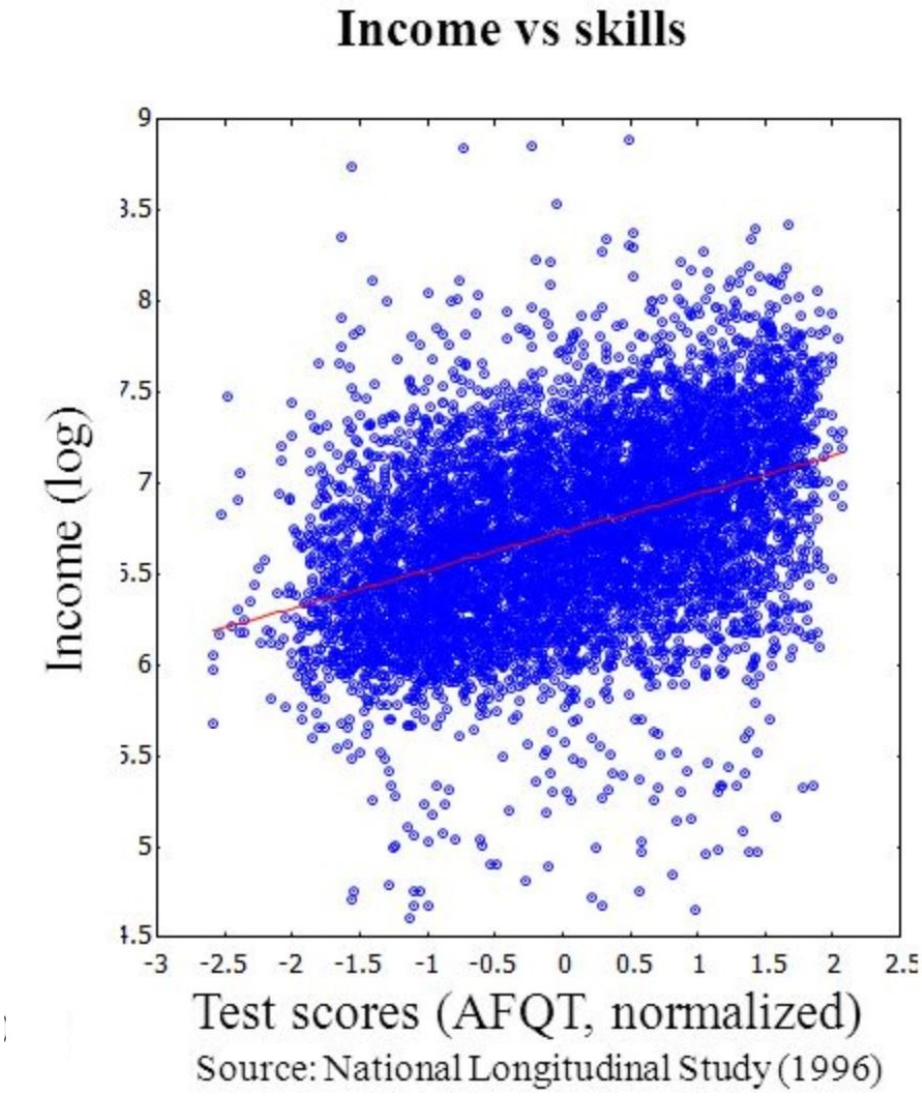
Measurement Errors and Identification

Measurement Error Illustration



Measurement Error Illustration

- "The data sets that economists typically use contain some kind of measurement errors, such as in income, expenditures, or skills. So, if you were to ignore the fact that there are errors in the data you observe, you may reach biased conclusions" **Susanne Schennach**

- "Measurement does not just mean data collection. Much of what economists measure is difficult or impossible to directly observe, like utility, ability, risk aversion, bargaining power, outside options, and joint consumption. Accurately measuring these and other inequality related concepts requires a combination of data collection, structural models of behavior, and econometric identification and estimation." **Arthur Lewbel**

Measurement Issues in Economics

- Economist need noise in their models: humans behavior is hard to predict - we are not all alike - Not everything can be measured directly - etc...
- We need noise to account for: model imperfection - Heterogeneity - Data limitations (latent variables, measurement error)
- Proper statistical analysis of "errors" is crutial for identification and estimation in policy evaluation and econometrics in general
- This is now a very active field of research

