GMM and Treatment Effects

Enrique Pinzón

StataCorp LP

October 22, 2015
Madrid

Introduction

- Challenges solved by treatment effects estimators
- Estimation methodology behind treatment effects. The Generalized Method of Moments (GMM)
- Research idea that involves the use of GMM.

Treatment Effects

We are interested in the outcomes of receiving a treatment in scenarios were researchers have observational data.

For instance:

- The impact on public education outcomes for schools that received a transfer and those that did not.
- Employment outcomes for individuals that participated in a job training program and those that did not.
- The effect on birth weight for babies of mothers that smoked relative to those of mothers that did not.

Missing Values

bweight	mbsmoke	bweight0	bweight1
3119	nonsmoker	3119	.
3515	nonsmoker	3515	.
3147	nonsmoker	3147	.
4026	nonsmoker	4026	.
4366	nonsmoker	4366	.
3544	nonsmoker	3544	.
3500	smoker	.	3500
3289	smoker	.	3289
3430	smoker	.	3430
3147	smoker	.	3147
2778	smoker	.	2778
3884	smoker	.	3884

How We Approach Treatment Effects

- We cannot observe individuals in both states simultaneously
  - Design a random experiment
  - We cannot do this because of technical or ethical concerns
- We need to account for covariates that are correlated with the treatment
- We will think of the problem in terms of models that govern the treatment result and the outcome

How We Approach Treatment Effects

- We cannot observe individuals in both states simultaneously
  - Design a random experiment
  - We cannot do this because of technical or ethical concerns
- We need to account for covariates that are correlated with the treatment
- We will think of the problem in terms of models that govern the treatment result and the outcome

How We Approach Treatment Effects

- We cannot observe individuals in both states simultaneously
  - Design a random experiment
  - We cannot do this because of technical or ethical concerns
- We need to account for covariates that are correlated with the treatment
- We will think of the problem in terms of models that govern the treatment result and the outcome

How We Approach Treatment Effects

- We cannot observe individuals in both states simultaneously
  - Design a random experiment
  - We cannot do this because of technical or ethical concerns
- We need to account for covariates that are correlated with the treatment
- We will think of the problem in terms of models that govern the treatment result and the outcome

Notation and Definitions

- The potential outcome is denoted by the random variable y_τ with $\tau \in \{0, 1, \ldots, K\}$.
- Usually people think about the binary case where there are only two levels y_0 and y_1
- Potential outcome mean

$$POM = E(y_\tau)$$

- Average treatment effect

$$ATE = E(y_k - y_0)$$

- Average treatment effect on the treated

$$ATET = E(y_k - y_0 | \tau = k)$$

- From now on we will focus on binary treatments. All results are valid for multivariate treatments unless explicitly noted.

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$
\begin{aligned}
y_0 &= x\beta_0 + \varepsilon_0 \\
y_1 &= x\beta_1 + \varepsilon_1 \\
y &= \tau y_1 + (1-\tau)\, y_0
\end{aligned}
$$

TREATMENT MODEL:

$$
\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}
$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$
\begin{aligned}
y_0 &= x\beta_0 + \varepsilon_0 \\
y_1 &= x\beta_1 + \varepsilon_1 \\
y &= \tau y_1 + (1-\tau)\, y_0
\end{aligned}
$$

TREATMENT MODEL:

$$
\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}
$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

General Framework Illustrated with a Linear Example

OUTCOME MODEL:

$$\begin{aligned} y_0 &= x\beta_0 + \varepsilon_0 \\ y_1 &= x\beta_1 + \varepsilon_1 \\ y &= \tau y_1 + (1-\tau)\, y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

- w refers to the covariates that determine the treatment
- y_0 and y_1 are not observed. Only y, x, w, and τ are observed
- The random disturbances η, ε_0, and ε_1 are independent
- The functional forms for the outcome model do not need to be linear
- All the estimators we will see arise from combinations of the outcome model and the treatment model

