Testing for Essential Heterogeneity

James Heckman
University of Chicago
American Bar Foundation
University College Dublin

Daniel Schmierer
University of Chicago

Sergio Urzua
Northwestern University

This draft, July 3, 2007

Objectives

- This paper examines the properties of instrumental variables (IV) applied to models with essential heterogeneity.

Objectives

- This paper examines the properties of instrumental variables (IV) applied to models with essential heterogeneity.
- We present several empirical examples demonstrating the importance of unobserved heterogeneity in economic applications.

Prototypical Model of Potential Outcomes

- We consider a setting where there are two possible outcomes for an individual, Y_1 or Y_0:

$$\begin{aligned} Y_1 &= \mu_1(X) + U_1 \\ Y_0 &= \mu_0(X) + U_0. \end{aligned} \tag{1}$$

Prototypical Model of Potential Outcomes

- We consider a setting where there are two possible outcomes for an individual, Y_1 or Y_0:

$$\begin{aligned} Y_1 &= \mu_1(X) + U_1 \\ Y_0 &= \mu_0(X) + U_0. \end{aligned} \tag{1}$$

- Then we can write the treatment effect as

$$Y_1 - Y_0 = \mu_1(X) - \mu_0(X) + U_1 - U_0.$$

OLS and IV under Essential Heterogeneity

- Then the outcome we observe for an individual is

$$\begin{aligned} Y &= DY_1 + (1 - D)Y_0 \\ &= Y_0 + (Y_1 - Y_0)D \qquad (2) \\ &= \mu_0(X) + (\mu_1(X) - \mu_0(X) + U_1 - U_0)D + U_0. \end{aligned}$$

OLS and IV under Essential Heterogeneity

- Then the outcome we observe for an individual is

$$\begin{aligned} Y &= DY_1 + (1 - D)Y_0 \\ &= Y_0 + (Y_1 - Y_0)D \qquad (2) \\ &= \mu_0(X) + (\mu_1(X) - \mu_0(X) + U_1 - U_0)D + U_0. \end{aligned}$$

- Rewriting this in regression notation,

$$Y_i = \alpha + \beta D_i + \varepsilon_i \qquad (3)$$

where $\alpha = \mu_0(X)$, $\beta = \mu_1(X) - \mu_0(X) + U_1 - U_0$ and $\varepsilon = U_0$.

For the case with homogeneous responses, if there is an instrument Z with the properties that

$$\mathrm{Cov}(Z, D) \neq 0 \tag{4}$$

and

$$\mathrm{Cov}(Z, \varepsilon) = 0 \tag{5}$$

then standard IV identifies β, at least in large samples:

$$\text{plim } \hat{\beta}_{\mathrm{IV}} = \frac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, D)} = \beta.$$

- However, it may be that even after conditioning on X there is still variation in β across individuals.

- However, it may be that even after conditioning on X there is still variation in β across individuals.
- If individuals make their decisions with at least partial knowledge of their idiosyncratic gain from the treatment, then the model contains essential heterogeneity.

- However, it may be that even after conditioning on X there is still variation in β across individuals.
- If individuals make their decisions with at least partial knowledge of their idiosyncratic gain from the treatment, then the model contains essential heterogeneity.
- Heckman Urzua and Vytlacil (2006) show that in models with essential heterogeneity, standard instrument variables does not identify any meaningful treatment parameters.

- However, it may be that even after conditioning on X there is still variation in β across individuals.
- If individuals make their decisions with at least partial knowledge of their idiosyncratic gain from the treatment, then the model contains essential heterogeneity.
- Heckman Urzua and Vytlacil (2006) show that in models with essential heterogeneity, standard instrument variables does not identify any meaningful treatment parameters.
- Therefore, a test for the presence of essential heterogeneity is necessary in order to determine whether IV can be used to recover a parameter that is meaningful to economists.