- Let's consider that we have a variable x^* that is unobserved
- We only observe $x = x^* + \Delta x$
- We cannot uncover the true variable x at the level of individual observations. Why?
- It's often sufficient to recover true x^* on average: how?
- In policy evaluation, we are interested in ATE on some population
- Taking averages: $(1/n) * \sum_{i=1}^{n} x_i = (1/n) * \sum_{i=1}^{n} x_i^* + (1/n) * \sum_{i=1}^{n} \Delta x_i$
- LLN $\implies$ averages converge asymptotically to expectations
- If we have classical measurement error: $E(\Delta X) = 0$
- This is the basic idea behind ME models but in more sophisticated way

Measurement Error Models

- In practice, things are often more complicated:
  - Model may be non-linear in mis-measured variables (even basic OLS)
    - Non-linearity in mis-measured variables $\Longrightarrow$ we cannot separate the measured variable between signal and noise
  - We may have non-classical ME $E(\Delta X) \neq 0$

- How can we deal with that?
  - **Assume knowledge of the distribution of the measurement error (mostly statisticians) and use deconvolution theorem to correct for ME**

Measurement Error Models

- In practice, things are often more complicated:
  - Model may be non-linear in mis-measured variables (even basic OLS)
    - Non-linearity in mis-measured variables $\Longrightarrow$ we cannot separate the measured variable between signal and noise
  - We may have non-classical ME $E(\Delta X) \neq 0$

- How can we deal with that?
  - **Assume knowledge of the distribution of the measurement error (mostly statisticians) and use deconvolution theorem to correct for ME**
  - **Use of validation data: extra dataset with perfectly measure variables (not always easy to get)**

Measurement Error Models

- In practice, things are often more complicated:
  - Model may be non-linear in mis-measured variables (even basic OLS)
    - Non-linearity in mis-measured variables $\Longrightarrow$ we cannot separate the measured variable between signal and noise
  - We may have non-classical ME $E(\Delta X) \neq 0$

- How can we deal with that?
  - **Assume knowledge of the distribution of the measurement error (mostly statisticians) and use deconvolution theorem to correct for ME**
  - **Use of validation data: extra dataset with perfectly measure variables (not always easy to get)**
  - **Use repeated measurement or instruments (when available)**
    - Survey same individual over time, survey couples about same thing etc..

Measurement Error Models

- In practice, things are often more complicated:
  - Model may be non-linear in mis-measured variables (even basic OLS)
    - Non-linearity in mis-measured variables $\Longrightarrow$ we cannot separate the measured variable between signal and noise
  - We may have non-classical ME $E(\Delta X) \neq 0$

- How can we deal with that?
  - **Assume knowledge of the distribution of the measurement error (mostly statisticians) and use deconvolution theorem to correct for ME**
  - **Use of validation data: extra dataset with perfectly measure variables (not always easy to get)**
  - **Use repeated measurement or instruments (when available)**
    - Survey same individual over time, survey couples about same thing etc..
  - Use higher order moments beyond mean and covariance (often strong assumptions)

Measurement Error Models

- In practice, things are often more complicated:
  - Model may be non-linear in mis-measured variables (even basic OLS)
    - Non-linearity in mis-measured variables $\Longrightarrow$ we cannot separate the measured variable between signal and noise
  - We may have non-classical ME $E(\Delta X) \neq 0$

- How can we deal with that?
  - **Assume knowledge of the distribution of the measurement error (mostly statisticians) and use deconvolution theorem to correct for ME**
  - **Use of validation data: extra dataset with perfectly measure variables (not always easy to get)**
  - **Use repeated measurement or instruments (when available)**
    - Survey same individual over time, survey couples about same thing etc..
  - Use higher order moments beyond mean and covariance (often strong assumptions)
  - Bound the parameters of interest under plausible assumptions: set identification

Distributional Assumptions

Kernel Deconvolution

- If we are willing to make assumptions on the distribution of ME, we can use deconvolution to recover error-free quantities from error-contaminated data (in classical ME)
- Let's first introduce the notion of characteristic function (*cf* - Fourier transform for probability measures)
- $\phi_X(\zeta) = E[e^{i\zeta \cdot X}]$, $\forall \zeta \in \mathcal{R}^{d_X}$ and $i^2 = -1$
- This *cf* has many properties including the convolution theorem $X = X^* + \Delta X$ with $X^* \perp \Delta X \implies \phi_X(\zeta) = \phi_{X^*}(\zeta)\phi_{\Delta X}(\zeta)$
- We have:

$$\phi_{X^*}(\zeta) = \frac{\phi_X(\zeta)}{\phi_{\Delta X}(\zeta)}$$

- We get f using inverse Fourier transform

$$f_{X^*}(x^*) = (2\pi)^{-d_X} \int \phi_{X^*}(\zeta) e^{-i\zeta \cdot x^*} d\zeta$$