Treatment Effects in Stata

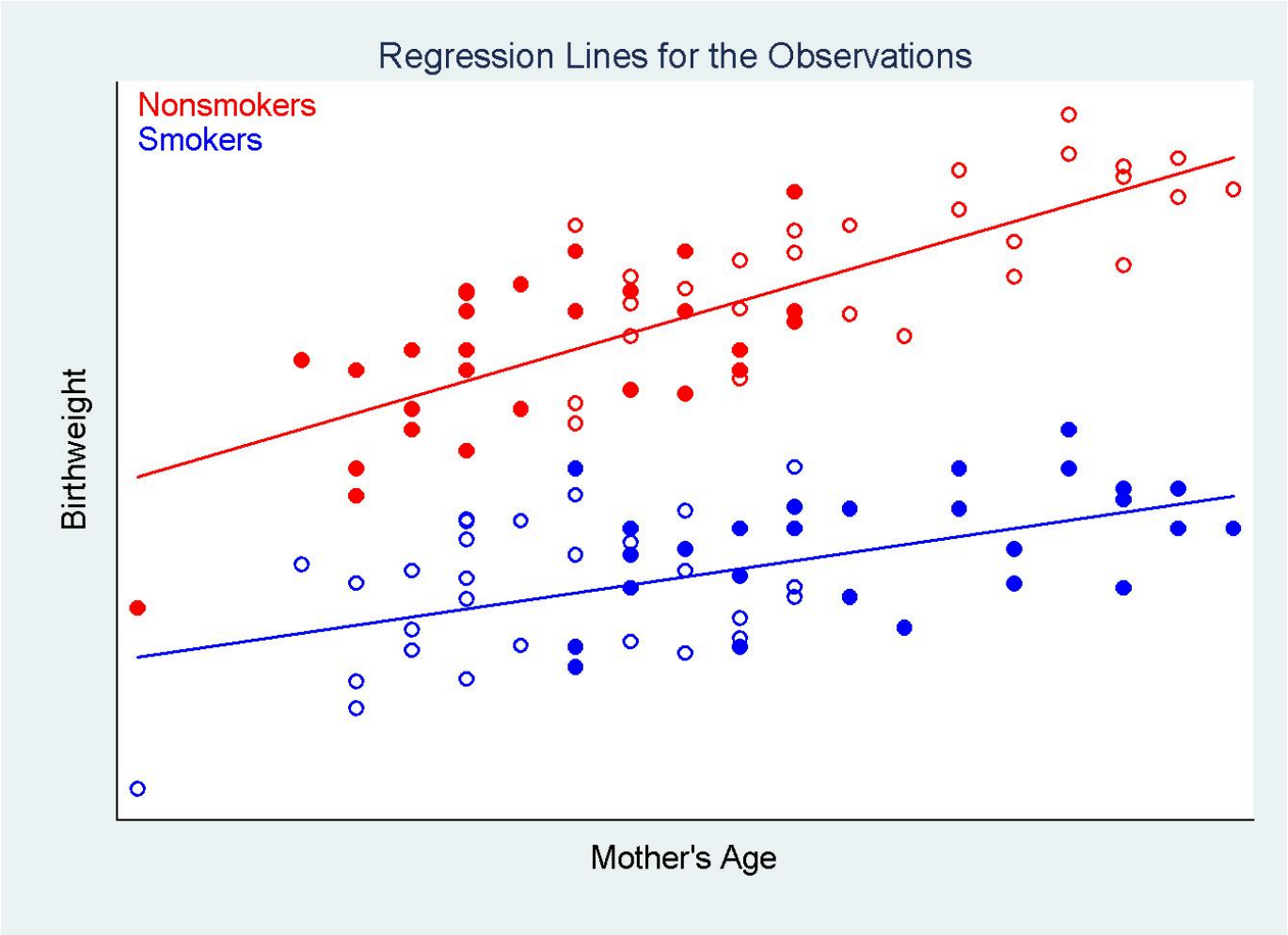
STATA 13

- Regression Adjustment (RA)
- Inverse Probability Weighting (IPW)
- Augmented Inverse Probability Weighting (AIPW)
- Inverse Probability Weighted Regression Adjustment (IPWRA)
- Nearest Neighbor Matching
- Propensity Score Matching

STATA 14

- Endogenous Treatment Effects (Control Function)
- Survival Outcome Treatment Effects
- Balancing

Graphical Representation of RA Estimation



Missing Values Solved

bweight	mbsmoke	bweight0	bweight1	hatbw1	hatbw0
3572	nonsmoker	3572	.	3179.558	3499.417
3289	nonsmoker	3289	.	3194.299	3494.323
3430	nonsmoker	3430	.	3120.591	3519.791
3119	nonsmoker	3119	.	3153.674	3244.215
3374	nonsmoker	3374	.	3147.075	3437.555
3760	nonsmoker	3760	.	3128.563	3179.807
2722	smoker	.	2722	3112.234	3315.544
3402	smoker	.	3402	3225.37	3353.492
3289	smoker	.	3289	3142.703	3512.151
2580	smoker	.	2580	3020.785	3274.057
3714	smoker	.	3714	3135.933	3177.26
3175	smoker	.	3175	3006.043	3279.151

Data from Cattaneo (2010) Journal of Econometrics

`bweight`:	infant birth weight (grams)
`lbweight`:	1 if low birthweight baby
`mbsmoke`:	1 if mother smoked
`prenatal`:	trimester of first prenatal care visit
`fbaby`:	1if first baby
`mmarried`:	1 if mother married
`mage`:	mother's age
`fage`:	father's age
`alcohol`:	1 if alcohol consumed during pregnancy

- Sample of newborns from the United States from 1997

Data from Cattaneo (2010) Journal of Econometrics

`bweight`:	infant birth weight (grams)
`lbweight`:	1 if low birthweight baby
`mbsmoke`:	1 if mother smoked
`prenatal`:	trimester of first prenatal care visit
`fbaby`:	1if first baby
`mmarried`:	1 if mother married
`mage`:	mother's age
`fage`:	father's age
`alcohol`:	1 if alcohol consumed during pregnancy

- Sample of newborns from the United States from 1997

RA Linear Outcome Average Treatment Effect (ATE)

```
. teffects ra (bweight prenatal1 mmarried mage fbaby) (mbsmoke)
Iteration 0:   EE criterion =  7.734e-24
Iteration 1:   EE criterion =  1.196e-25
Treatment-effects estimation                    Number of obs      =      4642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
mbsmoke (smoker vs nonsmoker)	-239.6392	23.82402	-10.06	0.000	-286.3334	-192.945
POmean						
mbsmoke nonsmoker	3403.242	9.525207	357.29	0.000	3384.573	3421.911

RA Average Treatment Effect on the Treated (ATET)

```
. teffects ra (bweight prenatal1 mmarried mage fbaby) (mbsmoke), atet
Iteration 0:   EE criterion =  7.629e-24
Iteration 1:   EE criterion =  2.697e-26
Treatment-effects estimation                    Number of obs      =      4642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATET						
mbsmoke (smoker vs nonsmoker)	-223.3017	22.7422	-9.82	0.000	-267.8755	-178.7278
POmean						
mbsmoke nonsmoker	3360.961	12.75749	263.45	0.000	3335.957	3385.966

Inverse Probability Weighting (IPW)

- In contrast to RA estimators, IPW estimate models for the treatment
- We fit a model for the treatment and compute the probabilities of treatment
- We then compute a weighted average, using the inverse of the probability of being in each group.

