LABOUR LECTURES

March 2011

# Lecture 1, Part B

# A Theoretical Foundation For Using Selection on Observed Variables to Assess Selection on Unobserved Variables

Joseph Altonji

Yale University

# 1 Introduction and Overview

- Goal: Provide estimation strategies when strong prior information is unavailable regarding the exogeneity of the variable of interest or instruments for that variable.

- Key Idea: Use the degree of selection on observables as a guide to the degree of selection on the unobservables.
  - Researchers often examining the relationship between the instrumental variable and a set of observed characteristics

- Provide formal analysis confirming the intuition that such evidence can be informative in some situations.

- Provide ways to quantitatively assess the degree of selection bias or omitted variables bias
  - Apply to Measuring the Effectiveness of Catholic Schools and a medical procedure
  - Assessing Validity of an IV strategy. (Apply to Catholic Schools Literature.)

- Provide two bounds estimators
    - Apply one to Catholic Schools and to assessment of a medical procedure.

## Model

$$\begin{aligned} Y &= \alpha T + X'\mathsf{\Gamma}_X + W^{c\prime}\mathsf{\Gamma}^c \\ &= \alpha T + X'\mathsf{\Gamma}_X + W'\mathsf{\Gamma} + \varepsilon, \end{aligned} \tag{1}$$

where

$T$ is potentially endogenous.

$\alpha$ is parameter of interest.

$X$ is a vector of *observed* variables.

$W^c$ is the vector of additional characteristics (observed *and* unobserved) that determine $Y$.

$W$ is the subvector $W^c$ that is observed, $\mathsf{\Gamma}$ is the corresponding subvector of $\mathsf{\Gamma}^c$,

$\varepsilon$ is an index of the unobserved variables.

- Catholic Schools Case: $Y_i$ is high school graduation. $T_i = CH_i$

- Health Application: $Y_i$ is mortality 90 days after admission, $T_i =$1 if patient received catheter.

- Present the general case with an instrument $Z$. A special case is $Z = T$.

Many studies assume $Z_i$ is correlated with some variable of interest $T_i$, but

$$cov(Z_i, \varepsilon_i) = 0.$$

A key special case of this model is OLS, in which $Z_i = T_i$.

Virtually all causal empirical work in economics makes some analogous assumption .

- The best justification for the instrument is random assignment

- If $Z_i$ was truly randomly assigned, it should not be correlated with the observable covariates either

- Researchers have recognized this for a long time.

- Common to run a regression of $Z_i$ on $W_i$ and test whether these are related.
  - eg. compare means of control variables by $Catholic$, as in lecture 1.

- Rejecting the null doesn't mean the assumption is not approximately true, because
  - $W_i$ is a control set
  - We care about the size of the bias in $\hat{\alpha}$, not the $F-$statistic

- Failing to reject may not mean much either.
  - The observed covariates we look at may not be representative of the unobserved factors

- We provide a foundation for the practice of using relationship between an endogenous variable or an instrumental variable and the observables to make inferences about the relationship between these variables and the unobservables.

- We formalize what it means to say “selection on observable covariates is the same as selection on unobservable covariates.”

- Provide a way to quantitatively assess the importance of the bias from the unobservables and to construct bounds for estimates

### 1.0.1 Note:

- What one really needs is for $Z_i$ to be uncorrelated with $\varepsilon_i$ conditional on the observed covariates

- In particular some $X$'s may be related to $Z_i$ by design or may be “special” in that they have an extremely large effect on $T$ or on $Y$.

  - We account for this.

- Correlation between $T$ and $\varepsilon$ will lead to bias. But so will correlation between $W$ and $\varepsilon$.

### 1.0.2 The Degree of Selection on Observables

- Consider $T = Z$ case. Consider the linear projection of $T$ onto $X$, $W'\Gamma$ and $\varepsilon$ :

$$\text{Proj}(T|X, W'\Gamma, \varepsilon) = \phi_0 + X'\phi_X + \phi W'\Gamma + \phi_\varepsilon \varepsilon. \tag{2}$$

- Our formalization of the idea that, after controlling for $X$, "selection on the unobservables is the same as selection on the remaining observables" leads to:

$$\textbf{Condition 1} : \phi_\varepsilon = \phi$$

In contrast the usual OLS orthogonality conditions which imply:

$$\textbf{Condition 2} : \phi_\varepsilon = 0$$

- Condition 1 says that conditional on $X$, the part of $Y$ that is related to the observables and the part related to the unobservables have the **same** relationship with $T$.

- Condition 2 says that the part of $Y$ related to the unobservables has **no** relationship with $T$.

We present a set of assumptions regarding how $W$ is chosen from $W^c$ that imply:

**Condition 3:**

$$\begin{aligned} 0 &\leq \phi_\varepsilon \leq \phi \text{ if } \phi > 0 \\ 0 &\geq \phi_\varepsilon \geq \phi \text{ if } \phi < 0 \end{aligned}$$

Suppose the data collector chose the variables knowing that we were going to do OLS

- As the number of variables gets large, get $\phi_\varepsilon = \phi$ if the data collector had no idea what he was doing and chose what to include in $W$ at random, assuming the number of variables is large.

- Get OLS condition $\phi_\varepsilon = 0$ if we had a perfect data collector. That person would collect all of the variables that were correlated with both $T_i$ and $Y_i$ so that the only unobservables left would be uncorrelated with $T_i$.

- The truth will be in between in most cases.

We propose two estimators that use Condition 3.

They differ in how they model link between $W$ and $\varepsilon$.

## “OU” Estimator:

- Estimate $\alpha$ essentially treating $W$ as exogenous.

- Requires a high level assumption that implies that Condition 3 holds for $\phi$, $\phi_e$ in

$$\text{Proj}(T|X, W'G, e) = \phi_0 + X'\phi_X + \phi W'G + \phi_e e. \tag{3}$$

where $G$ and $e$ are defined so that

$$\begin{aligned} Y &= \alpha T + X'\Gamma_X + W'\Gamma + \varepsilon \\ &= \alpha T + X'\Gamma_X + W'G + e, \end{aligned}$$

and $E(e|W) = 0$.

- (3) and condition 3 provides bounds on the amount of selection.

- OU has been applied in Altonji, Elder and Taber (2005a, 2005b, 2008, hereafter AET) and several other studies.

- Basis for sensitivity analysis in a number of recent papers

## OU-Factor Estimator

- Method of Moments Procedure

- Models the covariance between the observable and unobservable covariates with a factor structure.

- Use factor structure infer properties of unobserved covariates based on the observed correlation structure of the observed covariates $W$, $T$, and $Y$.

- The estimator consistently identifies a set that contains $\alpha$

- Provide a general bootstrap procedure that may be used to construct a confidence interval for the set.

- Less computationally demanding bootstrap procedure that seems to works well in practice.

# 2 Outline of the Rest of Lecture 1b and Lecture 2

- Related Literature, with applications. (mostly skip due to time constraints)

- Discussion of how observables are chosen, and a formal model

- Implications for $\phi$ and $\phi_\varepsilon$. Establish Condition 3 on $\phi$ and $\phi_\varepsilon$

**Lecture 2:**

- The OU Estimator

- Application of OU to Catholic School Effect, Swan Ganz procedure

- Sensitivity analysis related to the OU Estimator, with applications to Catholic school effect and Swan-Ganz

- brief discussion of heterogenous treatment effects case (very preliminary, will probably skip)

- The OU-Factor Estimator

- Consistency of OU-Factor

- Constructing Confidence Intervals

- Monte Carlo Evidence

- Conclusion

# 3 Related Literature

### 3.0.3 Sensitivity Analysis (Rosenbaum and Rubin (1983), Rosenbaum (1995))

- Consider the bivariate probit formulation considered earlier distinguishing $X$ and $W$.

$$\begin{aligned} CH_i &= 1(X_i'\beta_X + W'\beta + u > 0) \\ Y_i &= 1(X_i'\Gamma_X + W_i'\Gamma + \alpha CH_i + \varepsilon > 0) \\ (u, \varepsilon) &\sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}). \end{aligned}$$

- With linear indices and normal error terms this model is technically identified without an exclusion restriction

- Instead, treat model as if we are short by one parameter $(\rho)$

Display estimates of Catholic schooling effects that correspond to various assumptions about $\rho$,

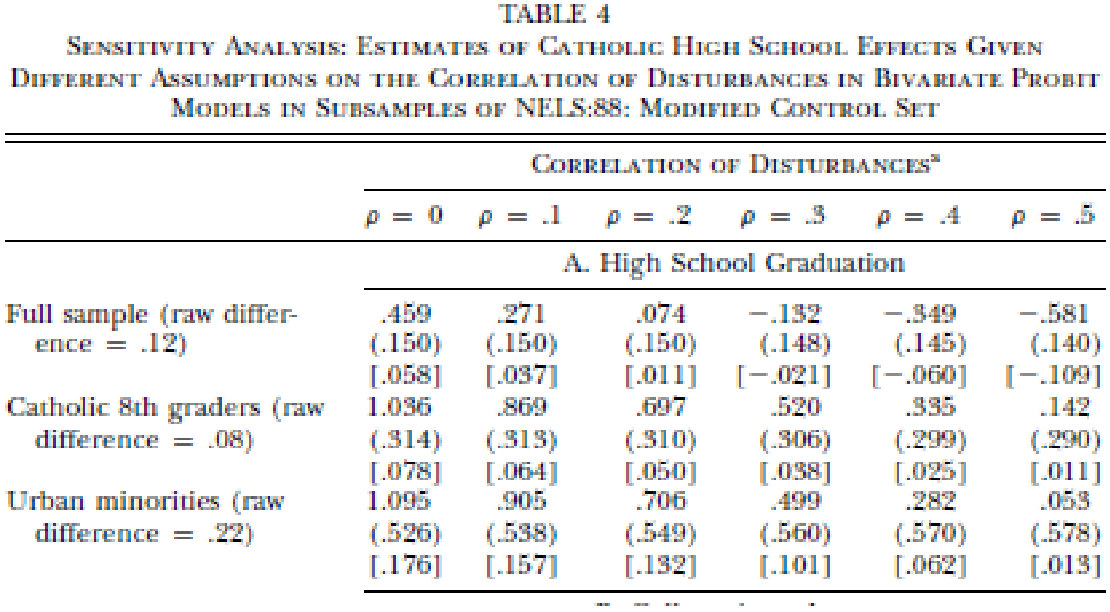


TABLE 4

SENSITIVITY ANALYSIS: ESTIMATES OF CATHOLIC HIGH SCHOOL EFFECTS GIVEN DIFFERENT ASSUMPTIONS ON THE CORRELATION OF DISTURBANCES IN BIVARIATE PROBIT MODELS IN SUBSAMPLES OF NELS:88: MODIFIED CONTROL SET