- This identification result can be turned into an estimator by using the fact that kernel smoothing is also a type of convolution

- The Fourier transform of a KDE is
$\hat{\phi}_X(\zeta) = \dot{\phi}_X(\zeta)\phi_K(h\zeta)$

- where $\dot{\phi}_X(\zeta) = \frac{1}{n}\sum_{j=1}^{n} e^{-i\zeta \cdot X_j}$ is the *empirical cf* of $(X_1, ..., X_n)$ and $\phi_K(\epsilon)$ is the Fourrier transform of the kernel, h is the bandwidth

- This connection with kernel smoothing and deconvolution leads to the kernel deconvolution estimator

$$\hat{f}_{X^*,h}(x^*) = (2\pi)^{-d_X} \int \dot{\phi}_X(\zeta) \frac{\phi_K(h\zeta)}{\phi_{\Delta X}(\zeta)} e^{-i\zeta \cdot x^*} d\zeta$$

Instruments and Repeated Measurements

Linear/non-linear model with classical ME models

- IV estimators can be used to estimate $Y = \theta X + \epsilon$, where $\epsilon \not\perp X$
- This correlation of error terms could be due to ME (also endogeneity)
- Going back to our measurement model: $X = X^* + \Delta X$ where $E(\Delta X) = 0$ and $\Delta X \perp X^*$
- True model $Y = \theta X^* + \Delta Y$ is related to observed model $Y = \theta X + \epsilon$

$$Y = \theta X^* + \Delta Y = \theta X - \theta \Delta X + \Delta Y = \theta X + \epsilon$$

- where $\epsilon = -\theta \Delta X + \Delta Y$ is correlated with X and can be corrected by an IV
- This AS in the error term fails in non-linear case so IV doesn't work anymore : Why?

Linear/non-linear model with classical ME models

- IV estimators can be used to estimate $Y = \theta X + \epsilon$, where $\epsilon \not\perp X$
- This correlation of error terms could be due to ME (also endogeneity)
- Going back to our measurement model: $X = X^* + \Delta X$ where $E(\Delta X) = 0$ and $\Delta X \perp X^*$
- True model $Y = \theta X^* + \Delta Y$ is related to observed model $Y = \theta X + \epsilon$

$$Y = \theta X^* + \Delta Y = \theta X - \theta \Delta X + \Delta Y = \theta X + \epsilon$$

- where $\epsilon = -\theta \Delta X + \Delta Y$ is correlated with X and can be corrected by an IV
- This AS in the error term fails in non-linear case so IV doesn't work anymore : Why?

$$Y = g(X^*) + \Delta Y = g(X - \Delta X) + \Delta Y$$

- Need different method to deal with non-linear ME

- Several ways to deal with it (See Schennach, ARE 2016)
- One can use repeated measurement to correct for ME in this case: $X = X^* + \Delta X$ and $Z = X^* + \Delta Z$
- This approach use the **Kotlarski identity** (Kotlarski 1967, Roa 1992):
- If $X, Z \in \mathcal{R}$ with $\Delta X, \Delta Z, X^*$ mutually independent and $E(\Delta X) = 0$, then

$$f_{X^*}(x^*) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} exp\Big(\int_0^{\omega} \frac{iE[xe^{i\zeta Z}]}{E[e^{i\zeta Z}]} d\zeta \Big) e^{-k\omega x^*} d\omega \tag{1}$$

- This expresses the pdf of X^* just in terms of moments of the observables without assuming that we know the pdf of the ME
- requires only 2 error contaminated measurements
- Extensions: Multivariate in Li & Vuong (1998), regression setting in Li (2002), weaker assumptions (see in Schennach, 2016)