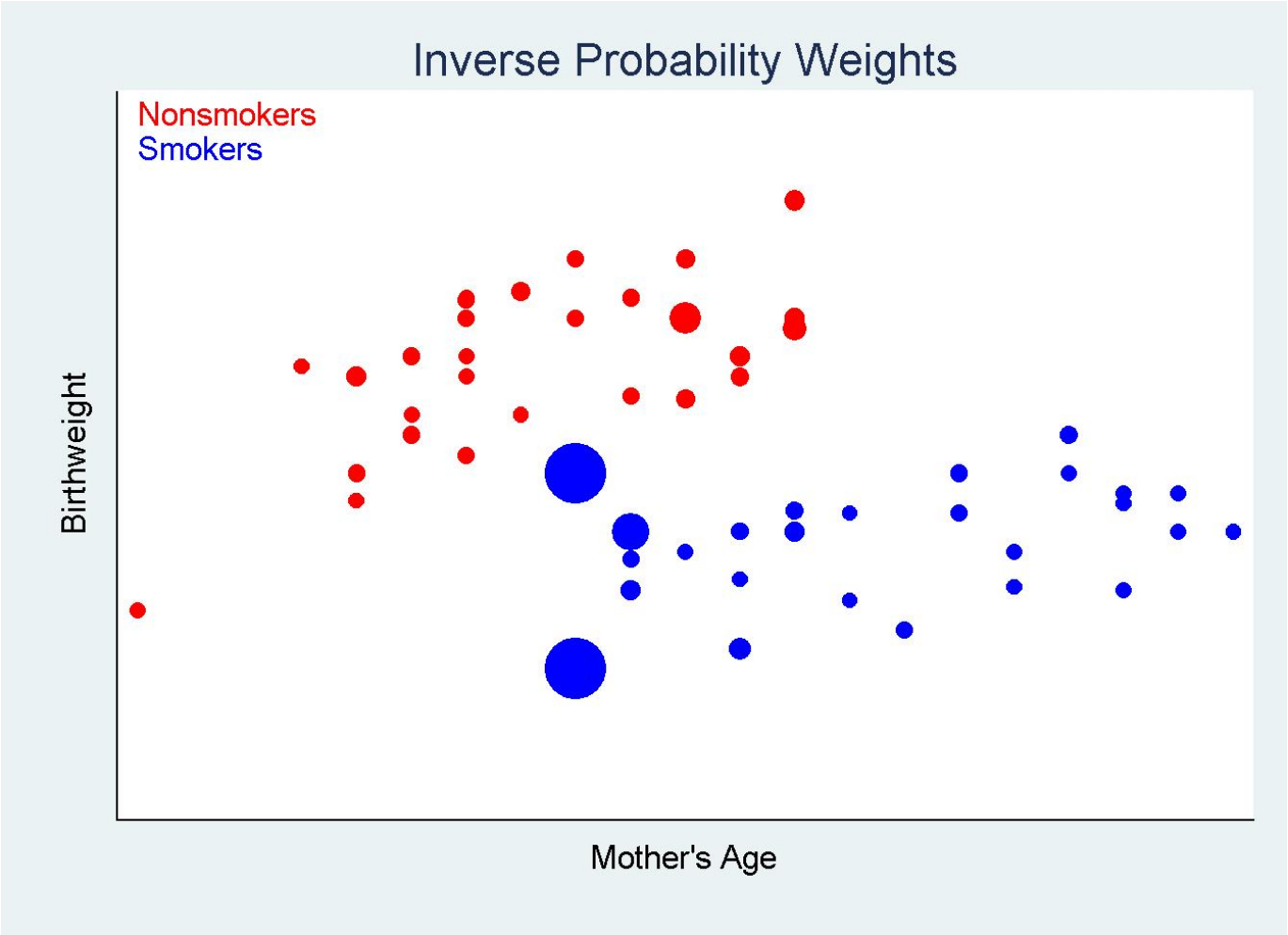
Inverse Probability Weight Calculation

```
. logistic mbsmoke mmarried alcohol mage fedu
Logistic regression                                     Number of obs   =         60
                                                        LR chi2(4)      =      46.50
                                                        Prob > chi2     =     0.0000
Log likelihood = -18.339432                             Pseudo R2       =     0.5590

------------------------------------------------------------------------------
     mbsmoke | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    mmarried |   .0785086   .0909212    -2.20   0.028     .0081122    .7597976
     alcohol |   18.81727   27.98003     1.97   0.048     1.020649    346.9259
        mage |   2.147569    .459327     3.57   0.000      1.41218    3.265909
        fedu |   .8189843   .1157528    -1.41   0.158     .6208252    1.080393
       _cons |   4.46e-07   2.12e-06    -3.07   0.002     3.96e-11    .0050329
------------------------------------------------------------------------------
```

```
. predict ps
(option pr assumed; Pr(mbsmoke))
. replace ps = 1/ps if mbsmoke==1
(30 real changes made)
. replace ps = 1/(1-ps) if mbsmoke==0
(30 real changes made)
```

Inverse Probability Weighting Graphically



Inverse Probability Weighting Estimation

```
. teffects ipw (bweight) (mbsmoke mmarried c.mage##c.mage fbaby medu)
Iteration 0:   EE criterion =  1.713e-21
Iteration 1:   EE criterion =  4.794e-27
Treatment-effects estimation                    Number of obs      =      4642
Estimator      : inverse probability weighted
Outcome model  : weighted mean
Treatment model: logit
```

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
mbsmoke (smoker vs nonsmoker)	-231.7203	25.17975	-9.20	0.000	-281.0717	-182.3689
POmean						
mbsmoke nonsmoker	3403.527	9.576358	355.41	0.000	3384.757	3422.296

Double Robust Estimators Inverse Probability Weighted Regression Adjustment (IPWRA)

- Estimate a treatment model and compute inverse-probability weights
- Use the estimated inverse-probability weights and fit weighted regression models of the outcome for each treatment level
- Compute the means of the treatment-specific predicted outcomes

ATET for Inverse Probability Weighted Regression Adjustment

```
. teffects ipwra (bweight prenatal1 mmarried mage fbaby) ///
> (mbsmoke mmarried c.mage##c.mage fbaby medu), atet
Iteration 0:   EE criterion =  4.620e-21
Iteration 1:   EE criterion =  1.345e-26
Treatment-effects estimation                    Number of obs      =      4642
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: logit
```

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATET						
mbsmoke (smoker vs nonsmoker)	-224.0108	23.846	-9.39	0.000	-270.7481	-177.2735
POmean						
mbsmoke nonsmoker	3361.671	14.54939	231.05	0.000	3333.154	3390.187

Challenge of observational data

1. Can only observe individuals when they are treated or controlled (Missing Data)
2. Control and treatment groups are different (Balancing)

CONTROLLED EXPERIMENT (STAR)

- Student/Teacher Achievement Ratio (STAR)
- Tennessee experiment 11,598 kindergarten students (4 cohorts)
- Treatment was small classroom and outcomes were test scores
- Done in the 1985-1986 school year