The Choice Model and the IV Approach

- We assume that choices are generated by a latent variable D^*, where
$$D^* = \mu_D(Z) - V \text{ and } D = \mathbf{1}(D^* \geq 0).$$

The Choice Model and the IV Approach

- We assume that choices are generated by a latent variable D^*, where

$$D^* = \mu_D(Z) - V \text{ and } D = \mathbf{1}(D^* \geq 0).$$

- Then the propensity score, or choice probability is

$$P(z) = Pr(D = 1|Z = z) = Pr(\mu_D(z) \geq V) = F_V(\mu_D(z))$$

- Note that

$$\begin{aligned} D = \mathbf{1}(D^* \geq 0) &= \mathbf{1}(\mu_D(z) \geq V) \\ &= \mathbf{1}(F_V(\mu_D(z)) \geq F_V(V)) \\ &= \mathbf{1}(P(z) \geq U_D) \end{aligned}$$

where $U_D = F_V(V)$ is a Uniform$[0,1]$ random variable.

- Note that

$$\begin{aligned} D &= \mathbf{1}(D^* \geq 0) = \mathbf{1}(\mu_D(z) \geq V) \\ &= \mathbf{1}(F_V(\mu_D(z)) \geq F_V(V)) \\ &= \mathbf{1}(P(z) \geq U_D) \end{aligned}$$

  where $U_D = F_V(V)$ is a Uniform$[0,1]$ random variable.

- The MTE is defined, for a given value of $X = x$, as

$$MTE(x, v) = E(Y_1 - Y_0 | X = x, V = v)$$

- That is, it is simply the mean treatment effect when the observables X are fixed at a value x and the unobservable V is fixed at a value v.

- That is, it is simply the mean treatment effect when the observables X are fixed at a value x and the unobservable V is fixed at a value v.
- Other mean effects:

$$ATE = E(Y_1 - Y_0) \tag{6}$$

$$TT = E(Y_1 - Y_0 | D = 1) \tag{7}$$

Basic Idea Behind Our Test

IV-1

$Z \perp\!\!\!\perp \left(Y_0, Y_1, \{D(z)\}_{z \in \mathcal{Z}}\right)$ where $\mathcal{Z}$ is the set of possible values of Z. (**Independence**)

IV-2

$\Pr(D = 1 \mid Z)$ depends on Z (**Rank**).

- Thus,

$$
\begin{aligned}
&E(Y|X = x, P(Z) = p) \\
&\quad = E(DY_1 + (1 - D)Y_0|P(Z) = p, X = x) \\
&\quad = E(Y_0|X = x) + E(D(Y_1 - Y_0)|X = x, P(Z) = p) \\
&\quad = E(Y_0|X = x) + E(Y_1 - Y_0|X = x, D = 1)p \\
&\quad = E(Y_0|X = x) + \int_0^p E(Y_1 - Y_0|X = x, U_D = u_D)du_D
\end{aligned}
$$

- Thus,

$$\begin{aligned} &E(Y|X=x, P(Z)=p) \\ &\quad = E(DY_1 + (1-D)Y_0|P(Z)=p, X=x) \\ &\quad = E(Y_0|X=x) + E(D(Y_1 - Y_0)|X=x, P(Z)=p) \\ &\quad = E(Y_0|X=x) + E(Y_1 - Y_0|X=x, D=1)p \\ &\quad = E(Y_0|X=x) + \int_0^p E(Y_1 - Y_0|X=x, U_D=u_D)du_D \end{aligned}$$

- In the absence of essential heterogeneity, we know that $\beta \perp\!\!\!\perp D$, that is $(Y_1 - Y_0) \perp\!\!\!\perp D$ and therefore

$$\begin{aligned} &E(Y|P(Z)=p, X=x) \\ &\quad = E(Y_0|X=x) + \int_0^p E(Y_1 - Y_0|X=x, U_D=u_D)du_D \\ &\quad = E(Y_0|X=x) + E(Y_1 - Y_0|X=x)p. \end{aligned}$$

- However, under essential heterogeneity, $E(Y|P(Z) = p, X = x)$ will be a general nonlinear function of p.