Non-Classical Measurement Error

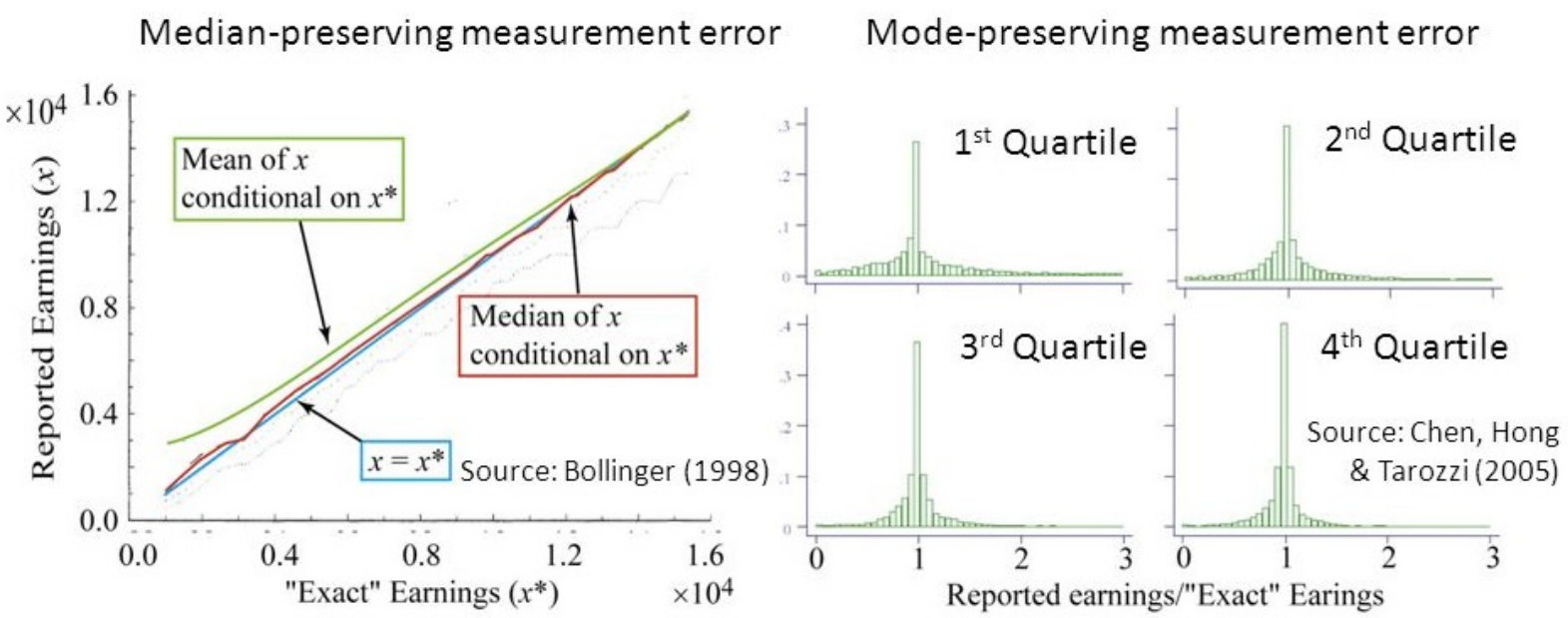
Non-Classical Measurement Error

- So far we were dealing with classical ME but we may have $E(\Delta X) \neq 0$
- We can relax the classical ME assumption which complicates things
- It is however crutial for policy evaluation
- When regressor X is a dummy variable for participating to a program, we have **misclassification**.
- This is a non-classical ME setting. Why?
- classical ME assumptions fails in this setting

Non-Classical Measurement Error

Ways to deal with non-classical ME

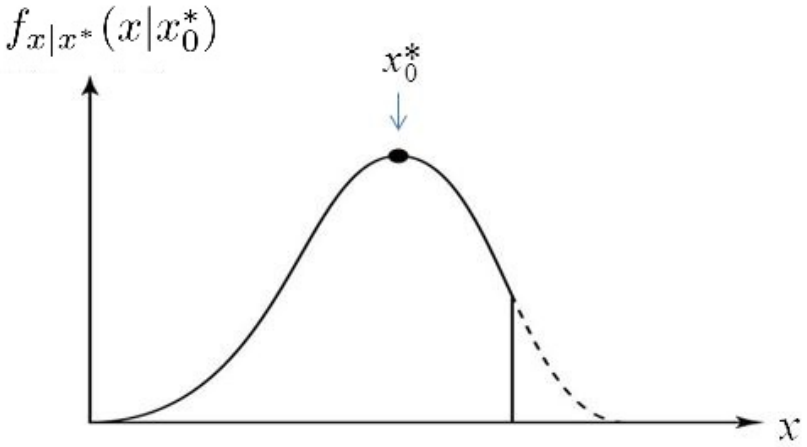
- Centering condition: Replace zero mean by - Measurement error has zero Mode, Median, Quantile, etc.



