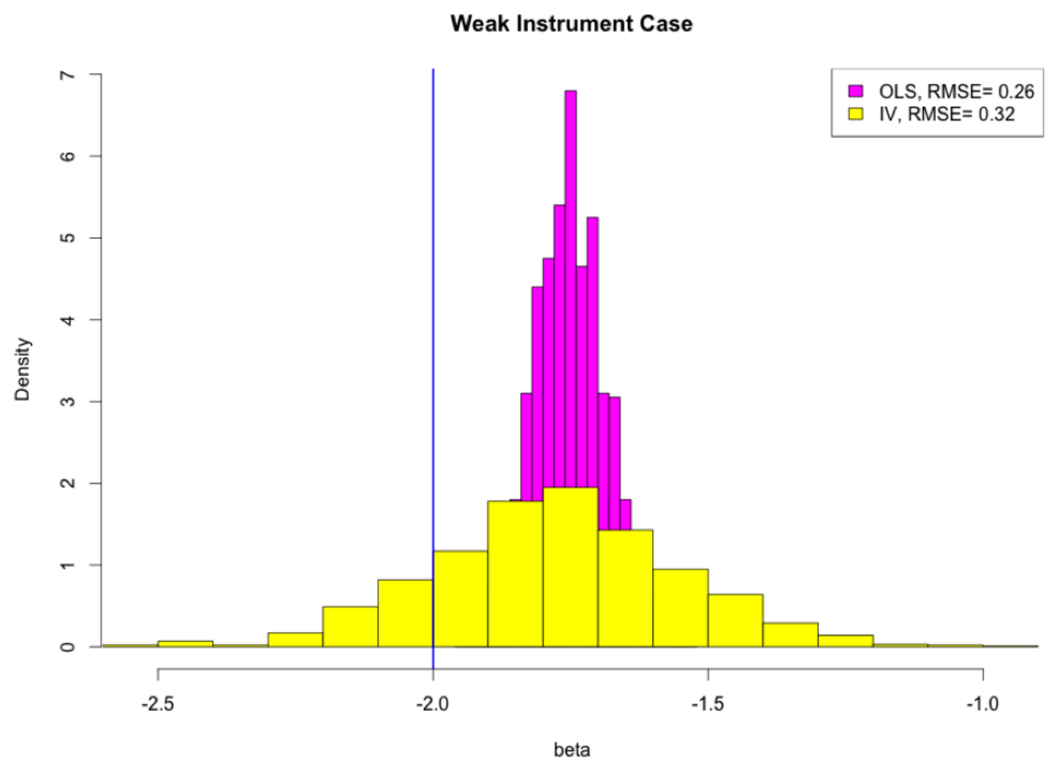
Are we testing or are we making predictions?

Perhaps, we might prefer an estimator with bias but smaller variance.

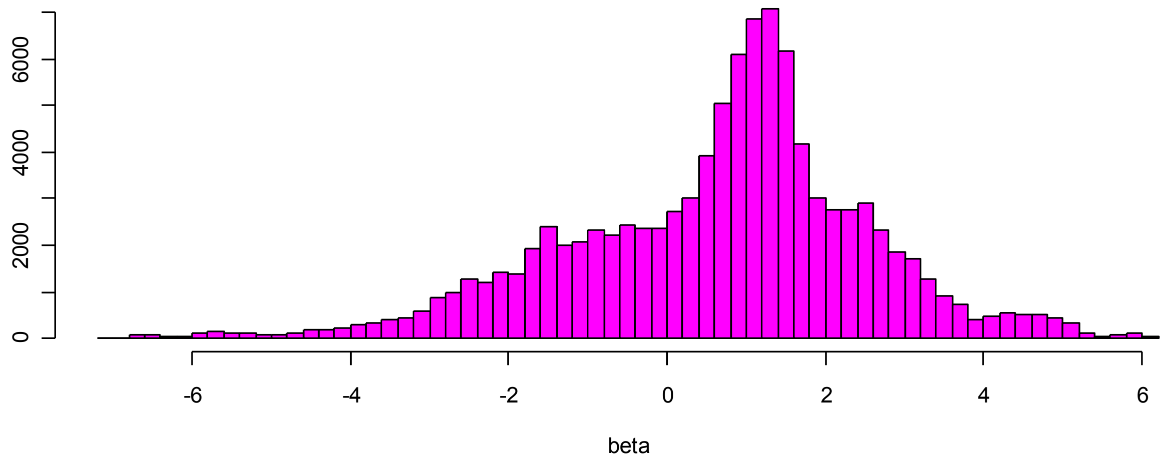
For marketing mix problems, what we really want is a good estimate of the profit function and how this varies with the marketing mix variables. For some situations, this can be obtained without the use of instruments!