| | CORRELATION OF DISTURBANCES[a] | | | | | |
|---|---|---|---|---|---|---|
| | $\rho = 0$ | $\rho = .1$ | $\rho = .2$ | $\rho = .3$ | $\rho = .4$ | $\rho = .5$ |
| | A. High School Graduation | | | | | |
| Full sample (raw difference = .12) | .459 | .271 | .074 | −.132 | −.349 | −.581 |
| | (.150) | (.150) | (.150) | (.148) | (.145) | (.140) |
| | [.058] | [.037] | [.011] | [−.021] | [−.060] | [−.109] |
| Catholic 8th graders (raw difference = .08) | 1.036 | .869 | .697 | .520 | .335 | .142 |
| | (.314) | (.313) | (.310) | (.306) | (.299) | (.290) |
| | [.078] | [.064] | [.050] | [.038] | [.025] | [.011] |
| Urban minorities (raw difference = .22) | 1.095 | .905 | .706 | .499 | .282 | .053 |
| | (.526) | (.538) | (.549) | (.560) | (.570) | (.578) |
| | [.176] | [.157] | [.132] | [.101] | [.062] | [.013] |

- In Catholic 8th grade sample, the effect of $CH$ declines from 0.078 when $\rho = 0$ to 0.038 when $\rho = 0.3$. It is still positive when $\rho = 0.5$.

- The estimate of effect of $CH$ on College is negative when $\rho = 0.3$.

- Since table of means suggests only limited selection on the observables, "selection on unobservables" would have to large to explain away the high school effect.

### 3.0.4 Partial Identification, Bounds Estimation

- Large, Rapidly Growing Literature. Many of the papers address selection bias.

- We use theoretical methods of Chernozhukov, Hong, and Tamer (2007)

### 3.0.5 Tangentially Related :

- Literature on Non-Response, Missing Data on Dependent Variables or Covariates for Some observations.

    - Recent Example: Kline and Santos (2010).

- We ignore *item* non-response. We focus on missing *variables*

Our work builds on many, many papers that informally examine patterns in the observables and drawing inferences about selection bais, as we illustrated earlier.

This is how we started.

But what theory says that observed variables provide information about the unobserved variables?

How does one turn examination of patterns into quantitative statements of $\alpha$?

# 4 A "Theory" of What Variables Are Chosen

- Large scale data sets (PSID, British Panel, the German Socioeconomic Panel) are multipurpose

- Content is a compromise among the interests of multiple research, policy making, and funding constituencies.

- Burden on the respondents, budget, and access to administrative data sources serve as constraints.

- Content is also shaped by

  - what is known about what matters for particular outcomes

  - variation in the feasibility of collecting useful information on particular topics.

- Due to constraints and lack of scientific knowledge, many elements of $W^c$ are left out. (low R-squareds)

- Explanatory variables that influence a large set of important outcomes (such as family income, race, education, gender, or geographical information) or are interesting outcomes, are more likely to be collected.

- The optimal survey design for estimation of $\alpha$ would be to assign the highest priority to variables that are important determinants of *both* $T$ and $Y$.

  - BUT: many factors that influence $Y$ and are correlated with $T$ are left out. (Consider low $R^2$)

- *Alternative View*: constraints on data collection are sufficiently severe that it is better to think of the elements of $W$ as a more or less random subset of the elements of $W^c$ rather than a set that has been systematically chosen to eliminate bias.

  - Many variables that affect $Y$ are determined after $T$..

  - Measurement error, random influences (eg., test scores)

- The truth is probably in between optimal variable choice and random variable choice in most cases.

## 4.1 Implications for What is Observed

- Partition $W^c$ into two categories of variables.
  - $W^*$, consists of $K^*$ variables that affect $Y$ and potentially $T$ (and possibly $Z$)
    - $*$ Subvector $W$ of $W^*$ is observed. $W^u$ is not.
  - $W^{**}$. These variables have a 0 probability of being observed and used. Some determined after $T$
  - Index the $W_j$ so that $j = 1, ..., K^*$ corresponds to $W^*$ and $j = K^* + 1, ..., K^c$ corresponds to $W^{**}$.

Let $S_j = 1$ if variable $j$ is observed and 0 otherwise. We can write

$$\begin{aligned} W'\Gamma &= \sum_{j=1}^{K^c} S_j W_j \Gamma_j = \sum_{j=1}^{K^*} S_j W_j \Gamma_j \\ \varepsilon &= \sum_{j=1}^{K^c} \left(1 - S_j\right) W_j \Gamma_j = \sum_{j=1}^{K^*} \left(1 - S_j\right) W_j \Gamma_j + \sum_{j=K^*+1}^{K^c} W_j \Gamma_j = W^{u\prime}\Gamma^u + \xi \end{aligned}$$

$\Gamma^u$ is the subvector of $\Gamma^c$ that corresponds to $W^u$

$\xi = W^{**\prime}\Gamma^{**}$.

- We assume that $\xi$ is orthogonal to $(W^*, T, Z)$.

- For this reason, we use Condition 3

$$\begin{aligned} 0 &\leq \phi_\varepsilon \leq \phi \text{ if } \phi > 0 \\ 0 &\geq \phi_\varepsilon \geq \phi \text{ if } \phi < 0 \end{aligned} \tag{4}$$

as the basis for the estimation strategies developed below.

- 3rd category: $X$
  - factors that play an essential role in determining $Y$ and potentially $Z$ and $T$.
  - Example: Catholic religion in our study of the effects of attending Catholic school on high school graduation.

## 4.2 Implications of Random Selection of Observables

- Allow the number of covariates in $W^c$ to get large and derive the probability limit of $\phi_\varepsilon/\phi$.

- For individual $i$, we define $Y_i$ and $Z_i$ as outcomes for a sequence of models indexed by $K^*$ where $K^*$ is the number of elements of $W^*$.

- The dimensions of $X$ and $W^{**}$ are fixed.

- $\mathcal{G}^{K^*}$ consists of the realization of the $S_j$, the $\Gamma_j$, and the joint distribution of $W_{ij}$ conditional on $j = 1, ..., K^*$.
  - First the “model” is drawn, represented by $\mathcal{G}^{K^*}$ .
  - Then individual data are drawn from the model.

The two steps combined generate $Y_i$ as is represented in Assumption 1.

**Assumption 1:**

$$Y_i = \alpha T_i + X_i' \Gamma + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} W_{ij} \Gamma_j + \xi_i \tag{5}$$

where $(W_{ij}, \Gamma_j)$ is unconditionally stationary (indexed by $j$) and $X_i$ includes an intercept.

Scaling by $\frac{1}{\sqrt{K^*}}$ guarantees that no particular covariate dominates $Y$.

(Dominant variables are in $X$.)

- Take residuals to remove $X$ . Call $\widetilde{W}_{ij}, \tilde{T}_i, \tilde{Z}_i, \tilde{Y}_i$

- Let $\sigma_{j,\ell}^{K^*} = E\left(\widetilde{W}_{ij}\widetilde{W}_{i\ell} \mid \mathcal{G}^{K^*}\right)$.

**Assumption 2**

$$0 < \lim_{K^* \to \infty} \frac{1}{K^*} \sum_{j=1}^{K^*} \sum_{\ell=1}^{K^*} E(\sigma_{j,\ell}^{K^*} \Gamma_j \Gamma_\ell) < \infty$$

and

$$\lim_{K^* \to \infty} Var\left(\frac{1}{K^*} \sum_{j=1}^{K^*} \sum_{\ell=1}^{K^*} \sigma_{j,\ell}^{K^*} \Gamma_j \Gamma_\ell)\right) \to 0 \ .$$

The next two assumptions guarantee that $cov(Z_i, Y_i)$ is well behaved as $K^*$ grows.

**Assumption 3** For any $j = 1, ..., K^*$,define $\mu_j^{K^*}$ so that

$$E\left(\widetilde{Z}_i \widetilde{W}_{ij} | \mathcal{G}^{K^*}\right) = \frac{\mu_j^{K^*}}{\sqrt{K^*}}$$

then

$$E(\mu_j^{K^*} \Gamma_j) < \infty.$$

and

$$\lim_{K^* \to \infty} Var\left(\frac{1}{K^*} \sum_{j=1}^{K^*} \mu_j^{K^*} \Gamma_j\right) \to 0\ .$$

To consider Assumption 3, we need a model for $Z$.

**Assumption 4**

$$Z_i = X_i'\beta_X + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} W_{ij}\beta_j + \psi_i, \tag{6}$$

Convenient to rewrite the model for $Z$ as

$$Z_i = X_i'\beta_x + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K} \tilde{W}_{ij}\beta_j + u_i \tag{7}$$

where $u_i = \frac{1}{\sqrt{K^*}} \sum_{j=K+1}^{K^*} \tilde{W}_{ij}\beta_j + \psi_i$

**Assumption 5** For $j = 1, ..., K^*$, $S_j$ is independent and identically distributed with $0 < \Pr\left(S_j = 1\right) \equiv P_s \leq 1$ . $S_j$ is also independent of all other random variables in the model. If $var(\xi) \equiv \sigma_\xi^2 = 0$, then $P_S < 1$.