Balancing (GENDER)

Table: Balancing Statistics for STAR: GENDER

Statistic	Estimate	Robust S.E.	C.I.
Mean Treated	0.4857	0.0148	[0.4568, 0.5148]
Mean Control	0.4863	0.0085	[0.4696, 0.5030]
Difference	-0.0005	0.0191	[-0.038, 0.0369]
Ratio Variances	0.9994	0.02747	[0.946, 1.053]

Balancing (WHITE)

Table: Balancing Statistics for STAR: WHITE

Statistic	Estimate	Robust S.E.	C.I.
Mean Treated	0.681	0.017	[0.648, 0.714]
Mean Control	0.664	0.009	[0.646, 0.682]
Difference	0.017	0.023	[-0.28, 0.061]
Ratio Variances	0.974	0.033	[0.909, 1.040]

Balancing for Observational Data

- Health data on newborn babies in the United States (subset of Cattaneo 2010)
- Treatment is smoking during pregnancy
- Outcome of interest is low birthweight or birthweight

Balancing (Mother's Education)

Table: Balancing Statistics for : MOTHER'S EDUCATION

Statistic	Estimate	Robust S.E.	C.I.
Mean Treated	11.639	0.364	[10.924, 12.353]
Mean Control	12.923	0.099	[12.734, 13.125]
Difference	-1.291	0.456	[-2.184, -0.397]
Ratio Variances	0.731	0.095	[0.545, 0.917]

Balancing (Number of Prenatal Visits)

Table: Balancing Statistics for : Number of Prenatal Visits

Statistic	Estimate	Robust S.E.	C.I.
Mean Treated	9.862	0.335	[9.206, 10.518]
Mean Control	10.962	0.096	[10.774, 11.151]
Difference	-1.101	0.409	[-1.904, -0.298]
Ratio Variances	1.428	0.133	[1.169, 1.689]

Before Balancing

```
. teffects ipwra                                                        ///
> (bweight mage medu i.mrace i.mmarried i.prenatal1 i.frace fage fedu) ///
> (mbsmoke i.mrace mage i.mmarried i.alcohol i.prenatal1 i.fbaby       ///
> c.mage#(c.mage i.mmarried i.prenatal1))
Iteration 0:   EE criterion =  8.397e-20
Iteration 1:   EE criterion =  2.070e-26
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : IPW regression adjustment
Outcome model  : linear
Treatment model: logit
```

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
mbsmoke (smoker vs nonsmoker)	-236.0222	29.55747	-7.99	0.000	-293.9537	-178.0906
POmean						
mbsmoke nonsmoker	3406.35	9.580925	355.53	0.000	3387.572	3425.129

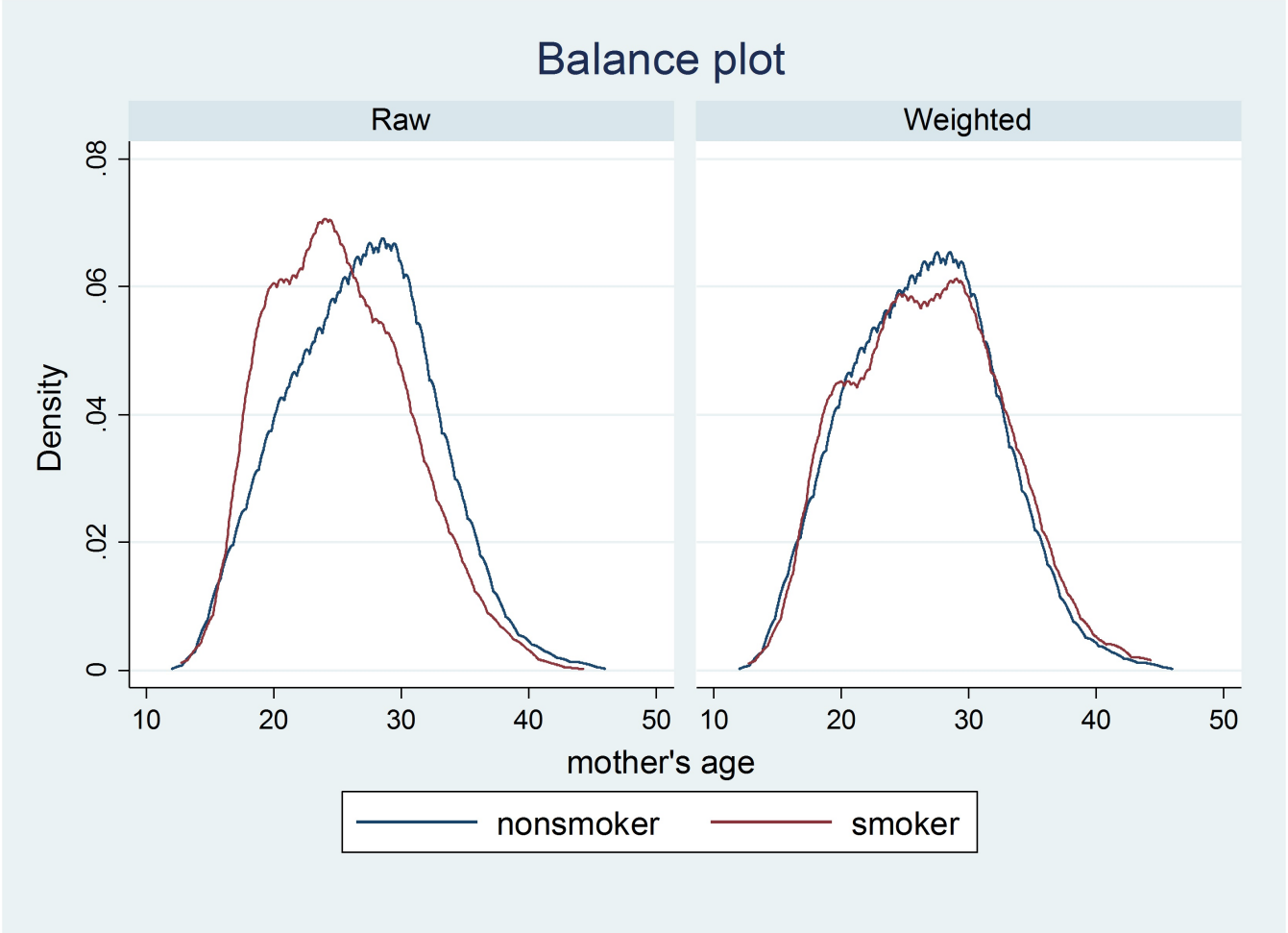
Diagnostic: Summary Statistics

```
. tebalance summarize
Covariate balance summary
```