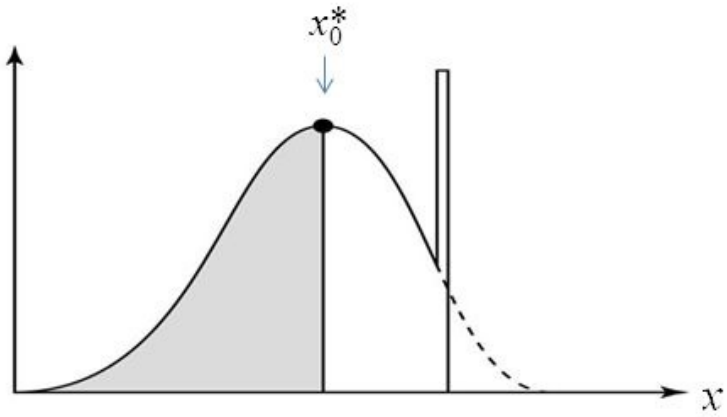
Other types of conditions

Nonclassical Measurement Error

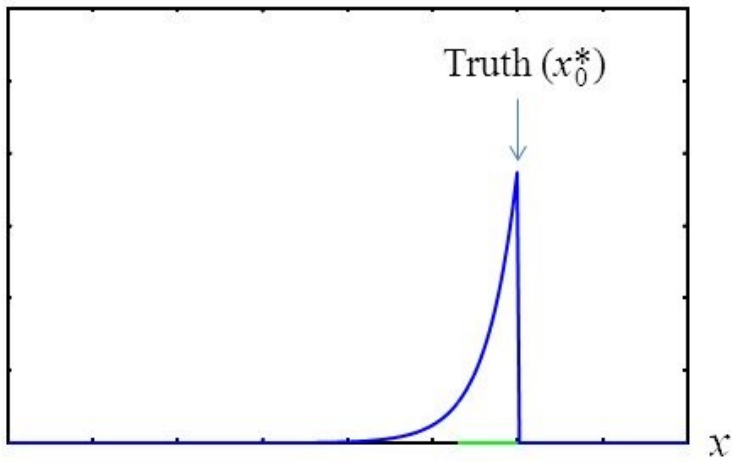
Truncation preserves mode



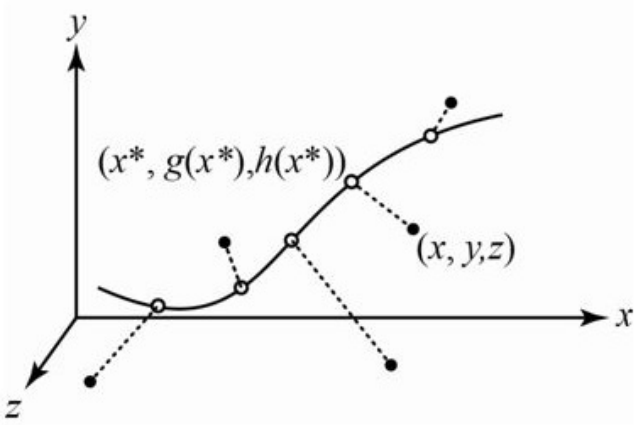
Censoring preserves median



Systematic Under-reporting:



Non-Classical Measurement Error: General result



- Assume we have access to 3 distorted error-contaminated measurements of X^* (continuous): X, Y, Z
- These measurements could be: Repeated measurements - Instruments - Dependent variable of regression
- Goal: use observed distribution of measurements to recover the distribution of the unobserved variable

Non-Classical Measurement Error: General result

- If X, Y, Z are mutually independent conditionnal on x^*, then

$$f_{YXZ}(y, x, z) = \int_{\chi^*} f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*) dx^*$$