**Assumption 6** $\xi$ is mean zero and uncorrelated with $Z$ and $W^*$.

(Can redefine $\xi$ so it uncorrelated with $Z$ and $W^*$)

**Theorem 1** *Define $\phi$ and $\phi_\varepsilon$ such that*

$$\mathit{Proj}\left(Z_i \mid X_i, \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j W_{ij}\Gamma_j, \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\left(1 - S_j\right)W_{ij}\Gamma_j + \xi; \mathcal{G}^K\right)$$
$$= X'\phi_X + \phi\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j W_{ij}\Gamma_j\right) + \phi_\varepsilon\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\left(1 - S_j\right)W_{ij}\Gamma_j + \xi_i\right).$$

*Then under assumptions 1-3 and 5-6, if the probability limit of $\phi$ is nonzero, then*

$$\frac{\phi_\varepsilon}{\phi}\ \underset{K^*\to\infty}{\xrightarrow{p}}\ \frac{\left(1 - P_s\right)A}{\left(1 - P_s\right)A + \sigma_\xi^2}$$

*where*

$$A \equiv \lim_{K^*\to\infty} E\left(\frac{1}{K^*}\sum_{j=1}^{K^*}\sigma_{j,j}^{K^*}\left(\Gamma_j\right)^2\right).$$

*If the probability limit of $\phi$ is zero, then the probability limit of $\phi_\varepsilon$ is also zero.*

**Corollary 1** *When* $\sigma_\xi^2 = 0$,

$$plim(\phi - \phi_\varepsilon) = 0.$$

- When $\sigma_\xi^2 = 0$, $W^c = W^*$, so $W$ is a random subset of all of elements of $W^c$.

- This is equality of selection on observed and unobserved variables—condition 1 above.

- Says that the coefficients of the projection of $Z_i$ onto $\frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} S_j W_{ij} \Gamma_j$

  and $\frac{1}{\sqrt{K^*}} \sum_{j=1}^{K^*} \left(1 - S_j\right) W_{ij} \Gamma_j$ approach each other with probability one as $K^*$ becomes large.

**Corollary 2** *When $P_s = 1$,*

$$plim(\phi_\varepsilon) = 0.$$

(OLS case—all variables that potentially affect both $Z$ and $Y$ are included in the model)

The next corollary establishes condition 3

**Corollary 3** *When* $0 < P_s < 1$ *and* $\sigma_\xi^2 > 0$,

*either*

$$0 < \ plim(\phi_\varepsilon) < plim(\phi),$$

*or*

$$plim(\phi) < plim(\phi_\varepsilon) < 0,$$

*or*

$$0 = plim(\phi_\varepsilon) = plim(\phi).$$

Key role in the estimators below.

## 4.3 Systematic Variation in $P_{s_j}$

**Assumption 7**

$$E\left(\mu_j\Gamma_j \mid S_j = 1\right) > E\left(\mu_j\Gamma_j \mid S_j = 0\right) > 0.$$

To make life simple, we also assume

**Assumption 8** $S_j$ is independent of $W_j\Gamma_j$.

(Not necessary)

**Theorem 2** *Define $\phi$ and $\phi_\varepsilon$ as in Theorem 1. Then under assumptions 1-3 and 5-8, as $K^*$ gets large.,*

$$0 < \phi_\varepsilon < \phi$$

The theorem implies $\phi_\varepsilon < \phi$ even when $\sigma_\xi^2 = 0$.

LABOUR Lectures

March 2011

# Lecture 2. Estimation Methods and Applications

Joseph G. Altonji

Yale University

# 5 Outline

- The OU Estimator

- Application of OU to Catholic School Effect, Swan Ganz procedure

- Sensitivity analysis related to the OU Estimator, with applications to Catholic school effect and Swan-Ganz

- brief discussion of heterogenous treatment effects case (very preliminary, will probably skip)

- The OU-Factor Estimator

- Consistency of OU-Factor

- Constructing Confidence Intervals

- Monte Carlo Evidence

- Conclusion

# 6 The OU Estimator

- KEY IDEA: Use $0 \leq \phi_{\varepsilon} \leq \phi$ as an additional restriction on the system of equations for $Y, T$ and $Z$.

- Suppress norming by $\sqrt{K^*}$. Consider the case

$$T = Z = X'\beta_x + W'\beta + u$$

- Problem: $0 \leq \phi_{\varepsilon} \leq \phi$ is not operational unless $E(\varepsilon|W) = 0$ because $\Gamma$ is not identified.

- observed and unobserved determinants of $Y$ are also likely to be correlated given that the $W_{ij}$ typically are correlated.

- AET consider the "reduced form"

$$E\left(\tilde{Y} - \alpha\tilde{T} \mid \widetilde{W}\right) \equiv \widetilde{W}'G \tag{8}$$

$$\tilde{Y} - E\left(\tilde{Y} - \alpha\tilde{T} \mid \widetilde{W}\right) \equiv e. \tag{9}$$

- Let $\phi_{W'G}$ and $\phi_e$ be the coefficients of the projection of $T$ on $W'G$ and $e$ (in a regression model that includes $X$).

**Assumption 9**

$$\frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j \widetilde{W}_{j-\ell}\right) E\left(\beta_j \mathsf{\Gamma}_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j \widetilde{W}_{j-\ell}\right) E\left(\mathsf{\Gamma}_j \mathsf{\Gamma}_{j-\ell}\right)} = \frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\beta_j \mathsf{\Gamma}_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{\widetilde{W}}_j \widetilde{\widetilde{W}}_{j-\ell}\right) E\left(\mathsf{\Gamma}_j \mathsf{\Gamma}_{j-\ell}\right)}, \tag{10}$$

where $\widetilde{\widetilde{W}}_j$ is the component of $W_j$ that is orthogonal to the observed variables $(X, W)$, for all elements of $W^*$.

- Roughly speaking (10) says that the regression of $T$ on $\widetilde{Y}-\alpha\widetilde{T}$ $-\xi$ is equal to the regression of the part of $\widetilde{T}$ that is orthogonal to $\widetilde{W}$ on the corresponding part of $\widetilde{Y}-\alpha\widetilde{T}-\xi$.

**Theorem 3** *Define $\phi_{W'G}$ and $\phi_e$ such that*

$$\textit{Proj}\left(\widetilde{Z}_i \mid \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_j\widetilde{W}_{ij}G_j, \frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\left(1-S_j\right)\widetilde{W}_{ij}\Gamma_j+\xi;\mathcal{G}^K\right)$$
$$=\phi_{W'G}\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*} S_jW_{ij}\Gamma_j\right)+\phi_e\left(\frac{1}{\sqrt{K^*}}\sum_{j=1}^{K^*}\left(1-S_j\right)W_{ij}\Gamma_j+\xi_i\right).$$

*Then under assumptions 1-6 and 9, as $K^*$ gets large, if the probability limit of $\phi$ is nonzero, then*

$$\frac{\phi_e}{\phi_{W'G}}\xrightarrow{p}\frac{\sum_{\ell=-\infty}^{\infty}E\left(\widetilde{W}_j\widetilde{W}_{j-\ell}\right)E\left(\Gamma_j\Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty}E\left(\widetilde{W}_j\widetilde{W}_{j-\ell}\right)E\left(\Gamma_j\Gamma_{j-\ell}\right)+\sigma_\xi^2}..$$

*If the probability limit of $\phi_{W'G}$ is zero then the probability limit of $\phi_e$ is also zero.*

Based on the argument that selection on unobservables is likely to be weaker than selection on observables, impose condition 3

$$\begin{aligned} 0 &\leq \phi_e \leq \phi \text{ if } \phi > 0 \\ 0 &\geq \phi_e \geq \phi \text{ if } \phi < 0 \end{aligned} \tag{11}$$

- OU estimator: work with the system

$$\begin{aligned} Y &= \alpha T + X'\mathsf{\Gamma}_X + W'G + e. \\ T &= X'\beta_X + W'\beta + u \\ 0 &\le \frac{cov(u,e)}{var(e)} \le \frac{Cov(\tilde{W}'\beta, \tilde{W}'G)}{Var(\tilde{W}'G)}. \end{aligned}$$

and estimate the set of $\alpha$ values that satisfy the above inequality restrictions.

- Perform statistical inference accounting to variation over $i$ conditional on which $W$ are observed in the usual way.

- No obvious way to account for random variation due to the draws of $S_j$.

## 6.1 Is Equality of Selection on Observables and Unobservables Enough to Identify $\alpha$?

**Theorem 4** *Suppose that $\varepsilon$ is independent of $W$. Under Condition 1, the true value of $\alpha$ is a root of a cubic polynomial. Thus the identified set contains one, two or three values.*

- Even if $Cov(\varepsilon, W'\Gamma) = 0$, there are typically either three solutions (i.e. three values of $\alpha^*$ that we can not distinguish between) or there is a unique solution that equals $\alpha$.

**Theorem 5** *If we impose the same model as above but use $T$ as an instrument for itself, the true value of $\alpha$ is a root of a quadratic polynomial with two roots:*

$$\begin{aligned} \alpha^* &= \alpha \\ \alpha^* &= \alpha + \frac{var(\varepsilon)}{cov(u,\varepsilon)}. \end{aligned}$$

- Have point identification if the researcher knows the sign of the bias, which is the sign of $cov(u,\varepsilon)$.

- Set $\hat{\alpha}$ to the larger root if believe $cov(u,\varepsilon) > 0$.

- However, equality of selection is unlikely to hold anyway. We focus on bounds

# 7 Applying the OU Estimator

## 7.1 Example 1: The Effect of Catholic Schools

- Consider

$$\begin{aligned} CH_i &= 1(X_i'\beta_X + W'\beta + u > 0) \quad (12) \\ Y_i &= 1(X_i'\Gamma_X + W'G + \alpha CH_i + e > 0) \quad (13) \\ u, e &\sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}). \quad (14) \end{aligned}$$

- In above bivariate probit, our restriction is

$$0 \leq \rho = cov(u, e)/var(e) \leq \frac{Cov(\tilde{W}'\beta, \tilde{W}'G)}{Var(\tilde{W}'G)}. \quad (15)$$

(AET used $W$ rather than $\tilde{W}$ in this restriction)

- Lower bound estimate is MLE value imposing equality of selection:

$$\rho = \frac{Cov(\tilde{W}'\beta, \tilde{W}'G)}{Var(\tilde{W}'G)}$$

- Upper bound: $\hat{\alpha}$ when $\rho = 0$ (essentially univariate probit).