- However, under essential heterogeneity, $E(Y|P(Z) = p, X = x)$ will be a general nonlinear function of p.
- Note also that the MTE can be found using local instrumental variables (LIV), defined as

$$LIV = \frac{\partial E(Y|X = x, P(Z) = p)}{\partial p} = E(Y_1 - Y_0|X = x, U_D = p).$$

- However, under essential heterogeneity, $E(Y|P(Z)=p, X=x)$ will be a general nonlinear function of p.
- Note also that the MTE can be found using local instrumental variables (LIV), defined as

$$LIV = \frac{\partial E(Y|X=x, P(Z)=p)}{\partial p} = E(Y_1 - Y_0|X=x, U_D=p).$$

- Therefore, in the absence of essential heterogeneity, the LIV estimator will give the constant $E(Y_1 - Y_0|X=x)$.

Testing for Linearity

- We keep the conditioning on X implicit in what follows.

Testing for Linearity

- We keep the conditioning on X implicit in what follows.
- We can write

$$E(Y|P(Z) = p) = h(p) \tag{8}$$

for some general nonlinear function $h(\cdot)$.

Testing for Linearity

- We keep the conditioning on X implicit in what follows.
- We can write

$$E(Y|P(Z)=p)=h(p) \tag{8}$$

for some general nonlinear function $h(\cdot)$.

- What we would like to do is allow for any possible functional form for $h(\cdot)$ and test the parametric null hypothesis

$$H_0: \; h(p)=a+bp \text{ for some } a,b\in\mathbb{R}$$

against the composite alternative

$$H_1: \; \text{not } H_0$$

- In implementing this test, we need to pick a specific alternative against which to test the null hypothesis of linearity.

- In implementing this test, we need to pick a specific alternative against which to test the null hypothesis of linearity.
- Our alternative specification (conditional on X) is

$$h(p) = \sum_{j=0}^{d} \phi_j p^j$$

- In implementing this test, we need to pick a specific alternative against which to test the null hypothesis of linearity.
- Our alternative specification (conditional on X) is

$$h(p) = \sum_{j=0}^{d} \phi_j p^j$$

- Then the test for linearity is simply a test of

$$\begin{aligned} &H_0 : \phi_j = 0 \text{ for } j = 2, \ldots, d \\ &H_1 : \text{not } H_0 \end{aligned}$$

Implementing the Test of Linearity

- We estimate the alternative specification as

$$Y_i = X_i\beta_0 + X_i(\beta_1 - \beta_0)P(Z_i) + \sum_{j=1}^{d} \phi_j P(Z_i)^j + \varepsilon_i \tag{9}$$

Implementing the Test of Linearity

- We estimate the alternative specification as

$$Y_i = X_i\beta_0 + X_i(\beta_1 - \beta_0)P(Z_i) + \sum_{j=1}^{d} \phi_j P(Z_i)^j + \varepsilon_i \tag{9}$$

- We choose to estimate the propensity scores using a probit.

- We propose a simple sequential method.

- We propose a simple sequential method.
- We suggest starting with just a linear term in P, then adding a quadratic term, then a cubic term, etc.

- We propose a simple sequential method.
- We suggest starting with just a linear term in P, then adding a quadratic term, then a cubic term, etc.
- If, after adding a quadratic term, one is already able to reject linearity, then one can stop and take that as evidence of essential heterogeneity.

- We propose a simple sequential method.
- We suggest starting with just a linear term in P, then adding a quadratic term, then a cubic term, etc.
- If, after adding a quadratic term, one is already able to reject linearity, then one can stop and take that as evidence of essential heterogeneity.
- If not, then one can add a cubic term in P and test for both the significance of that term individually, as well as the joint significance of the quadratic and cubic terms.

- We propose a simple sequential method.
- We suggest starting with just a linear term in P, then adding a quadratic term, then a cubic term, etc.
- If, after adding a quadratic term, one is already able to reject linearity, then one can stop and take that as evidence of essential heterogeneity.
- If not, then one can add a cubic term in P and test for both the significance of that term individually, as well as the joint significance of the quadratic and cubic terms.
- If either or both are significant, this provides some evidence of unobserved heterogeneity.

- Inherent in our test of linearity is the standard bias-variance tradeoff.