- Model is identified if there is a unique solution to this equation
- Hu and Schennach (Econometrica, 2008) show this is the case if $f_{X|X^*}(x|x^*)$ is centered around x^* + other technical conditions
- Y can be a dependent variable and X^* a regressor in a regression model
- X, Y, Z can also be a measurement system in a general nonlinear (dynamic) factor model
- This generalizes many identification results: linear factor models (Anderson & Rubin 1956) and models with hidden discrete variable (Kruskal, 1977)

Misclassification

- Let's go back to identifying ATE from binary treatment variable
- If treatment dummy is measured with errors, this will be a non-classical ME model
- In general, when X takes only finite values, extreme values of X^* can only be mismeasured in one direction
- $E(\Delta X)$ cannot be zero
  treatment is misclassified and an instrument is available

Measurement Error in Dummy Variables

- consider $y_i = \alpha + \beta d_i + \epsilon_i$ where $d_i \in \{0, 1\}$
- Eg: effect of union membership on wages
- Instead of d we observe $\tilde{d}$: misclassifies some observations

$$E(y_i \mid \widetilde{d}_i = 1) = \alpha + \beta P(d_i = 1 \mid \widetilde{d}_i = 1)$$

$$E(y_i \mid \widetilde{d}_i = 0) = \alpha + \beta P(d_i = 1 \mid \widetilde{d}_i = 0)$$

$$\text{plim } \widehat{\beta} = \beta \left[P(d_i = 1 \mid \widetilde{d}_i = 1) - P(d_i = 1 \mid \widetilde{d}_i = 0) \right]$$

(1)

- β will be attenuated because some (high wage) union members are classified nonmembers while some (low wage) nonmembers are classified as members
- Define $q_1 = P(\tilde{d}_i = 1 | d_i = 1)$ probability that we observe somebody to be a union member when he truly is

- and $q_0 = P(\tilde{d}_i = 1|d_i = 0)$: probability that a nonmember is misclassified
- $\pi = P(d_i = 1)$: true membership rate
- The estimate of π by $\hat{\pi} = N^{-1} \sum \tilde{d}_i$ satisfies plim $\hat{\pi} = \pi q_1 + (1 - \pi) q_0$
- By Bayes rule:

$$P(d_i = 1 \mid \widetilde{d}_i = 1) = \frac{P(\widetilde{d}_i = 1 \mid d_i = 1) \cdot P(d_i = 1)}{P(\widetilde{d}_i = 1)} = \frac{\pi q_1}{\pi q_1 + (1 - \pi) q_0}$$

and

$$P(d_i = 1 \mid \widetilde{d}_i = 0) = \frac{\pi(1 - q_1)}{\pi(1 - q_1) + (1 - \pi)(1 - q_0)}$$

So we have

$$
\begin{aligned}
\operatorname{plim} \widehat{\beta} &= \beta \left[\frac{\pi q_1}{\pi q_1 + (1-\pi) q_0} - \frac{\pi (1-q_1)}{\pi (1-q_1) + (1-\pi)(1-q_0)} \right] \\
&= \beta \left[\frac{\pi q_1}{\widehat{\pi}} - \frac{\pi (1-q_1)}{1-\widehat{\pi}} \right] \\
&= \beta \frac{(1-\widehat{\pi}) \pi q_1 - \widehat{\pi} \pi (1-q_1)}{\widehat{\pi}(1-\widehat{\pi})} \\
&= \beta \frac{\pi \left[(1-\widehat{\pi}) q_1 - \widehat{\pi}(1-q_1) \right]}{\widehat{\pi}(1-\widehat{\pi})} \\
&= \beta \frac{\pi \left(q_1 - \widehat{\pi} \right)}{\widehat{\pi}(1-\widehat{\pi})}
\end{aligned}
$$

- Absent knowledge about q_1 and q_0 we cannot identify the true β and π from our data, i.e. from the estimates $\hat{\beta}$ and $\hat{\pi}$
- If we have panel data available and we are willing to impose the restriction that q_1 and q_0 do not change over time, all coefficients will be identified
- Even with just a two period panel there is already one overidentifying restriction

Identification from Repeated Measurement

- To see this, notice that there are now not two states for union status but four possible transitions: continuous union members, continuous nonmembers, union entrants and union leavers
- The key is that there have to be some switchers in the data
- Then we can observe separate changes in y over time for each of the four transition groups
- Furthermore, we observe three independent transition probabilities
  - This makes a total of 7 moments calculated from the data.
  - From these we have to identify β, q_1, q_0 and the 3 true transition probabilities: 6 parameters
- See Card (1996) for detailed algebra and Krueger and Summers (1988) for equivalent results on measurement error in multinomial variables