Weak Instruments

Consider a situation with a good deal of endogeneity (1/4 variation in the structural error co-moves with x). First stage R-squared of 1 percent. N=200



Weak Ins Ex

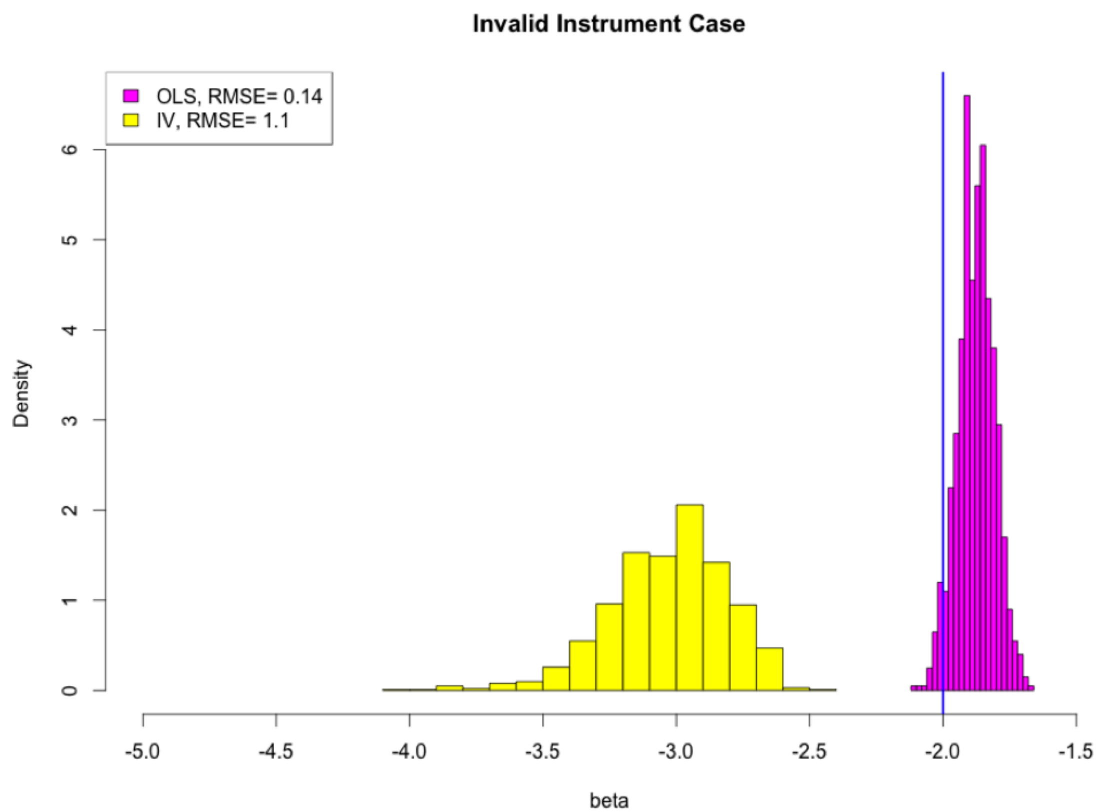


Posterior distribution showing pronounced non-normality and huge long tails.

You can see why asymptotics fail. Asymptotics are based on local curvature of the likelihood (or moment) criterion function. This likelihood is very curved locally!

Invalid Instruments

An instrument is a variable that moves around x but has no direct effect on y. Suppose we use an invalid instrument.



When Should We Worry about Endogeneity?