	Raw	Weighted
Number of obs =	4,642	4,642.0
Treated obs =	864	2,325.1
Control obs =	3,778	2,316.9

	Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted
mrace				
1	-.1029446	.0045151	1.198452	.9915814
mage	-.300179	.0392884	.8818025	1.068054
mmarried				
married	-.5953009	.0120047	1.335944	.9894832
alcohol				
1	.3222725	.0001366	4.509207	1.000699
prenatal1				
Yes	-.3242695	.0032052	1.496155	.9951704
fbaby				
Yes	-.1663271	-.0012409	.9430944	.9996803
mage#				
mage	-.3028275	.0451331	.8274389	1.097548
mmarried#				
mage				
married	-.6329701	.0228823	1.157026	1.024093
prenatal1#				
mage				
Yes	-.4053969	.0180678	1.226363	1.026796

Diagnostic: Graphical



Testing

```
 . tebalance overid
Iteration 0:   criterion =  .09613375
Iteration 1:   criterion =  .09621125  (backed up)
Iteration 2:   criterion =  .09631507  (backed up)
Iteration 3:   criterion =  .09640924  (backed up)
Iteration 4:   criterion =  .09641617  (backed up)
Iteration 5:   criterion =  .09730774
Iteration 6:   criterion =  .09845507
Iteration 7:   criterion =  .09974706
Iteration 8:   criterion =  .10042566
Iteration 9:   criterion =  .10053085
Iteration 10:  criterion =  .10073921
Iteration 11:  criterion =   .1009058
Iteration 12:  criterion =  .10090754
Iteration 13:  criterion =  .10091067
Iteration 14:  criterion =  .10091127
Iteration 15:  criterion =  .10091128
Overidentification test for covariate balance
         H0: Covariates are balanced:
         chi2(10)     =   8.9097
         Prob > chi2  =   0.5407
```

Endogeneity in Treatment Effects

- Greatest challenges to establish causal relationships
- There is an unobservable (omitted) component that affects treatment assignment and outcome.
  - Tax reform on investment (tax evasion)
  - Increased gun ownership on violent crime (preference for non-violent crime an improvement in police effectiveness)
  - Smoking on birth weight (healthy life style)
- This is also true for controlled experiments
  - STAR experiment attrition and parental pressure

Endogeneity in Treatment Effects

- Greatest challenges to establish causal relationships
- There is an unobservable (omitted) component that affects treatment assignment and outcome.
  - Tax reform on investment (tax evasion)
  - Increased gun ownership on violent crime (preference for non-violent crime an improvement in police effectiveness)
  - Smoking on birth weight (healthy life style)
- This is also true for controlled experiments
  - STAR experiment attrition and parental pressure

Endogeneity in Treatment Effects

- Greatest challenges to establish causal relationships
- There is an unobservable (omitted) component that affects treatment assignment and outcome.
  - Tax reform on investment (tax evasion)
  - Increased gun ownership on violent crime (preference for non-violent crime an improvement in police effectiveness)
  - Smoking on birth weight (healthy life style)
- This is also true for controlled experiments
  - STAR experiment attrition and parental pressure

Endogeneity in Treatment Effects

- Greatest challenges to establish causal relationships
- There is an unobservable (omitted) component that affects treatment assignment and outcome.
  - Tax reform on investment (tax evasion)
  - Increased gun ownership on violent crime (preference for non-violent crime an improvement in police effectiveness)
  - Smoking on birth weight (healthy life style)
- This is also true for controlled experiments
  - STAR experiment attrition and parental pressure

Endogeneity

OUTCOME MODEL:

$$\begin{aligned} y_0 &= g(x\beta_0 + \varepsilon_0) \\ y_1 &= g(x\beta_1 + \varepsilon_1) \\ y &= \tau y_1 + (1-\tau) y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

ENDOGENEITY:

$$\begin{aligned} E(\varepsilon_0|\eta) &\neq 0 \\ E(\varepsilon_1|\eta) &\neq 0 \end{aligned}$$

Endogeneity

OUTCOME MODEL:

$$\begin{aligned} y_0 &= g(x\beta_0 + \varepsilon_0) \\ y_1 &= g(x\beta_1 + \varepsilon_1) \\ y &= \tau y_1 + (1-\tau) y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

ENDOGENEITY:

$$\begin{aligned} E(\varepsilon_0|\eta) &\neq 0 \\ E(\varepsilon_1|\eta) &\neq 0 \end{aligned}$$

Endogeneity

OUTCOME MODEL:

$$\begin{aligned} y_0 &= g(x\beta_0 + \varepsilon_0) \\ y_1 &= g(x\beta_1 + \varepsilon_1) \\ y &= \tau y_1 + (1-\tau) y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

ENDOGENEITY:

$$\begin{aligned} E(\varepsilon_0|\eta) &\neq 0 \\ E(\varepsilon_1|\eta) &\neq 0 \end{aligned}$$

Endogeneity

OUTCOME MODEL:

$$\begin{aligned} y_0 &= g(x\beta_0 + \varepsilon_0) \\ y_1 &= g(x\beta_1 + \varepsilon_1) \\ y &= \tau y_1 + (1-\tau) y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

ENDOGENEITY:

$$\begin{aligned} E(\varepsilon_0|\eta) &\neq 0 \\ E(\varepsilon_1|\eta) &\neq 0 \end{aligned}$$

Endogeneity

OUTCOME MODEL:

$$\begin{aligned} y_0 &= g(x\beta_0 + \varepsilon_0) \\ y_1 &= g(x\beta_1 + \varepsilon_1) \\ y &= \tau y_1 + (1-\tau) y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