- Can relax normality

## 7.2 Results: (AET (2005a) Table 6

- We use two alternative methods to estimate $G$.

- For Method 1, in the case of High School graduation,

Univariate probit estimate of marginal effect on graduation is 0.08 (.025)

The estimate of

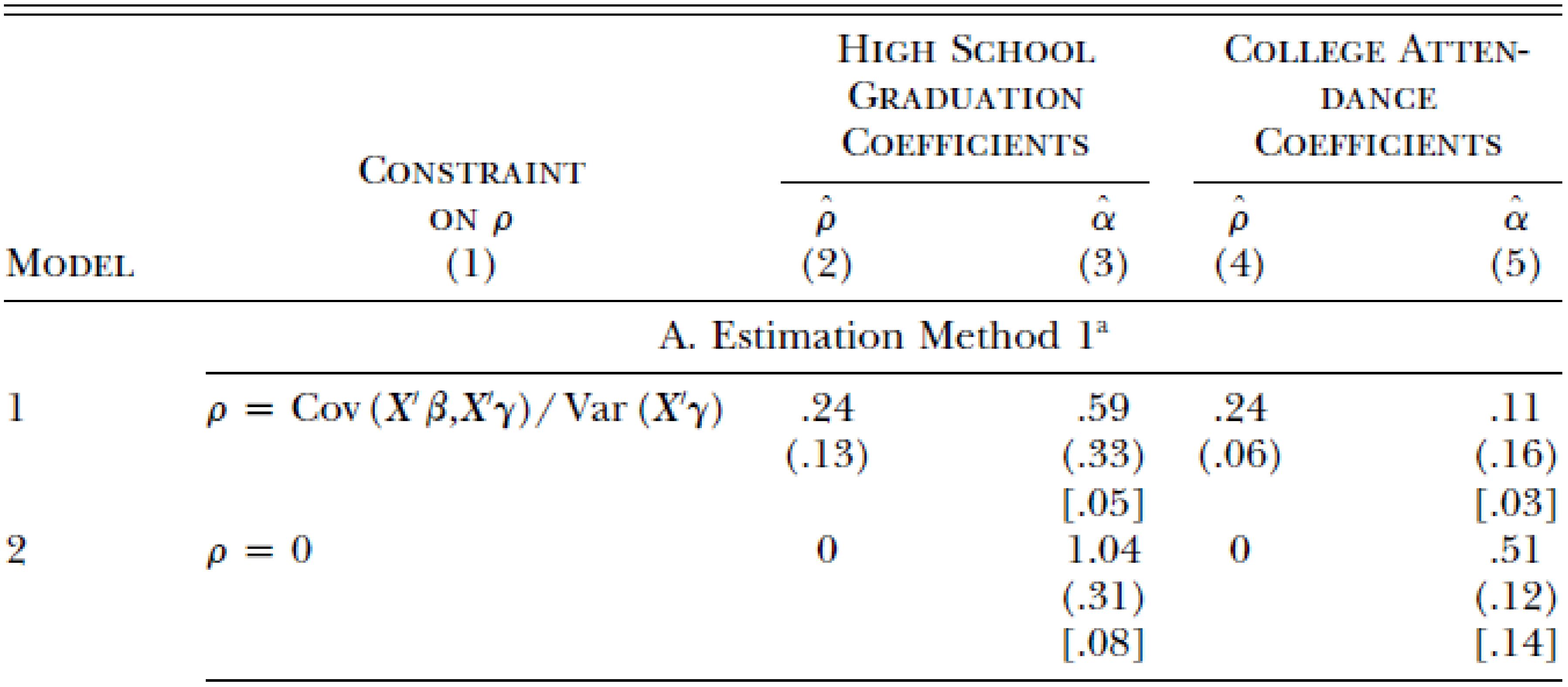
$\rho = cov(u, e)/var(u) = Cov(W_i'\beta, W_i'G)/Var(W'G) =$ 0.24 (0.13)

and the estimate of $\alpha$ falls somewhat.

The effect on graduation. prob.falls from .08 to .05

- For method 2, $\rho$ is only 0.09, and $\alpha$ is 0.94 (0.30)., effect on grad prob is .09

- Consequently, even with the lower bound estimate based on the extreme assumption of equal selection on observables and unobservables imposed, there is evidence for a substantial positive effect of attending Catholic high school on high school graduation.



|  |  | High School Graduation Coefficients |  | College Attendance Coefficients |  |
|---|---|---|---|---|---|
| Model | Constraint on $\rho$<br>(1) | $\hat{\rho}$<br>(2) | $\hat{\alpha}$<br>(3) | $\hat{\rho}$<br>(4) | $\hat{\alpha}$<br>(5) |
|  | A. Estimation Method 1[a] |  |  |  |  |
| 1 | $\rho = \mathrm{Cov}(X'\beta, X'\gamma) / \mathrm{Var}(X'\gamma)$ | .24 | .59 | .24 | .11 |
|  |  | (.13) | (.33) | (.06) | (.16) |
|  |  |  | [.05] |  | [.03] |
| 2 | $\rho = 0$ | 0 | 1.04 | 0 | .51 |
|  |  |  | (.31) |  | (.12) |
|  |  |  | [.08] |  | [.14] |

### 7.2.1 College Attendance:

The results for college attendance follow a similar pattern, but with the extreme assumption imposed most of the effect of $CH$ is gone.

### 7.2.2 Results Robust to Relaxing Normality

## 7.3 Alternative way to use information about selection on the observables

**Condition 4: (Suppress conditioning on** $X$, suppress tildas over the $W$)

$$
\begin{aligned}
& \frac{E(\epsilon_i \mid CH_i = 1) - E(\epsilon_i \mid CH_i = 0)}{Var(\epsilon_i)} \\
= \quad & \frac{E(W_i'G \mid CH_i = 1) - E(W_i'G \mid CH_i = 0)}{Var(W_i'G)}
\end{aligned}
$$

- Says difference by $CH$ in standardized means is the same for the index of observables ($W_i'G$) and the index of unobservables $e_i$. that determine $Y$ is the same.

- This condition is equivalent to $\phi = \phi^e$ .Can justify with random variable selection argument.

Assess evidence for a $CH$ effect by asking how large the ratio on the left side of Condition 4 would have to be relative to the ratio on the right to account for the entire estimate of $\alpha$ under the null hypothesis that $\alpha$ is zero.

- Ignore the fact that $Y$ is estimated by a probit and treat $\alpha$ as if it were estimated by a regression of the latent variable $Y^*$ on $X, W$ and $CH$.

- Let $\widetilde{CH}$ represent the residuals of a regression of $CH$ on $X$ and $W$ so that $CH = X'\beta_X + W'\beta + \widetilde{CH}$. Then,

$$Y^* = \alpha\widetilde{CH} + X'\Gamma_X + W'[G + \alpha\beta] + e.$$

- If the bias in a probit is close to the bias in OLS applied to the above model, then the fact that $\widetilde{CH}$ is orthogonal to $W$ leads to

$$\begin{aligned} \text{plim } \widehat{\alpha} - \alpha &\simeq \frac{cov(\widetilde{CH}, e)}{var\left(\widetilde{CH}\right)} \\ &= \frac{var\left(CH\right)}{var\left(\widetilde{CH}\right)}\left[E(e \mid CH = 1) - E(e \mid CH = 0)\right]. \end{aligned}$$

- Condition 4 allows us to use an estimate of $E(W'G \mid CH = 1) - E(W'G \mid CH = 0)$ to estimate the magnitude of $E(e \mid CH = 1) - E(e \mid CH = 0)$. Plug into the above formula to the bias.

- If $var(e)$ is very large relative to $var(W'G)$,what one can learn is limited, because even a small shift in $(E(e \mid CH = 1) - E(e \mid CH = 0))\,/var(e)$ is consistent with a large bias in $\alpha$.)

- Under the null hypothesis of no $CH$ effect, we can consistently estimate $G$, and thus $E(W'G \mid CH)$, from a separate model imposing $\alpha = 0$.

## 7.4 Results:

- Estimate of $(E(W'G \mid CH = 1) - E(W'G \mid CH = 0)) \, / Var(W'G)$ is 0.24.
  - Mean/variance of the probit index of $X$ variables that determine $HS$ is 0.24 higher for those who attend $CH$ than for those who do not.
  - Variance of $e$ is 1.00, so the implied estimate of $E(e \mid CH = 1) - E(e \mid CH = 0)$ if Condition 4 holds is 0.24
  - Multiplying by $var\left(CH_i\right) / var\left(\widetilde{CH}_i\right)$ yields a bias of 0.29.
  - The unconstrained estimate of $\alpha$ is 1.03
  - The ratio $\widehat{\alpha} / [\frac{var(CH)}{var\left(\widetilde{CH}\right)}(E(e \mid CH = 1) - E(e \mid CH = 0))] = 1.03 \; / \; 0.29 =$ 3.55.
  - So the normalized shift in the distribution of the unobservables would have to be 3.55 times as large as the shift in the observables to explain away the entire $CH$ effect.
  - Seems highly unlikely.

- College attendance: estimated ratio is 1.43

## 7.5 Assessing instrumental variables estimators (AET, 2005b).

- We can use the approach to take another look at the merits of estimate the effect of Catholic school on outcomes using two instrumental variables
  - Catholic religion
  - proximity of a Catholic school
- I focus specifically on the Catholic instrument $(C)$ and the high school graduation outcome $(CH)$.
- For simplicity, leave conditioning on $X$ implicit.
- Define

$$\begin{aligned}
\text{Proj}\,(CH_i \mid W, C_i) &= W_i'\beta + \lambda C_i \\
\widetilde{CH}_i &= \text{Proj}\,(CH_i \mid W_i, C_i) - W_i'\beta - \lambda C_i \\
\text{Proj}\,(C_i \mid X_i) &= W_i'\pi \\
\widetilde{C}_i &= C_i - W_i'\pi
\end{aligned}$$

- We can rewrite the theorem 1 expression

$$\text{Proj}(C|W_i'G, e) = \phi W_i'G + \phi e$$

as

$$\frac{cov(C_i, e_i)}{var(e_i)} = \frac{cov(W_i'\pi, W_i'G)}{var(W_i'G)}$$

We can use this expression to get an expression for the bias one gets from IV

- 2SLS estimate is huge—about .3 Implied bias also turns out to be huge—about .84.