One definition of endogeneity problem:

> Any situation in which there is an unobservable that influences an outcome (e.g. sales) that is observed by the firm and incorporated in the setting the rhs marketing mix variables (enter into the FOC for firm optimization).

In this sense, endogeneity is ubiquitous. It also applies at any level data aggregation (from market to the individual level).

Endogeneity of Price (it this a fad?)

How did we get so fixated on price endogeneity?

Fitted sales models have “too small” elasticities.

“too” small: implied optimal prices higher than actual prices.

Explanation:

1. Failure to include competing stores or chains.
2. Mis-specified functional forms and heterogeneity bias.
3. Endogeneity bias

Price Endogeneity

Consider a generic example:

$$y_{ijt} \mid p_{ijt}$$

i is the index of cross-sectional unit (consumer, store, market)

j is the brand index

t is time

Some Examples

Types of endogeneity concerns wrt to p_{ijt}.

Over j: unobserved product characteristics

solution: brand “intercepts” or brand specific parameters

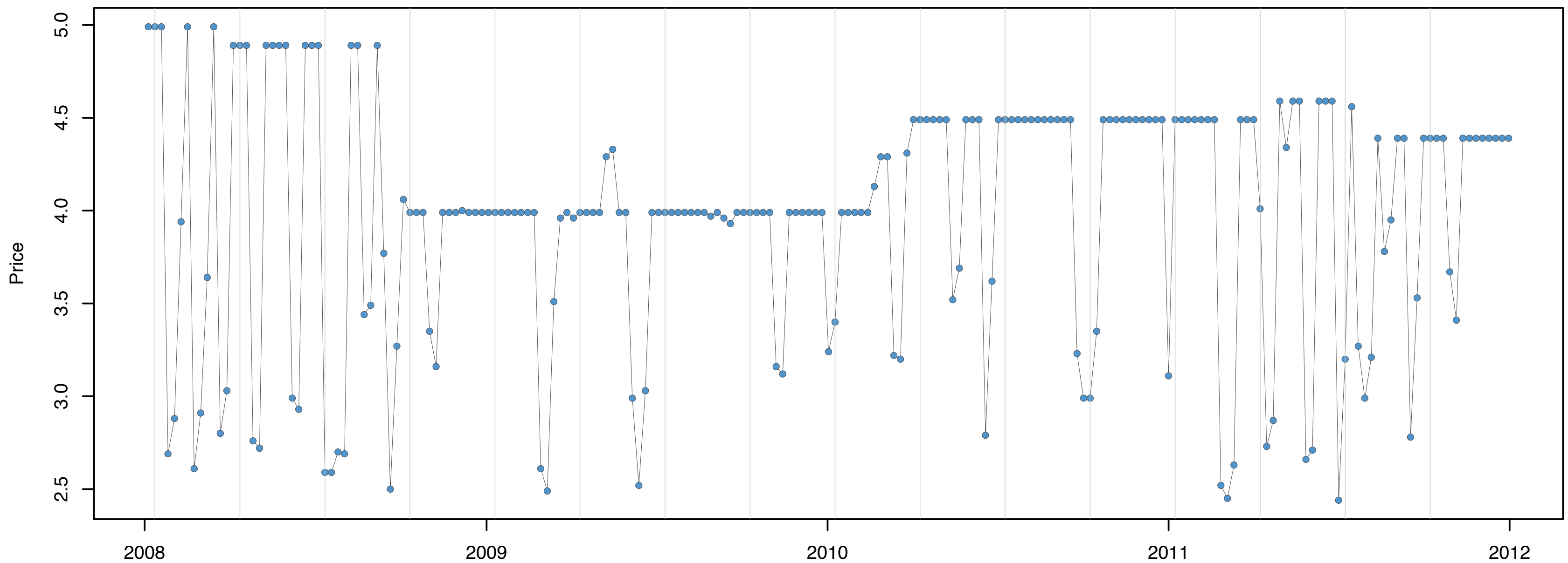
Over i: unobserved “market” characteristics.

solution: market-specific effects.