- Inherent in our test of linearity is the standard bias-variance tradeoff.
- That is, as we increase the number of polynomial terms we expect to get a more accurate approximation to the true MTE, however the standard errors will eventually begin increasing.

Implementing the Test of Linearity

- Inherent in our test of linearity is the standard bias-variance tradeoff.
- That is, as we increase the number of polynomial terms we expect to get a more accurate approximation to the true MTE, however the standard errors will eventually begin increasing.
- In order to help choose, then, the optimal number of polynomial terms to include, we suggest constructing a nonparametric (or semiparametric) estimate of the MTE to use as a reference.

- While it is unlikely that a test of linearity on this nonparametric estimate directly would be able to reject linearity, it is useful to see how close the polynomial approximations are to this more flexible estimate.

Implementing the Test of Linearity

- While it is unlikely that a test of linearity on this nonparametric estimate directly would be able to reject linearity, it is useful to see how close the polynomial approximations are to this more flexible estimate.
- In our estimation below, we use a local polynomial approximation as our flexible functional form to which we compare our polynomial estimates.

Implementing the Test of Linearity

- The statistical tests we use to test the coefficients in our regressions are t-tests and Wald tests.

- The statistical tests we use to test the coefficients in our regressions are t-tests and Wald tests.
- To get standard errors of the coefficients, we bootstrap 50 times and reestimate the propensity scores P, as well as the outcome equation $E(Y|P(Z) = p)$ in each bootstrap sample.

- The statistical tests we use to test the coefficients in our regressions are t-tests and Wald tests.
- To get standard errors of the coefficients, we bootstrap 50 times and reestimate the propensity scores P, as well as the outcome equation $E(Y|P(Z) = p)$ in each bootstrap sample.
- We then use t-tests to test for the significance of individual coefficients and Wald tests to test for the joint significance of all of the nonlinear terms in P.

Testing for Heterogeneity Using LATE

- Another way to test for the linearity of $E(Y|P(Z)=p)$ in p is to use the local average treatment effect (LATE) parameter of Imbens and Angrist (1994).

$$LATE(z', z) = \frac{E(Y|Z=z') - E(Y|Z=z)}{Pr(D=1|Z=z') - Pr(D=1|Z=z)}$$

- Another way to test for the linearity of $E(Y|P(Z)=p)$ in p is to use the local average treatment effect (LATE) parameter of Imbens and Angrist (1994).

$$LATE(z',z) = \frac{E(Y|Z=z') - E(Y|Z=z)}{Pr(D=1|Z=z') - Pr(D=1|Z=z)}$$

- Vytlacil (2002) shows that this can be written as

$$LATE(u'_D,u_D) = \frac{E(Y|P(Z)=u'_D) - E(Y|P(Z)=u_D)}{u'_D - u_D} \quad (10)$$

- Another way to test for the linearity of $E(Y|P(Z)=p)$ in p is to use the local average treatment effect (LATE) parameter of Imbens and Angrist (1994).

$$LATE(z', z) = \frac{E(Y|Z=z') - E(Y|Z=z)}{Pr(D=1|Z=z') - Pr(D=1|Z=z)}$$

- Vytlacil (2002) shows that this can be written as

$$LATE(u_D', u_D) = \frac{E(Y|P(Z)=u_D') - E(Y|P(Z)=u_D)}{u_D' - u_D} \tag{10}$$

- If $E(Y|P(Z)=p)$ were linear then $LATE(v, w)$ would be the same for any points $v, w \in \text{Supp}(P(Z))$.

- Therefore, another testable implication of the absence of essential heterogeneity is the equality of $LATE$ at all evaluation points in the support of P.

Testing for Heterogeneity Using LATE

- Therefore, another testable implication of the absence of essential heterogeneity is the equality of *LATE* at all evaluation points in the support of P.
- In practice, however, estimating the *LATE* over different intervals is difficult because it involves forming conditional expectations where we are conditioning on the value of a continuous variable (the propensity score).

- Therefore, we take a different approach to test for heterogeneity using $LATE$s.

- Therefore, we take a different approach to test for heterogeneity using *LATE*s.
- The linear IV estimator is just a weighted average of the *MTE* with weights integrating to 1.