- Bias overstated, because equality of selection almost certainly wrong.

- But conclude $C_i$ is not a good instrument.

- Proximity to a Catholic school looks even worse.

|  | Excluded Instruments |  |  |
|---|---|---|---|
|  | (1)<br>Catholic | (2)<br>Distance | (3)<br>Catholic × Distance |
| HS Graduation |  |  |  |
| 2SLS Coefficient | 0.34 (0.08) | –0.04 (0.10) | 0.09 (0.11) |
| Bias 1 | 0.52 (0.23) | 0.15 (0.16) | 0.14 (0.24) |
| Bias 2 | 0.84 (0.26) | 0.06 (0.14) | — |
| College in 1994 |  |  |  |
| 2SLS coefficient | 0.40 (0.10) | 0.31 (0.11) | –0.11 (0.12) |
| Bias 1 | 0.45 (0.21) | 0.46 (0.22) | 0.15 (0.26) |
| Bias 2 | 0.45 (0.21) | 0.40 (0.20) | — |
| Twelfth reading score |  |  |  |
| 2SLS Coefficient | 1.40 (1.54) | –1.09 (1.84) | 1.24 (1.82) |
| Bias 1 | 1.18 (1.06) | 2.49 (1.59) | 2.59 (1.14) |
| Bias 2 | 1.42 (1.07) | 2.11 (1.40) | — |
| Twelfth math score |  |  |  |
| 2SLS Coefficient | 2.64 (1.21) | 2.43 (1.45) | –2.63 (1.57) |
| Bias 1 | 2.02 (0.75) | 1.76 (1.03) | 1.42 (0.88) |
| Bias 2 | 1.87 (0.74) | 1.72 (0.98) | — |

# 8 Application 2: Does Swan-Ganz Catheterization Help or Hurt Patients

- Does use of Swan-Ganz catheter to monitor intensive care unit (ICU) patients raise mortality?

- Revisit applying methods of Altonji Elder and Taber (2002, 2005, hereafter AET) to data from the leading observational study.

- Our Main Conclusion: The data do not support strong conclusions about Swan-Ganz

## 8.1 Background

- Use of the catheter ($T$) popular in the 70s and 80s. Strong consensus that it was a safe way to monitor patients

- No random trial evaluation—viewed as unethical given strong consensus $T$ is beneficial

- Accumulation of evidence from observational studies suggested no benefit or harm

## 8.2 Prior Work

### 8.2.1 A.F. Connors et. al. (1996)

- use propensity score matching and multivariate models to assess $T$.

- Large sample, rich set of demographic characteristics and health status measures

- Find that $T$ within the first twenty four hours *raises* mortality rates,

- Provide impetus for two large-scale experimental evaluations of the approach that find that $T$ has no effect on mortality in a population that is less sick than Connors et al. (1996).

### 8.2.2 Bhattacharya, Shaikh, and Vytlacil (2007, hereafter, BSV)

- $T$ recipients are sicker on many observed dimensions.
  - propensity score matching ignores selection on unobservables
  - might overstate the negative consequences of $T$.
- BSV apply a set of bounds estimators, including an extension of Shaikh and Vytlacil (2004), that incorporate prior information that weekend admission to the hospital is a valid instrument for $T$.
- Results:
  - Bounds include possibility of a benefit over the first seven days,
  - estimates suggest that $T$ has either no effect or a harmful effect after 30 days.
- Issues:

  - Bounds quite wide
  - Exogeneneity of weekend admission controversial
  - Weekend admission not a very powerful instrument once necessary controls are included

- Interesting to consider alternative approaches, such as AET's $OU$ estimator.

# 9 Data

- From Connors et. al. (1996).

- Medical chart information, data from interviews with patients and proxy respondents.

- Demographic information and private insurance status.

- Outcomes: mortality in seven , 90, and 180 days.

- $T$ patients sicker on most dimensions at baseline

- Mortality rate for $T$ patients is 0.038 higher at seven days, 0.093 at 90 days, and 0.087 at 180 days.

- Connors et al show that controls reduce but do not eliminate differences.

- Is remaining effect due to Selection on Unobservables? BSV motivate their attention to selection on unobservables by noting the systematic pattern in the observables.

## 9.1 The Sensitivity of Probit Estimates of Catheterization to Correlation in Unobservables

- Let $Y = 1$ indicate death within $t$ days.

- Consider the model

$$\begin{aligned} T &= 1(T^* > 0) \equiv 1(W'\beta + u > 0) && (16) \\ Y &= 1(W'G + \alpha T + e > 0) && (17) \\ \begin{bmatrix} u \\ e \end{bmatrix} &\sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), && (18) \end{aligned}$$

- Estimate $\alpha$ under different assumptions about $\alpha$.

- Connors et. al. (1996) present a related calculation

- Results robust to relaxing normality.

- Conclusion: even a modest value of $\rho$ could eliminate the positive (harmful) effect of $T$ on mortality,

- But not clear what range of values of $\rho$ are plausible.

- Next, use the degree of selection on the observables as a guide.

Table 1: Sensitivity of Estimates of Swan-Ganz Treatment Effects to Variation in the Correlation of Disturbances in Bivariate Probit Models

| | Dependent Variable: Mortality in: | | |
|---|---|---|---|
| $\rho$ | 7 days | 90 days | 180 days |
| 0.0 | 0.137 | 0.231 | 0.219 |
| | (0.058) | (0.046) | (0.046) |
| | [0.025] | [0.074] | [0.071] |
| 0.1 | -0.029 | 0.065 | 0.053 |
| | (0.058) | (0.046) | (0.045) |
| | [-0.005] | [0.021] | [0.017] |
| 0.2 | -0.195 | -0.103 | -0.114 |
| | (0.057) | (0.045) | (0.045) |
| | [-0.036] | [-0.033] | [-0.037] |
| 0.3 | -0.363 | -0.270 | -0.282 |
| | (0.056) | (0.045) | (0.044) |
| | [-0.067] | [-0.086] | [-0.092] |

Note: cell entries are estimated Swan-Ganz treatment effects from bivariate probit models restricting the correlation between the disturbances in the treatment and outcome equations to the values given in the column headings. Standard errors are in parentheses and marginal effects are in brackets.

## 9.2 Estimates of the $T$ Effect Using Selection on the Observables to Assess Selection Bias

- Information on medical charts is collected because it is believed to be relevant for assessing health status and guiding treatment.

- Also, future shocks (e.g., infection) that lead to mortality are unknown when $T$ is chosen.

- Thus in Swan-Ganz application, selection on observables is likely to be stronger than selection on unoservables:

$$0 < \phi_e < \phi_{W'G}$$

## 9.3 Implimentation

- In bivariate probit case restrictions on $\phi_e$ correspond to

$$0 \leq \rho \leq \frac{Cov(W'\beta, W'G)}{Var(W'G)}. \tag{1}$$

- Table 2: MLE estimates of $\alpha$ and marginal effect imposing $\rho = \frac{Cov(W'\beta, W'G)}{Var(W'G)}$.

- Standard errors assume that (1) holds for the particular set of $X$ variables that we have.

- Ignores variation that would arise if the set of $X$ variables is too small for such variation to be non-negligible.

Table 2: Estimates of Swan-Ganz Treatment Effects Assuming Equality of Selection on Observable and Unobservable Determinants of Mortality

| | Dependent Variable: Mortality in: | | |
|---|---|---|---|
| Estimate of: | 7 days | 90 days | 180 days |
| $\alpha$ | -0.231 | -0.044 | -0.017 |
| | (0.286) | (0.174) | (0.176) |
| | [-0.042] | [-0.014] | [-0.005] |
| $\rho$ | 0.221 | 0.165 | 0.142 |

- Lower bound estimates are negative. (Shouldn't conclude from the table that $T$ is beneficial)

- Calls into question the strength of the evidence for a harmful effect.

### 9.3.1 Thinking about $\rho$

- AET (2008) distinguish between unobserved (by econometrician) mortality factors that are known and unknown to the doctor at baseline.

- Obtain expression for $\rho$ as product of
  - fraction of unobserved mortality factors that are known to doctors at baseline $\theta$
  - the degree $q$ that $C$ is selected on those factors relative to $\frac{Cov(W'\beta, W'G)}{Var(W'G)}$

- Example: if $\theta = .5$, $q = .7$, in 90 day case

$$\rho = \phi_e = q\frac{Cov(W'\beta, W'G)}{Var(W'G)} \cdot \theta = .7 \cdot 0.165 \cdot 0.5 = 0.0578.$$

- $\theta = .5$ implies Doctor's $R^2 = .655$.

- We lacked the expertise and data to use formula.

# 10 The Relative Amount of Selection on Unobservables Required to Explain the Swan-Ganz Catheter Effect

Consider

$$\frac{E(e \mid T=1) - E(e \mid T=0)}{var(e)} = \lambda \frac{E(W'G \mid T=1) - E(W'G \mid T=0)}{var(W'G)}.$$

- $\lambda$ is the strength of selection on unobservables and relative to selection on observables.

- Under the assumptions leading to $\phi_{W'G} = \phi_e$. $\lambda = 1$.

- How large does $\lambda$ have to be for bias to account for $\hat{\alpha}$ if $\alpha$ is actually zero?

- *Caution:* When $var(e)$ is very large relative to $var(W'G)$, one can't learn much unless one is confident in the choice of $\lambda$

## 10.1 Results: (Table 3)

- In the 90 day case, $\left(E(W'\hat{G} \mid T=1) - E(W'\hat{G} \mid T=0)\right) / Var(W'\hat{G})$ is 0.211,

- Using bias formula presented earlier, this implies 0.211 as an estimate of $E(e \mid T=1)$ $-\ E(e \mid T=0)$ if $\lambda = 1$

- Multiplying by $var\,(T)\,/var\left(\tilde{T}\right)$ yields a bias reported in the table of 0.288 (0.056).