Over t: what is this unobserved demand shock that varies over time? If t is weekly or higher frequency, this is difficult to believe.

solution: don’t do anything about this or iv?

Price Endogeneity



Prices of Kellogg’s Frosted Mini-Wheats over time at one store. What is the source of this variation?

Price Endogeneity

Variation in “Regular” or Base Prices

Temporary price promotions or sales

Those express concerns regarding endogeneity would have you believe that retailers change prices on a week to week basis because they anticipate market wide demand shocks.

What are these demand shocks? (at the product level, from week to week) Can they possibly be important? If they are advertising, isn’t this just a standard omitted var problem?

Price Endogeneity

It looks like all or at least the vast majority of price variation is related to long and short run price changes. If there is an endogeneity problem, it can't be of economic consequence.

If I instrument with wholesale prices, then I eliminate variation in prices due to temporary price cuts or "sales." That will reduce NOT increase price elasticity.

Isn't the more important question to understand how to model long run vs short run price changes? -- see the bald man and Erdem, Imai and Keane.

Some Suggestions

Use Fixed Effects and Controls instead of instruments where possible.

Don’t use weak instruments.

Obviously, well executed randomized experiments are the ultimate valid instruments. Ironically, we see little of this in marketing relative to economics.

Fixed Effects

Consider the general sales response model:

$$y_{ijt} = \alpha_{ijt} + \beta_{ijt} x_{ijt} + \varepsilon_{ijt}$$

The term “fixed” effects refers to the possibility that each unit of observation (market, product, or time period) could have unique parameters. In principle, all model parameters could be unit-specific. In this case, we simply estimate the model independently on all units.

Fixed Effects

Clearly this is inefficient and possibly infeasible (singular X matrix for some units). For this reason, methods which combine data from multiple units (pooling) are used. One possibility is to simply allow the intercepts to vary, only.

$$y_{ijt} = \alpha_{ijt} + \beta x_{ijt} + \varepsilon_{ijt}$$

In this model, we say there are market, product, and time specific “fixed” effects. This is clearly a saturated model. Let’s remove the time component.

Fixed Effects

$$y_{ijt} = \alpha_{ij} + \beta x_{ijt} + \varepsilon_{ijt}$$

Because this is a linear model, we can eliminate the intercept “fixed” effects parameters by differencing or centering:

$$y_{ijt} - y_{ij,t-1} = \beta\left(x_{ijt} - x_{ij,t-1}\right) + \left(\varepsilon_{ijt} - \varepsilon_{ij,t-1}\right)$$

$$y_{ijt} - \overline{y}_{ij.} = \beta\left(x_{ijt} - \overline{x}_{ij.}\right) + \left(\varepsilon_{ijt} - \overline{\varepsilon}_{ij.}\right)$$

Fixed Effects

What are we doing?

Centering: we now use ONLY variation for unit ij over time. All cross-sectional variation is thrown out.

Differencing: only changes over time within unit matter.

Simple to implement. However, errors are no longer iid and have autocorrelation within unit. A “cluster” covariance estimate should be used.

Fixed Effects Rationale

There is some unobservable that varies across units and effects both y and x. By including the fixed effect, we are actually estimating the level of the unobservable in a so-called "non-parametric" way (we make no restrictions on the distribution of the unobservable).

The presence of the unobservable means that units with similar *x*s are not necessarily similar. However, a unit is always similar to itself. So the valid "experiment" is time series or "within-in" variation in *x*.

Errors in the Variables

Let's review. If a regressor is measured with error, there is a downward bias in the OLS estimates.

$$y = \beta x^* + \varepsilon$$

$$x = x^* + \upsilon$$

or

$$y = \beta\left(x - \upsilon\right) + \varepsilon = \beta x + \left(\varepsilon - \beta\upsilon\right)$$

$$plim\left(\hat{\beta}_{OLS}\right) - \beta = \beta\left(\frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \sigma_{x^*}^2}\right)$$

Fixed Effects and Measurement Error

The Error-In-the-Variables asymptotic bias is minimized by maximizing total variation in the true construct. If you reduce variation in the true value of x, then you are increasing the error-in-the-variables bias.