ENDOGENEITY:

$$\begin{aligned} E(\varepsilon_0|\eta) &\neq 0 \\ E(\varepsilon_1|\eta) &\neq 0 \end{aligned}$$

Endogeneity

OUTCOME MODEL:

$$\begin{aligned} y_0 &= g(x\beta_0 + \varepsilon_0) \\ y_1 &= g(x\beta_1 + \varepsilon_1) \\ y &= \tau y_1 + (1-\tau) y_0 \end{aligned}$$

TREATMENT MODEL:

$$\tau = \begin{cases} 1 & \text{if} \quad w\gamma + \eta > 0 \\ 0 & \text{otherwise} \end{cases}$$

ENDOGENEITY:

$$\begin{aligned} E(\varepsilon_0|\eta) &\neq 0 \\ E(\varepsilon_1|\eta) &\neq 0 \end{aligned}$$

Solution: Control Function Approach

- Obtain residuals from treatment model $\hat{\eta}$
- Incorporate them into the estimation of outcome model.
- Estimate:

$$
\begin{aligned}
E\left(y_0|x,w,\eta\right) &= g\left(x\beta_0+\theta_0\eta\right)\\
E\left(y_1|x,w,\eta\right) &= g\left(x\beta_1+\theta_1\eta\right)
\end{aligned}
$$

Estimation

```
. eteffects                                                                  ///
> (bweight mage medu i.mrace i.mmarried i.prenatal1 i.frace fage fedu)      ///
> (mbsmoke i.mrace mage i.mmarried i.alcohol i.prenatal1 i.fbaby           ///
> c.mage#(c.mage i.mmarried i.prenatal1))
Iteration 0:   EE criterion =  8.584e-23
Iteration 1:   EE criterion =  6.151e-26
Endogenous treatment-effects estimation          Number of obs      =      4,642
Outcome model  : linear
Treatment model: probit
```

bweight	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ATE						
mbsmoke (smoker vs nonsmoker)	-347.454	136.1922	-2.55	0.011	-614.3858	-80.52218
POmean						
mbsmoke nonsmoker	3325.45	28.89296	115.10	0.000	3268.821	3382.079

Testing for Endogeneity

```
. estat endogenous
 Test of endogeneity
 Ho: treatment and outcome unobservables are uncorrelated
          chi2(  2) =    10.25
        Prob > chi2 =     0.0059
```

Generalized Method of Moments (GMM)

- All our treatment effects estimators use GMM
- We use GMM mainly to calculate standard errors, we know the true value of the parameters.
- We also use GMM for our overidentification test.

GMM Basics

- GMM solves systems of equations that have the form:

$$E\left[g\left(x,\theta\right)\right]=0$$

- A very wide range of problems can be written in this form
- The function $g(.)$ is known
- The expected value is replaced using a sample average

A Simple Example

- We have a random variable W with $E(W) = \mu$ and $Var(W) = 3\mu$
- We are interested in one parameter μ and we have two moment conditions

$$\begin{aligned} E\left(W - \mu\right) &= 0 \\ E\left[(1/3)\left(W - \mu\right)^2 - \mu\right] &= 0 \end{aligned}$$

- We know the solutions to this two equations
- We have more equations than unknowns so the solution is not unique
- We have a nonlinear function of the parameters
- GMM weights the two solutions, giving a higher weight to the more efficient solution.

A Simple Example

- We have a random variable W with $E(W) = \mu$ and $Var(W) = 3\mu$
- We are interested in one parameter μ and we have two moment conditions

$$\begin{aligned} E\left(W - \mu\right) &= 0 \\ E\left[(1/3)\left(W - \mu\right)^2 - \mu\right] &= 0 \end{aligned}$$

- We know the solutions to this two equations
- We have more equations than unknowns so the solution is not unique
- We have a nonlinear function of the parameters
- GMM weights the two solutions, giving a higher weight to the more efficient solution.