- Unconstrained estimate of $\alpha$ is 0.231 (0.046)

- $\widehat{\alpha}/[\frac{var(T)}{var\left(\widetilde{T}\right)}(E(e \mid T=1) - E(e \mid T=0))] = 0.231\ /\ 0.288$, or 0.801.

- so can attribute the entire positive $T$ effect to bias if the normalized shift with $T$ in the distribution of the unobservables is 0.801 as large as the shift in the observables ($\lambda = 0.801$).

- We suspect true value of $\lambda$ is lower for reasons discussed above.

- At 7 days, the ratio of selection on unobservables relative to selection on observables need only be 0.289 to explain away the positive mortality estimate.

# 11 Heterogenous Treatment Effects

- AET (2002) speculate on extension of consider treatment heterogeneity.

- A threshold crossing model with heterogeneous effects may be written as

$$\begin{aligned} T^* &= W'\beta + u \\ Y_t^* &= W'G_t + e_c \\ Y_{nc}^* &= W'G_{nt} + e_{nt} \\ T &= 1(T^* > 0) \\ Y &= 1(T \cdot Y_t^* + (1 - T)Y_{nt}^* > 0) \end{aligned}$$

- Apart from an intercept shift, we imposed $G_t = G_{nt}$ and $e_t = e_{nt}$.

- Doctors choose $T$ to minimize mortality, so $W'\beta$ is negatively related to $[W'G_t - W'G_{nt} + e_t - e_{nt}]$.

Table 3: The Amount of Selection on Unobservables Relative to Selection on Observables Required to Attribute the Entire S-G Effect to Selection Bias

| | Dependent Variable: Mortality in:. | | |
|---|---|---|---|
| | 7 days | 90 days | 180 days |
| Mean of Outcome | 0.136 | 0.419 | 0.475 |
| Univariate Probit Estimate | 0.137 | 0.231 | 0.219 |
| | (0.058) | (0.046) | (0.046) |
| | [0.025] | [0.074] | [0.071] |
| Implied Bias | 0.475 | 0.288 | 0.288 |
| | (0.111) | (0.056) | (0.056) |
| Ratio of Estimate to Bias | 0.289 | 0.801 | 0.759 |

Notes: a) The entries in the "Univariate Probit Estimate" row are the coefficients from univariate probit models relating mortality to binary indicators of Swan-Ganz catheterization.

b) The entries in the "Implied Bias" row correspond to the implied bias from Condition 4 in the text.

- Conjecture that reasoning and assumptions similar to homogenous case would lead to

$$\begin{aligned}\frac{Cov(W'\beta, W'G_t)}{var(W'G_t)} &= \frac{Cov(u, e_t)}{var(e_t)} \equiv \rho_{ue_t}\\ \frac{Cov(W'\beta, W'G_{nt})}{var(W'G_{nt})} &= \frac{Cov(u, e_{nt})}{var(e_{nt})} \equiv \rho_{ue_{nt}}\\ \frac{Cov(W'G_t, W'G_{nt})}{var(W'G_{nt})} &= \frac{Cov(e_t, e_{nt})}{var(e_{nt})} \equiv \rho_{e_te_{nt}}.\end{aligned}$$

- Given clear evidence that sickest patients receive $T$, one might want to impose

$$\frac{Cov(W'\beta, W'G_t)}{var(W'G_t)} > \frac{Cov(u, e_t)}{var(e_t)} \equiv \rho_{ue_t} > 0$$

- In addition, interactions are very large,

$$\frac{Cov(W'\beta, W'G_{nt})}{var(W'G_{nt})} > \frac{Cov(u, e_{nt})}{var(e_{nt})} \equiv \rho_{ue_{nt}} > 0$$

- $\rho_{e_te_{nt}}$ would have to be estimated or a sensitivity analysis conducted.

- Use these restrictions to help bound estimates of $G_t$ and $G_{nt}$ in a way that is analogous to our use of (1) in the homogeneous effects case?

- To my knowledge, no one has implemented

## 11.1 Conclusions from Swan-Ganz Analysis

- Conners et al. data not conclusive about Swan-Ganz

- Observable-Unobservable Bounds Estimator and Sensitivity Analysis might be usefully applied in epidemeology in situations where strong instruments are lacking, experiements are lacking.

# 12 The OU-Factor Estimator

- A Factor Model of the $W_{ij}$
- The Estimator
- Consistency
- Statistical Inference Based on the Bootstrap
- Monte Carlo Evidence

## 12.1 A Factor Model of $\widetilde{W}_{ij}$

$$\widetilde{W}_{ij} = \frac{1}{\sqrt{K^*}}\widetilde{F}_i^{\prime}\Lambda_j + \upsilon_{ij},\ j = 1, ..., K^* \tag{2}$$

where $\tilde{F}_i$ is an $r$ dimensional vector.

$r$ doesn't grow with the number of $W_{ij}$

$Var(\tilde{F}_i)$ is the identity matrix.

$\sigma_j^2 \equiv E(v_{ij}^2 \mid j)$.

Continue to assume

$$Z_i = X_i' \beta_x + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K} \tilde{W}_{ij} \beta_j + u_i$$

and analogously

$$T_i = X_i' \delta_X + \frac{1}{\sqrt{K^*}} \sum_{j=1}^{K} \tilde{W}_{ij} \delta_j + \omega_i$$

**Assumption 10** (i) $\left(\Gamma_j, \beta_j, \Lambda_j, \sigma_j^2\right)$ is i.i.d with fourth moments; (ii) The components $\xi_i$ and $\psi_i$ of $Y_i$ and $Z_i$ respectively are independent of $W_i^*$ and of each other. (iii) $\xi_i$ is independent of $X_i$.

## 12.2 The OU-Factor Estimator of an Admissible Set for $\alpha$

- Observe $K$ (but not $K^*$) and the joint distribution of $Y_i$, $Z_i$, $T_i$ , $X_i$ and $\left\{W_{ij} : S_{ij} = 1\right\}$

- $K/K^* \longrightarrow P_{s0}$.

- $\frac{K^*}{N} \longrightarrow 0$, so that we can take sequential limits.

- Let $\theta = \{\alpha, \phi, P_s, \sigma_\xi^2)$.

  - Abstract from parameters that are point identified and parameters that are point identified given $\theta$.

- The true value of $\theta$ is $\theta_0 = \{\alpha_0, \phi_0, P_{s0}, \sigma_{\xi 0}^2)$ which lies in the compact set $\bar{\Theta}$.

- We estimate a set $\widehat{\Theta}$ that asymptotically will contain the true value $\theta_0$.

- The key restrictions are

$$0 < P_{s0} \leq 1 \tag{3}$$
$$\sigma^2_{\xi 0} \geq 0. \tag{4}$$

- $P_{s0} = 1$ is the standard IV case

- $\sigma^2_{\xi 0} = 0$ is the "unobservables are like observables" case.

- Estimate the set of values for $\alpha$ by first estimating the set of $\theta$ that satisfy all of the conditions. Then projecting the set onto the $\alpha$ dimension.

- The upper bound and lower bound of the estimated set do not have to occur at $P_{s0} = 1$ and $\sigma^2_{\xi 0} = 0$, but in practice we have found that they do.

### 12.2.1 Stage 1 : Estimate Factor Model $\Lambda_1, .., \Lambda_K$ and $\sigma_1^2, ..., \sigma_K^2$.

- Use sample analogues to the $K$ moment conditions

$$E\left(\widetilde{W}_{ij_1}\widetilde{W}_{ij_2}\right) = \frac{1}{K^*}\Lambda_{j_1}^2 + \sigma_{j_1}^2;\ j_1 = 1, ..., K,\ j_1 = j_2 \tag{5}$$

and the $K \cdot (K-1)/2$ conditions

$$E\left(\widetilde{W}_{ij_1}\widetilde{W}_{ij_2}\right) = \frac{1}{K^*}\Lambda_{j_1}^2\ ;\ j_1,\ j_2 = 1, ..., K,\ j_1 \neq j_2 \tag{6}$$

- Standard GMM problem.

- Let $\widehat{\lambda_j}$ be the GMM estimate of the parameter $\sqrt{K} \times \frac{1}{\sqrt{K^*}}\Lambda_j \approx \sqrt{P_{S0}}\Lambda_j$. $\hat{\lambda}$ is the vector of $\widehat{\lambda_j}$.

### 12.2.2 Stage 2

If we knew $\alpha_0$ we could estimate $\Gamma$ conditional on $\alpha_0$ using moment condition

$$\begin{aligned}\sqrt{K^*}E\left[\widetilde{W}_{ij}\left(\widetilde{Y}_i-\alpha_0\widetilde{T}_i)\right)\right] &= \sqrt{K^*}E\left[\begin{array}{c}\left(\frac{1}{\sqrt{K^*}}\widetilde{F}_i\Lambda_j+v_{ij}\right)\cdot\\ \left(\frac{1}{\sqrt{K^*}}\sum_{\ell=1}^{K^*}\frac{1}{\sqrt{K^*}}\widetilde{F}_i\Lambda_\ell\Gamma_\ell+\frac{1}{\sqrt{K^*}}\sum_{\ell=1}^{K^*}v_{ij}\Gamma_\ell\right)\end{array}\right]\\ &= \Lambda_j\left(\frac{1}{K^*}\sum_{\ell=1}^{K^*}\Lambda_\ell\Gamma_\ell\right)+\sigma_{vj}^2\Gamma_j\\ &\xrightarrow{p} \Lambda_j'E(\Lambda_\ell\Gamma_\ell)+\sigma_{vj}^2\Gamma_j.\end{aligned}$$

- Basically, we are using the factor model to fill in averages of moments involving the missing $W_{ij}$

- Sample analog is

$$\left[\ \sqrt{K^*}\frac{1}{N}\widetilde{W}'\left(\tilde{Y}-\alpha_0\tilde{T}\right)\ \right]=\left[\ \frac{1}{K}\frac{1}{P_{s0}}\hat{\lambda}\hat{\lambda}'\mathbf{\Gamma}+\mathbf{\Sigma}\mathbf{\Gamma}\ \right]$$