- Therefore, we take a different approach to test for heterogeneity using *LATE*s.
- The linear IV estimator is just a weighted average of the *MTE* with weights integrating to 1.
- Therefore, we consider forming an IV estimate using just the data from a given interval of our propensity score.

- Therefore, we take a different approach to test for heterogeneity using $LATE$s.
- The linear IV estimator is just a weighted average of the MTE with weights integrating to 1.
- Therefore, we consider forming an IV estimate using just the data from a given interval of our propensity score.
- This estimate will be some weighted average of the MTE.

- If we form another IV estimate over a different interval of P that will be a weighted average of a different portion of the MTE.

- If we form another IV estimate over a different interval of P that will be a weighted average of a different portion of the MTE.
- However, the absence of essential heterogeneity implies that these IV estimates must be the same, because they are both weighted averages of the same quantity with weights summing to 1.

- If we form another IV estimate over a different interval of P that will be a weighted average of a different portion of the MTE.
- However, the absence of essential heterogeneity implies that these IV estimates must be the same, because they are both weighted averages of the same quantity with weights summing to 1.
- This suggests a test of equality of the IV estimates across different intervals of P as a way to test for essential heterogeneity.

- To implement this test we form, for two specified intervals$[\underline{p}_1, \overline{p}_1]$ and $[\underline{p}_2, \overline{p}_2]$,

$$
\begin{aligned}
IV\left(\underline{p}_1, \overline{p}_1\right) &= \frac{\operatorname{Cov}\left(Y, P(Z) \mid P(Z) \in [\underline{p}_1, \overline{p}_1]\right)}{\operatorname{Var}\left(P(Z) \mid P(Z) \in [\underline{p}_1, \overline{p}_1]\right)} \\
IV\left(\underline{p}_2, \overline{p}_2\right) &= \frac{\operatorname{Cov}\left(Y, P(Z) \mid P(Z) \in [\underline{p}_2, \overline{p}_2]\right)}{\operatorname{Var}\left(P(Z) \mid P(Z) \in [\underline{p}_2, \overline{p}_2]\right)}
\end{aligned}
$$

and then test

$$
\begin{aligned}
H_0 &: IV\left(\underline{p}_1, \overline{p}_1\right) = IV\left(\underline{p}_2, \overline{p}_2\right) \\
H_1 &: IV\left(\underline{p}_1, \overline{p}_1\right) \neq IV\left(\underline{p}_2, \overline{p}_2\right)
\end{aligned}
$$

- Because there is estimation error from two stages (estimating $P(Z)$ and constructing this IV estimate), we bootstrap the difference between these estimates and check whether 0 lies in the tail of our bootstrapped distribution of the difference between the estimates.

The Power of the Tests

- In this section we provide the results of carrying out our tests on simulated data which is generated from a fairly restrictive model – namely the Generalized Roy Model, where all errors are normal.

The Power of the Tests

- In this section we provide the results of carrying out our tests on simulated data which is generated from a fairly restrictive model – namely the Generalized Roy Model, where all errors are normal.
- We simply use this as our base case because it allows for the simple parameterization of essential heterogeneity.

- Our potential outcomes are

$$\begin{aligned} Y_0 &= \alpha_0 + \beta_{10} X_1 + \beta_{20} X_2 + U_0 \\ Y_1 &= \alpha_1 + \beta_{11} X_1 + \beta_{21} X_2 + U_1. \end{aligned}$$

- Our potential outcomes are

$$\begin{aligned} Y_0 &= \alpha_0 + \beta_{10}X_1 + \beta_{20}X_2 + U_0 \\ Y_1 &= \alpha_1 + \beta_{11}X_1 + \beta_{21}X_2 + U_1. \end{aligned}$$

- Our choice equation is

$$D = \mathbf{1}(\alpha_d + \beta_d Z \geq V)$$

where

$$\begin{pmatrix} U_1 \\ U_0 \\ V \end{pmatrix} \sim N\left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array}, \begin{pmatrix} \sigma_1^2 & \sigma_{10} & \sigma_{1V} \\ \sigma_{10} & \sigma_0^2 & \sigma_{0V} \\ \sigma_{1V} & \sigma_{0V} & \sigma_V^2 \end{pmatrix} \right)$$