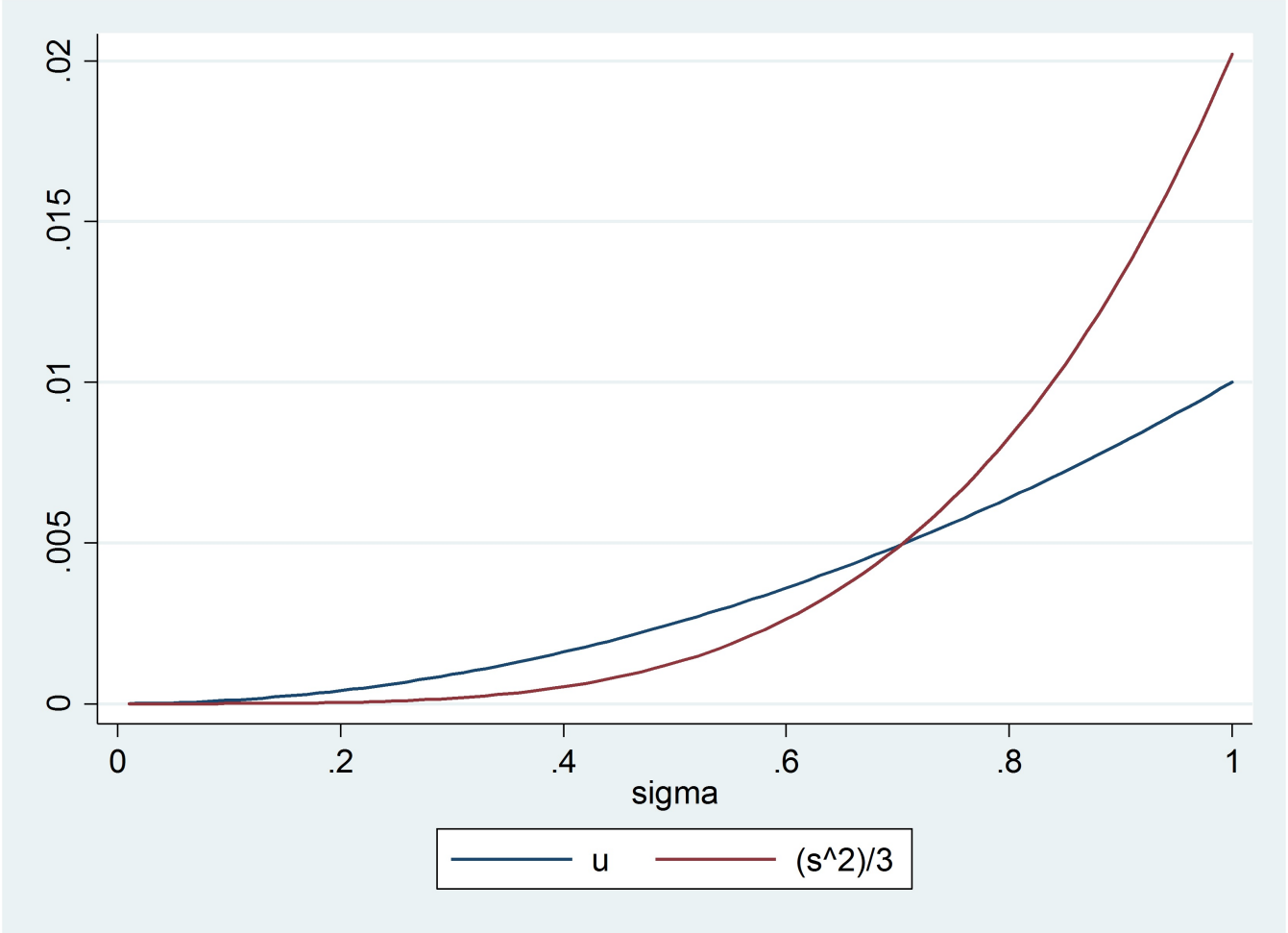
A Simple Example

- We have a random variable W with $E(W) = \mu$ and $Var(W) = 3\mu$
- We are interested in one parameter μ and we have two moment conditions

$$\begin{aligned} E\left(W - \mu\right) &= 0 \\ E\left[(1/3)\left(W - \mu\right)^2 - \mu\right] &= 0 \end{aligned}$$

- We know the solutions to this two equations
- We have more equations than unknowns so the solution is not unique
- We have a nonlinear function of the parameters
- GMM weights the two solutions, giving a higher weight to the more efficient solution.

Efficiency



DGP and Mean Estimate

```
. generate y  = rnormal(3,3)
. summarize y
```

Variable	Obs	Mean	Std. Dev.	Min	Max
y	1,000	2.929767	3.003027	-7.028034	12.90658

Estimation

```
. gmm (eq1: y - {a})(eq2: (1/3)*(y - {a})^2 - {a}), ///
> instruments(eq1 eq2:) winitial(unadjusted, independent) twostep nolog
Final GMM criterion Q(b) =  .0001936
GMM estimation
Number of parameters =   1
Number of moments    =   2
Initial weight matrix: Unadjusted                 Number of obs   =      1,000
GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/a	2.954494	.0766991	38.52	0.000	2.804167	3.104822

```
Instruments for equation eq1: _cons
Instruments for equation eq2: _cons
```

Weight Matrix