- Given $\theta$, can construct the estimator

$$\widehat{\mathbf{\Gamma}}(\theta)\ \approx\left[\frac{1}{P_sK}\hat{\lambda}\hat{\lambda}'+\widehat{\mathbf{\Sigma}}\right]^{-1}\frac{1}{N}\widetilde{W}'\left(\tilde{Y}-\alpha\tilde{T}\right) \tag{7}$$

- $\widehat{\mathbf{\Sigma}}$ is the diagonal matrix of the idiosyncratic variances $\widehat{\sigma_j^2}$ from the factor model of $W$

$$\phi_0=\frac{\left[E(\mathbf{\Gamma}_j\mathbf{\Lambda}_j)E(\beta_j\mathbf{\Lambda}_j)+E(\mathbf{\Gamma}_j\beta_j\sigma_j^2)\right]\left[P_{s0}\left(1-P_{s0}\right)E(\mathbf{\Gamma}_j^2\sigma_j^2)+P_{s0}\sigma_{\xi 0}^2\right]}{\sigma_{\xi 0}^2\left[P_{s0}^2E(\mathbf{\Gamma}_j\mathbf{\Lambda}_j)^2+P_{s0}E(\mathbf{\Gamma}_j^2\sigma_j^2)\right]+\left[E(\mathbf{\Gamma}_j\mathbf{\Lambda}_j)^2+E(\mathbf{\Gamma}_j^2\sigma_j^2)\right]\left(1-P_{s0}\right)P_{s0}E(\mathbf{\Gamma}}$$

Using this fact, we define our estimator of $\theta$ based on the following system of equations.

$$
\begin{aligned}
q^{1}_{N,K^*}(\theta) =& \frac{1}{N}\sum_{i=1}^{N} \widetilde{W}_i' \widehat{\boldsymbol{\Gamma}}(\theta) \times \qquad (8)\\
& \left[ \widetilde{Z}_i - \phi \widetilde{W}_i' \widehat{\boldsymbol{\Gamma}}(\theta) - \phi \frac{(1-P_s)\widehat{\boldsymbol{\Gamma}}(\theta)' \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\Gamma}}(\theta)}{(1-P_s)\widehat{\boldsymbol{\Gamma}}(\theta) \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\Gamma}}(\theta) + P_s \sigma_\xi^2} \left( \widetilde{Y}_i - \alpha \widetilde{T}_i - \widetilde{W}_i' \widehat{\boldsymbol{\Gamma}}(\theta) \right) \right] \\
q^{2}_{N,K^*}(\theta) =& \frac{1}{N}\sum_{i=1}^{N} \left( \left( \widetilde{Y}_i - \alpha \widetilde{T}_i - \widetilde{W}_i' \widehat{\boldsymbol{\Gamma}}(\theta) \right) \right) \times \qquad (9)\\
& \left[ \widetilde{Z}_i - \phi \widetilde{W}_i' \widehat{\boldsymbol{\Gamma}}(\theta) - \phi \frac{(1-P_s)\widehat{\boldsymbol{\Gamma}}(\theta)' \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\Gamma}}(\theta)}{(1-P_s)\widehat{\boldsymbol{\Gamma}}(\theta)' \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\Gamma}}(\theta) + P_s \sigma_\xi^2} \left( \widetilde{Y}_i - \alpha \widetilde{T}_i - \widetilde{W}_i' \widehat{\boldsymbol{\Gamma}}(\theta) \right) \right] \\
q^{3}_{N.K^*}(\theta) =& \frac{1}{N}\sum_{i=1}^{N} \left( \widetilde{Y}_i - \alpha \widetilde{T}_i \right)^2 - \left( \frac{\widehat{\boldsymbol{\Gamma}}(\theta)' \widehat{\lambda}}{P_s} \right)^2 - \frac{\widehat{\boldsymbol{\Gamma}}(\theta)' \widehat{\boldsymbol{\Sigma}} \widehat{\boldsymbol{\Gamma}}(\theta)}{P_s} - \sigma_\xi^2 \qquad (10)
\end{aligned}
$$

subject to $\theta \in \bar{\boldsymbol{\Theta}}$.

- At $\theta = \theta_0$, right hand sides of these equations converge to zero as $N$ and $K^*$ grow.

### 12.2.3 Intuition for first two equations:

When $\sigma_{\xi}^{2}=0$ they reduce to

$$
\begin{aligned}
q_{N,K^{*}}^{1}\left(\theta\right) &= \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{W}_{i}^{\prime}\widehat{\Gamma}\left(\theta\right)\left[\tilde{Z}_{i}-\phi\widetilde{W}_{i}^{\prime}\widehat{\Gamma}\left(\theta\right)-\phi\left(\tilde{Y}_{i}-\alpha\tilde{T}_{i}-\widetilde{W}_{i}^{\prime}\widehat{\Gamma}\left(\theta\right)\right)\right]\right) \\
q_{N,K^{*}}^{2}\left(\theta\right) &= \frac{1}{N}\sum_{i=1}^{N}\left(\left(\tilde{Y}_{i}-\alpha\tilde{T}_{i}-\widetilde{W}_{i}^{\prime}\widehat{\Gamma}\left(\theta\right)\right)\left[\tilde{Z}_{i}-\phi\widetilde{W}_{i}^{\prime}\widehat{\Gamma}\left(\theta\right)-\phi\left(\tilde{Y}_{i}-\alpha\tilde{T}_{i}-\widetilde{W}_{i}^{\prime}\widehat{\Gamma}\left(\theta\right)\right)\right]\right.
\end{aligned}
$$

These are the classic moment conditions of a regression of $\tilde{Z}_{i}$ on $(\widetilde{W}_{i}^{\prime}\widehat{\Gamma}\left(\theta\right))$ and $(\tilde{Y}_{i}-\alpha\tilde{T}_{i}-\widetilde{W}_{i}^{\prime}\widehat{\Gamma}\left(\theta\right))$ when the regression coefficients are restricted to be the same.

Empirical analog of Corollary 1 of Theorem 1.

In the general case the error term $\xi$ leads to attenuation bias.

- When $P_S = 1$, the second equation is

$$q^2_{N,K^*}(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\tilde{Y}_i - \alpha\tilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)\right)\left[\tilde{Z}_i - \phi\widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)\right]\right)$$

In this case $\widehat{\Gamma}\left(\theta\right)$ could be estimated as the coefficient of a regression of $\tilde{Y}_i - \alpha\tilde{T}_i$ on $\widetilde{W}_i$.

- In $P_S = 1$ case $\widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)$ would have to be orthogonal to the error term, so equation is the standard IV moment condition:

$$q^2_N(\alpha,\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{Y}_i - \alpha\tilde{T}_i - \widetilde{W}_i'\widehat{\Gamma}\left(\theta\right)\right) \times Z_i$$

- $q^3_{N.K^*}\left(\theta\right)$ is the difference between the sample value of $var\left(\tilde{Y}_i - \alpha\tilde{T}_i\right)$ for the hypothesized value of $\alpha$ and the variance implied by the model estimate.

The estimator $\widehat{\Theta}$ is the set of values of $\theta$ that minimize the criterion function

$$Q_{N,K^*}(\theta) = q_{N,K^*}(\theta)'\Omega q_{N,K^*}(\theta)$$

where

$$q_{N,K^*}(\theta) = \left[\begin{array}{ccc} q^1_{N,K^*}(\theta) & q^2_{N,K^*}(\theta) & q^3_{N,K^*}(\theta) \end{array}\right]'$$

and $\Omega$ is some predetermined positive definite weighting matrix.

## 12.3 Consistency of the Estimator

- Prove consistency using the standard methods from Chernozhukov, Hong, and Tamer (2007).

- Define $Q_0(\theta)$ as the probability limit of $Q_{N,K^*}(\theta)$ as $N$ and $K^*$ get large. Sequential limits assuming that $N$ grows faster than $K^*$.

- The identified set, $\Theta_I$, is defined as the set of values that minimize $Q_0(\theta)$.

- We verify the conditions in Chernozhukov, Hong, and Tamer (2007) to show that the Hausdorff distance between $\widehat{\Theta}$ and $\Theta_I$ converges in probability to zero and that $\theta_0 \in \Theta_I$. Thus as the sample gets large our estimate of $\widehat{\Theta}$ will contain the true value with probability approaching 1.

**Assumption 11** $F_i, \xi_i$, and $\psi_i$ are all mean 0 and i.i.d. across individuals and are independent of each other with finite second moments. $\omega_i$ is i.i.d. across individuals with finite second moments, is independent of $F_i$, but may be correlated with $\xi_i$ and/or $\psi_i.v_{ij}$ is mean zero and i.i.d. across individuals and covariates with finite variance. The vector $(\Gamma_j, \Lambda_j, \beta_j, \delta_j, \sigma_j^2)$ is i.i.d. across covariates with finite second moments.

**Assumption 12** $\bar{\Theta}$ is compact with the support of $P_s$ bounded below by $p_s^\ell > 0$.

**Assumption 13** The dimension of $F_i$ is 1

Let $d_h(\cdot,\cdot)$ to be Hausdorff distance as defined in Chernozhukov, Hong, and Tamer (2007).

**Theorem 6** *Under Assumptions 11-13, $d_h(\widehat{\Theta}, \Theta_I)$ converges in probability to zero and $\theta_0 \in \Theta_I$.*

The set estimator for $\alpha_0$ is the projection of $\widehat{\Theta}$ onto $\alpha$.

$$\widehat{A} \equiv \left\{ \alpha : \text{ there exists some value of } (\phi, P_s, \sigma_\xi^2) \text{ such that } \{\alpha, \phi, P_s, \sigma_\xi^2\} \in \widehat{\Theta} \right\}$$

## 12.4 Constructing Confidence Intervals

### 12.4.1 The General Approach

- Construct confidence set for $(\alpha_0, \phi_0, P_S^0, \sigma_\xi^0)$ by “inverting a test statistic.” The confidence set for $\alpha$ is the set of values of $\alpha$ in that set.

- We construct a test statistic $T(\theta)$ with known distribution under the null: $\theta = \theta_0$.

- For each potential $\theta$, construct an acceptance region of the test.