Identification from IV

- Suppose we have another binary variable z_i available, which has the same properties as the mismeasured dummy variable $\tilde{d}_i$
- Can we use z_i as an instrument in the estimation of β?
- IV will not yield a consistent estimate of β in this case
- Reason: measurement error can only be either -1 or 0 (when $d_i = 1$), or 1 or 0 (when $d_i = 0$).
- This means that the measurement errors in two mismeasured variables will be positively correlated: violates exclusion restriction
- Define $h_1 = P(z_i = 1|d_i = 1)$ and $h_0 = P(z_i = 1|d_i = 0)$
- IV estimator in this case is simply the Wald estimator so that

Identification from IV

$$\text{plim}\ \widehat{\beta}_{IV} = \frac{E(y_i \mid z_i = 1) - E(y_i \mid z_i = 0)}{E(\widetilde{d}_i \mid z_i = 1) - E(\widetilde{d}_i \mid z_i = 0)}.$$

- The numerator has the same form as Eq (1) with z_i replacing $\tilde{d}_i$
- Terms in denominator can also be derived

$$\begin{aligned} E(\widetilde{d}_i \mid z_i = 1) &= P(\widetilde{d}_i = 1 \mid z_i = 1) \\ &= \frac{P(\widetilde{d}_i = 1, z_i = 1)}{P(z_i = 1)} \\ &= \frac{P(\widetilde{d}_i = 1, z_i = 1 \mid d_i = 1)P(d_i = 1) + P(\widetilde{d}_i = 1, z_i = 1 \mid d_i = 0)P(d_i = 0)}{P(z_i = 1 \mid d_i = 1)P(d_i = 1) + P(z_i = 1 \mid d_i = 0)P(d_i = 0)} \\ &= \frac{q_1 h_1 \pi + q_0 h_0 (1 - \pi)}{h_1 \pi + h_0 (1 - \pi)} \end{aligned}$$

- and similarly for $E(\tilde{d}_i | z_i = 0)$. Substituting everything back

Identification from IV

$$\text{plim}\,\widehat{\beta}_{IV} = \frac{\beta\left[\frac{\pi h_1}{h_1\pi+h_0(1-\pi)} - \frac{\pi(1-h_1)}{(1-h_1)\pi+(1-h_0)(1-\pi)}\right]}{\frac{q_1 h_1\pi+q_0 h_0(1-\pi)}{h_1\pi+h_0(1-\pi)} - \frac{q_1(1-h_1)\pi+q_0(1-h_0)(1-\pi)}{(1-h_1)\pi+(1-h_0)(1-\pi)}}.$$

With some elementary algebra this simplifies to

$$\text{plim}\,\widehat{\beta}_{IV} = \frac{\beta}{q_1 - q_0}.$$

- IV estimate of β is biased by a factor $1/(q_1 - q_0)$
- The bias only depends on the misclassification rates in the variable $\tilde{d}_i$ which is being used as the endogenous regressor
- This is because more misclassification in the instrument will lead to a smaller first stage coefficient
- Since in general $1 > q_1 > q_0 > 0$, IV will be biased upwards
- Hence, **OLS and IV estimation could be used to bound the true**

Instrument as repeated measurement

- The true coefficient is actually identified from the data, using an idea analogous to the panel data case above [Kane, Rouse, and Staiger (1999)].
- There are 7 sample moments which can be computed from the data.
  - There are 4 cells defined by the cross-tabulation of d_i and z_i. The mean of y_i can be computed for each of these cells.
  - In addition, we have three independent sampling fractions for the cross-tabulation.
- From these moments we have to identify 7 parameters: $\alpha, \beta, \pi, q_0, q_1, h_0,$ and h_1
- These parameters are indeed just identified and can be estimated by method of moments.

Misclassification

- These are standard textbook results
- Other papers deal with relaxing assumptions to fit empirical cases
- Mahajan (Econometrica, 2006) shows that a binary instruments can be used to identify and estimate an index model (non parametric) with a misclassified binary regressor and other perfectly measured regressors
- Lewbel (Econometrica, 2007): treatment effect models when the treatment is misclassified and an instrument is available
- etc...