Fixed effects procedures radically reduce total variation (particularly if you put in unit fixed effects) and, therefore, maximize errors-in-the-variables bias.

Many report that estimates of beta decline as you put in fixed effects. Does this mean you have unobservables or errors-in-the-variables?

Fixed Effects and Measurement Error

We may be in much better shape than many economists in that we might argue that price is measured with little error. Although this is not true for many datasets where prices are merged in from other sources.

Are promotion and advertising variables measured with little error?

All variables collected via a survey style method may have a serious EIV problem.

An Example Involving Fixed Effects

Larry (JP), Moe (Peter), and Curly (Guenter) are interested in the determinants of demand for Private Labels. In particular, we want to estimate the income expenditure elasticity.

$$S_{it} = \eta \ln\left(Income_{it}\right) + x_{it}^{'}\gamma + \varepsilon_{it}$$

S is PL share of total expenditure for household i in month t. Panel of about 35,000 households observed for at least one year. *X* are covariates such as education etc.

An Example Involving Fixed Effects

There are two sources of variation: 1. across panelists and 2. across time for the same panelist. There is a great deal of income variation across panelists and there is a very large sample of panelists.

JP and Guenter have received conventional applied econometrics/IO. Their training has taught them to be suspicious of effects estimated by cross-sectional regressions.

An Example Involving Fixed Effects

Why?

Standard argument: there is some unobservable that is correlated with both income and demand for private labels. Perhaps, subscription to CR. Or lower income folks shop at stores where the relative price of PL is lower/different than for higher income folks.

Larry and Curly want to throw out ALL of the cross-sectional variation in income by putting in household fixed effects.

$$S_{it} - S_{i.} = \eta \ln\left(Inc_{it} - Inc_{i.}\right) + \left(x_{it} - x_{i.}\right)' \gamma + \varepsilon_{it}$$

An Example Involving Fixed Effects

To be fair, Larry and Curly don't necessarily reject the cross-sectional results but they wouldn't accept them as valid unless the cross-section results agree with the fixed effects results. They regard this as proof that there aren't any unobservables.

Problems with this argument:

1. Assumes that income effect measured across households has the same economic meaning as the effect measured over time in the same household.
2. Assumes that measurement error in income is small.

An Example Involving Fixed Effects

Moe thinks there is a great deal of measurement error in income (panelists don't aren't forced to update income each year and it is not clear what year the income applies to).

His preferred specification is a cross-sectional regression with average income to reduce measurement error. Average income probably should be interpreted as a proxy for permanent income.

What about responses to year to year changes in income? This is transitory income.

An Example Involving Fixed Effects

In this example, we expect the cross-sectional regression to yield higher income elasticities than the "within" or fixed effects regression from an elementary application of the theory of permanent income.

It doesn't mean there is a correlated unobservable lurking out there!

Before you throw in fixed effects, I think you have to have a good reason. A better set of controls is always more valuable than "throwing in the towel" and putting in fixed effects.

The Dangers of Fixed Effects

Fixed effects estimation methods throw out a great deal of the variation in the data.

Measurement error problems are frequently ignored.

Fixed effects approach does not extend in any meaningful way to nonlinear models like choice models.

If results change when you dump in fixed effects, this does not mean there is an unobservable!

You need a strong argument that your controls are not adequate.

Conclusions

We are very fortunate in marketing to have such excellent data.

The key characteristic of our data is the large variation in marketing mix variables.

We should be very cautious about adopting methods which are profligate with data.

There has to be a strong argument that we might expect not only endogeneity but a high degree of endogeneity bias before resorting to IV methods.

Conclusions

Ultimately, the only way to properly evaluate instruments is in the context of a joint model of supply and demand.

There may be few valid and/or useful instruments for many Marketing Mix vars like advertising!

There is a stronger argument that fixed effects approaches are more useful that IV methods in marketing applications, particularly ones where the rhs variable of interest is measured well.

Conclusions

Theoretically, arguments for endogeneity do not depend on the level of data aggregation. However, it is very hard to make the argument that endogeneity bias is large with individual consumer data as the relative variance of the common component to the total error variation will be small.