```
. mat list e(W)
symmetric e(W)[2,2]
                 eq1:        eq2:
               _cons       _cons
eq1:_cons   .11113597
eq2:_cons   .00293237    .0558026
```

Standard Errors

- GMM estimates the system of equations simultaneously
- It gets correct standard errors where a multistep estimator would get them incorrectly

Solution: Control Function Approach

- Obtain residuals from treatment model $\hat{\eta}$
- Incorporate them into the estimation of outcome model.
- Estimate:

$$
\begin{aligned}
E\left(y_0|x,w,\eta\right) &= g\left(x\beta_0+\theta_0\eta\right)\\
E\left(y_1|x,w,\eta\right) &= g\left(x\beta_1+\theta_\eta\eta\right)
\end{aligned}
$$

How

$$\min_{\theta,W} \left[\bar{g}\left(x,\theta\right)\right]' W \left[\bar{g}\left(x,\theta\right)\right]$$

$$\bar{g}\left(x,\theta\right) \equiv N^{-1} \sum_{i=1}^{N} g_i\left(x_i,\theta\right)$$

If we have K parameters and L equations with $L = K$

$$\left[\bar{g}\left(x,\widehat{\theta}\right)\right]' \widehat{W} \left[\bar{g}\left(x,\widehat{\theta}\right)\right] = 0$$

If $L > K$ we know how far from zero we are and:

$$\left[\bar{g}\left(x,\widehat{\theta}\right)\right]' \widehat{W} \left[\bar{g}\left(x,\widehat{\theta}\right)\right] \xrightarrow{d} \chi^2_{L-K}$$

We can use this to test:

$$\mathrm{H_o} : E\left[g\left(x,\theta\right)\right] = 0$$

Overid Test

- We have the score equations
- Probability of being treated $p(x_i, \theta)$
- T_i indicator for being in treatment group
- Weighted observations $\frac{T_i}{p(x_i,\theta)}$ and $\frac{1-T_i}{1-p(x_i,\theta)}$
- It follows that:

$$E\left\{\frac{T_i}{p(x_i,\theta)}x_i - \frac{1-T_i}{1-p(x_i,\theta)}x_i\right\} = 0$$

$$E\left\{\frac{T_i}{p(x_i,\theta)}g(x_i) - \frac{1-T_i}{1-p(x_i,\theta)}g(x_i)\right\} = 0$$

Overid Test

- We have the score equations
- Probability of being treated $p(x_i, \theta)$
- T_i indicator for being in treatment group
- Weighted observations $\frac{T_i}{p(x_i,\theta)}$ and $\frac{1-T_i}{1-p(x_i,\theta)}$
- It follows that:

$$E\left\{\frac{T_i}{p(x_i,\theta)}x_i - \frac{1-T_i}{1-p(x_i,\theta)}x_i\right\} = 0$$

$$E\left\{\frac{T_i}{p(x_i,\theta)}g(x_i) - \frac{1-T_i}{1-p(x_i,\theta)}g(x_i)\right\} = 0$$

Overid Test

- We have the score equations
- Probability of being treated $p(x_i, \theta)$
- T_i indicator for being in treatment group
- Weighted observations $\frac{T_i}{p(x_i,\theta)}$ and $\frac{1-T_i}{1-p(x_i,\theta)}$
- It follows that:

$$E\left\{\frac{T_i}{p(x_i,\theta)}x_i - \frac{1-T_i}{1-p(x_i,\theta)}x_i\right\} = 0$$

$$E\left\{\frac{T_i}{p(x_i,\theta)}g(x_i) - \frac{1-T_i}{1-p(x_i,\theta)}g(x_i)\right\} = 0$$

Overid Test

- We have the score equations
- Probability of being treated $p(x_i, \theta)$
- T_i indicator for being in treatment group
- Weighted observations $\frac{T_i}{p(x_i,\theta)}$ and $\frac{1-T_i}{1-p(x_i,\theta)}$
- It follows that:

$$E\left\{\frac{T_i}{p(x_i,\theta)}x_i - \frac{1-T_i}{1-p(x_i,\theta)}x_i\right\} = 0$$

$$E\left\{\frac{T_i}{p(x_i,\theta)}g(x_i) - \frac{1-T_i}{1-p(x_i,\theta)}g(x_i)\right\} = 0$$

Overid Test

- We have the score equations
- Probability of being treated $p(x_i, \theta)$
- T_i indicator for being in treatment group
- Weighted observations $\frac{T_i}{p(x_i,\theta)}$ and $\frac{1-T_i}{1-p(x_i,\theta)}$
- It follows that:

$$E\left\{\frac{T_i}{p(x_i,\theta)}x_i - \frac{1-T_i}{1-p(x_i,\theta)}x_i\right\} = 0$$

$$E\left\{\frac{T_i}{p(x_i,\theta)}g(x_i) - \frac{1-T_i}{1-p(x_i,\theta)}g(x_i)\right\} = 0$$

Test of Strict Exogeneity

$$
\begin{aligned}
y_{it} &= x_{it}\beta + \alpha_i + \varepsilon_{it} && (1)\\
E(\alpha_i | x_{it}) &= E(\varepsilon_{it} | x_{it}) = 0 && (2)\\
E(\alpha_i | x_{i1}, \ldots, x_{iT}) &= 0 && (3)\\
E(\varepsilon_{it} | x_{i1}, \ldots, x_{iT}, \alpha_i) &= 0 && (4)
\end{aligned}
$$

- (3) and (4) represent strict exogeneity
- Ordinary least squares (OLS) is consistent under (1) an (2) Random Effects (RE) is NOT
- (OLS) and Random Effects (RE) are consistent under (3) and (4)

Overid Test

The moment conditions of OLS and Random Effects are given by:

$$
\begin{aligned}
E\left[\left(y_{it} - x_{it}\beta\right) x_{it}\right] &= 0 \quad \text{OLS} \\
E\left[\left(\tilde{y}_{it} - \tilde{x}_{it}\beta\right) x_{it}\right] &= 0 \quad \text{RE}
\end{aligned}
$$

- Say β has dimensions $K \times 1$
- We have $2K$ moment conditions and K overidentifying restrictions (OVERID TEST)
- We could also use GMM with $2K$ moments assuming parameters from OLS and RE are different. Then test $\beta_{RE} = \beta_{OLS}$.

Weight Matrix

```
 . reover y x1 x2, nolog
Final GMM criterion Q(b) =  .0009141
GMM estimation
Number of parameters =   8
Number of moments    =  11
Initial weight matrix: Unadjusted                 Number of obs   =       3000
GMM weight matrix:     Cluster (id)
                                  (Std. Err. adjusted for 1,000 clusters in id)
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ols_re						
x1	.9660433	.0189498	50.98	0.000	.9289023	1.003184
x2	.9762426	.0215371	45.33	0.000	.9340307	1.018454
_cons	1.041375	.0361566	28.80	0.000	.9705095	1.112241
within						
x1	.966068	.0189482	50.98	0.000	.9289302	1.003206
x2	.9761178	.0215073	45.39	0.000	.9339641	1.018271
between						
x1	.9663328	.0189535	50.98	0.000	.9291846	1.003481
x2	.9766673	.0215971	45.22	0.000	.9343377	1.018997
_cons	1.041571	.0361565	28.81	0.000	.9707059	1.112437

```
. estat overid
  Test of overidentifying restriction:
  Hansen´s J chi2(2) = 2.74223 (p = 0.2538)
```

Monte Carlo

$$
\begin{aligned}
T &= 3 \\
N &= 1000 \\
e_{it} &\sim \frac{(\chi^2_3 - 3)}{\sqrt{6}} \\
u_i &= \frac{(\chi^2_5 - 5)}{\sqrt{10}} \\
\epsilon_j &\sim N(0,1) \quad j \in \{1,2\} \\
x_{1it} &= \epsilon_1 + \mathbb{I}(t > 1) * (e_{it}\rho) \\
x_{2it} &= \epsilon_2 + \mathbb{I}(t > 1) * (e_{it}\rho) \\
\rho &\in \{0, .05, 0.1, 0.15, 0.2\} \\
y_{it} &= 1 + x_{1it} + x_{2it} + u_i + e_{it}
\end{aligned}
$$

- Violations of strict exogeneity occur when $\rho > 0$ in the equations above
- Results for $\rho = 0$ asses the size of the tests
- Levels of $\rho > 0$ are used to study the power of the tests
- We use 2000 repetitions for our simulations

Results

Strict Exogeneity Parameter	Test Statistic	Rejection Rate
$\rho = 0$	overid (Sargan)	0.044
	exact id (Wald)	0.046
$\rho = 0.05$	overid (Sargan)	0.146
	exact id (Wald)	0.148
$\rho = 0.10$	overid (Sargan)	0.455
	exact id (Wald)	0.457
$\rho = 0.15$	overid (Sargan)	0.799
	exact id (Wald)	0.798
$\rho = 0.20$	overid (Sargan)	0.965
	exact id (Wald)	0.964

Conclusion

- I talked about the problems inherent in obtaining treatment effects and tools in Stata to address them
- I explained the workings of GMM and how it is used in our estimations
- I showed how that intuition can be extended to answer a relevant research question.