- Let $T_{N,K^*}(\theta)$ be the estimated value of the test statistic and let $T^c(\theta)$ the critical value. Confidence set is defined as

  $$\widehat{C}_{N,K^*} = \left\{\theta \in \Theta \mid \widehat{T}(\theta) \leq T^c(\theta)\right\},$$

  Confidence region for $\alpha$ can be written as

  $$\widehat{C}_\alpha = \left\{\alpha \in \mathbb{R} \mid (\alpha, \Theta) \cap \widehat{C}_N \neq \emptyset\right\}.$$

### 12.4.2 Algorithm based on the Bootstrap

- Consider testing the null hypothesis $\theta = \theta_0$.

We use normalized criteria function so that

$$T_{N,K^*}(\theta) = K \cdot Q_{N,K^*}(\theta)$$

1. Estimate parameters to be used in generating data for bootstrap.

   From the joint distribution of $(X_i, W_i)$ estimate

   (a) $\Sigma, \Lambda, \Lambda_X$, and data generating processes for $F_i$ and $v_{ij}$

   (b) Estimate

$$\frac{\widehat{\Gamma}(\theta)}{\sqrt{K^*}} \equiv \left[\frac{1}{P_s K}\widehat{\lambda}'\widehat{\lambda} + \widehat{\Sigma}\right]^{-1} \frac{1}{N}\widetilde{W}'\left(\tilde{Y} - \alpha\tilde{T}\right)$$
$$\frac{\widehat{\beta}(\theta)}{\sqrt{K^*}} \equiv \left[\frac{1}{P_s K}\widehat{\lambda}'\widehat{\lambda} + \widehat{\Sigma}\right]^{-1} \frac{1}{N}\widetilde{W}'\tilde{Z}$$

   (c) Given knowledge of $P_S$ estimate the distribution of $(\xi_i, \psi_i, \omega_i)$

2. Generate $N_B$ bootstrap samples. For each sample:

   (a) Draw $K$ observable covariates from the actual set of covariates (with replacement) with appropriate $\left(\widehat{\mathbf{\Gamma}_j}, \widehat{\beta}_j, \widehat{\lambda}_j, \widehat{\mathbf{\Sigma}}_{jj}\right)$

   (b) Draw $(K^* - K)$ unobservable covariates from the actual set of covariates (with replacement) with appropriate $\left(\widehat{\mathbf{\Gamma}_j}, \widehat{\beta}_j, \widehat{\lambda}_j, \widehat{\mathbf{\Sigma}}_{jj}\right)$

   (c) For $i = 1, ..., N$ generate $(X_i, W_i^*)$ using DGP for $F_i$ and $v_{ij}$.

   (d) Using DGP for $\psi_i$ and $\xi_i$ generate $Z_i$ and $(Y_i - \alpha_0 T_i)$

   (e) Given generated bootstrap data construct the test statistic $Q_{N,K^*}(\theta)$. (involves the intermediate steps of estimating $\mathbf{\Sigma}, \lambda$ and $\mathbf{\Gamma}$as well.)

3. From the bootstrap sample, estimate the distribution of the test statistic and calculate the critical value given the size of the test.

   - To reduce computation burden, combine simulations of $T_{N,K^*}(\theta)$ for grid of values of $\theta$ and estimate conditional quantile function corresponding to desired confidence level.

- We conjecture the bootstrap distribution of $T_{N,K^*}(\theta_0)$ provides a consistent estimate of the actual distribution of $T_{N,K^*}(\theta_0)$. (Proof is in progress.)

## The Distribution of $T_{N,K^*}(\theta_0)$

$$\chi_j \equiv \Big[\ \Lambda_j\Gamma_j \quad \Lambda_j\beta_j \quad \Gamma_j\sigma_j^2\Gamma_j \quad \Gamma_j\sigma_j^2\beta_j \quad S_j\frac{\Lambda_j^2}{\sigma_j^2} \quad S_j\Gamma_j\Lambda_j \quad S_j\Gamma_j\Lambda_j\sigma_j^2 \quad S_j\beta_j\Lambda_j \quad S_j\beta_j\Lambda_j\sigma_j^2 \quad S$$

- The limit of $Q_{N,K^*}(\theta_0)$ as $N$ gets large turns out to be a known function of only $\theta$ and $E\left(\chi_j\right)$.

### 12.4.3 A Simplified Parametric Boot Strap Procedure

- Testing the null over a four dimensional grid is computationally very demanding.

- In simulations, we consistently find a compact region:

  - one end of the region at $(P_S = 1$

  - The other end at the "observable like unobservable restriction" $(\sigma_\xi = 0)$.

- Assume positive selection bias so that the upper bound occurs under the constraint $P_S = 1$ and minimum value occur at $\sigma_\xi$.

- parametric bootstrap procedure to construct a one sided confidence interval estimators for $\alpha_{\text{min}}$ and $\alpha_{\text{max}}$.

- $\hat{\alpha}_{.10\,\text{min}}$ has 10% probability of being below $\alpha_{\text{min}}$.

- $\hat{\alpha}_{.10,\text{max}}$ has a 10% nominal probability of exceeding $\alpha_{\text{max}}$.

**Sketch of Simplied Boot Strap to construct** $\hat{\alpha}_{.10\,\mathrm{min}}$ 1. Fit distributions that do not constrain second and fourth moments to the random components that determine the $W$ components, including the common factors $\theta$ and the idiosyncratic components $v_{ij}$

2. Sample with replacement $\hat{K}^*$ values from the $K$ $\hat{\Gamma}_j$ , $\hat{\lambda}_j$, $\hat{\sigma}_v$, $\hat{\beta}_j$ and the distributions. Treat the first $K$ as corresponding to the observables.

3. Generate $\hat{K}^*$ 1 x N vectors $W_j$ using the draws of $\hat{\lambda}_j$, $\hat{\sigma}_v$, $\hat{\beta}_j$, etc

4. Given $W^*$, and estimate of $\alpha$ and $P_s$ when $\hat{\sigma}_\xi^2 = 0$, generate $Y$, $T$, and $Z$.

5. Estimate $\hat{\alpha}$ with $\hat{\sigma}_\xi^2 = 0$

6. Repeat lots of times.

# 13 Monte Carlo Evidence

- One Factor Case. $Z = T$.

- The base specification is random assignment in which case we should obtain tight bounds at the true value.

- Lets first check if we find tight bounds around the truth.

- OLS $\hat{\alpha}$:
  - 10th percentile: 0.9863
  - Median: 1.002
  - 90th Percentile: 1.0171
- OU $\hat{\alpha}_{\text{min}}$:
  - 10th percentile: 0.9806
  - Median: 1.0062
  - 90th Percentile: 1.0211
- OU-Factor $\hat{\alpha}_{\text{min}}$
  - 10th percentile: 0.9811
  - Median: 0.9941
  - 90th Percentile: 1.0103

## 13.1 Additional Monte Carlo Cases

- We studied models with covariates that have a factor structure and a nonzero covariance between $\beta_j W_j$ and $\Gamma_j W_j$ .

- Bounds depend on the design, but in many cases, $OU$ and $OU - Factor$ seems to be informative.

- First, the medians of $\widehat{\alpha}_{\text{min}}$ and $\widehat{\alpha}_{OU}$ are close to 1 when the assumption of equality of selection on observed and unobserved variables is correct ($R^2_\xi = 0$).
  - Relative performance of $\widehat{\alpha}_{\text{min}}$ and $\widehat{\alpha}_{OU}$ depends upon the specifics of the experiment, particularly the strength of the factor structure, but overall the two perform similarly.
  - The sampling variances are narrower when the factor structure is stronger, i.e., when $E[Corr(W_{ij}, W_{ij'})] = 0.2$.

- Second, both $\widehat{\alpha}_{\text{min}}$ and $\widehat{\alpha}_{OU}$ typically lie below the value of $\alpha_0$ when $\phi > \phi_\varepsilon$. This is to be expected, because both estimators are based on the assumption that $\phi = \phi_\varepsilon$ and are to be interpreted as lower bound estimators if $\phi > \phi_\varepsilon > 0$ ( in the case $\phi > 0$).

- Third, the gap between the lower bound estimators and $\alpha_0$ declines with $P_S$, which is also to be expected.

- Fourth, the $\widehat{\alpha}_{\text{min}}$ and $\widehat{\alpha}_{OU}$ estimators are usually less precise than $\alpha_{OLS}$ is.

  - The loss of precision depends on the design and is negligible in the case in which $T$ is randomly assigned (as in Table 1).

  - For some designs, such as some of the cases with a strong factor structure, the sampling variance of $\widehat{\alpha}_{\text{min}}$ is actually smaller than that of $\hat{\alpha}_{OLS}$.

- Overall, the distribution of $\widehat{\alpha}_{\text{min}}$ and $\widehat{\alpha}_{OU}$ are sufficiently precise to provide useful information about $\alpha$ in all of the cases that we consider.

- We have not estimated confidence sets using the general procedure yet.

- Preliminary monte carlo evidence assuming $\hat{\alpha}_{\text{min}}$ occurs at $\hat{\sigma}^2_{\xi} = 0$ using simplified parametric bootstrap produces confidence interval estimates with close to nominal values when equality of selection holds.

- lower value is below true value more than specified nominal probability when $\sigma_{\xi}^2 > 0$ , as it should be.

# 14 Conclusions and Caveats

- Systematically examining pattern of selection based on a rich set of observables is helpful in bounding estimates, assessing potential for bias, assessing IV strategies.

- Only beginning. We think of $OU$ and $OU - Factor$ as a start for investigation into a broader class of estimators based on the idea that if one has some prior information about how the observed variables were arrived at, then the joint distribution of the outcome, the treatment variable, the instrument, and the observed explanatory variables are informative about the distribution of the unobservables.

- The basic idea of using observables to say something about unobservables can be extended to other models and one can try alternative assumptions. Factor model is just one approach.

- heterogenous treatment effects

- Warning: potential for misuse of the idea of using observables to draw inferences about selection bias.
  - Dangerous to infer too much about selection on the unobservables from selection on the observables if
    - observables are small in number and explanatory power,
    - they are unlikely to be representative of the full range of factors that determine an outcome.

- Problem in studies that informally examine correlation between $T$ or $Z$ and a small set of covariates