- We also generate the regressors and the instrument as normal random variables with distribution

$$\begin{pmatrix} X_1 \\ X_2 \\ Z \end{pmatrix} \sim N\left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array}, \begin{pmatrix} \sigma^2_{X_1} & \sigma_{X_1X_2} & \sigma_{X_1Z} \\ \sigma_{X_1X_2} & \sigma^2_{X_2} & \sigma_{X_2Z} \\ \sigma_{X_1Z} & \sigma_{X_2Z} & \sigma^2_{Z} \end{pmatrix} \right)$$

- We also generate the regressors and the instrument as normal random variables with distribution

$$\begin{pmatrix} X_1 \\ X_2 \\ Z \end{pmatrix} \sim N\left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array}, \begin{pmatrix} \sigma^2_{X_1} & \sigma_{X_1 X_2} & \sigma_{X_1 Z} \\ \sigma_{X_1 X_2} & \sigma^2_{X_2} & \sigma_{X_2 Z} \\ \sigma_{X_1 Z} & \sigma_{X_2 Z} & \sigma^2_Z \end{pmatrix} \right)$$

- In this model, the marginal treatment effect is given by

$$\begin{aligned} MTE(X = x, P(Z) = p) \\ &= (\alpha_1 - \alpha_0) + (\beta_{11} - \beta_{10}) X_1 + (\beta_{21} - \beta_{20}) X_2 \\ &\quad + (\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0) \Phi^{-1}(p) \end{aligned}$$

where $\Phi^{-1}(\cdot)$ is the inverse of a standard normal CDF.

- We also generate the regressors and the instrument as normal random variables with distribution

$$\begin{pmatrix} X_1 \\ X_2 \\ Z \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{X_1} & \sigma_{X_1 X_2} & \sigma_{X_1 Z} \\ \sigma_{X_1 X_2} & \sigma^2_{X_2} & \sigma_{X_2 Z} \\ \sigma_{X_1 Z} & \sigma_{X_2 Z} & \sigma^2_Z \end{pmatrix} \right)$$

- In this model, the marginal treatment effect is given by

$$\begin{aligned} & MTE(X = x, P(Z) = p) \\ & \quad = (\alpha_1 - \alpha_0) + (\beta_{11} - \beta_{10}) X_1 + (\beta_{21} - \beta_{20}) X_2 \\ & \qquad + (\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0) \Phi^{-1}(p) \end{aligned}$$

  where $\Phi^{-1}(\cdot)$ is the inverse of a standard normal CDF.

- It is $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0$ index which lets us vary the degree of heterogeneity of treatment effects and trace out the power function in this dimension.

- The test we are using in both the test of linearity and the testing using *LATE* is a Wald test.

- The test we are using in both the test of linearity and the testing using *LATE* is a Wald test.
- In order to get the distribution of the test statistic under the alternative hypothesis (of essential heterogeneity), we need to somehow restrict the form of heterogeneity and completely specify the data generating process.

- The test we are using in both the test of linearity and the testing using *LATE* is a Wald test.
- In order to get the distribution of the test statistic under the alternative hypothesis (of essential heterogeneity), we need to somehow restrict the form of heterogeneity and completely specify the data generating process.
- In our model, we can parameterize the amount of heterogeneity using the term $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0$ and hence we also know that under the null hypothesis $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.

- The test we are using in both the test of linearity and the testing using *LATE* is a Wald test.
- In order to get the distribution of the test statistic under the alternative hypothesis (of essential heterogeneity), we need to somehow restrict the form of heterogeneity and completely specify the data generating process.
- In our model, we can parameterize the amount of heterogeneity using the term $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0$ and hence we also know that under the null hypothesis $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.
- For our model we can also simulate the exact distribution of the test statistic under the null using a bootstrap procedure.

- So our test becomes

$$\begin{aligned} H_0 &: \text{ Generalized Roy model with } \rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0 \\ H_1 &: \text{ Generalized Roy model with } \rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k \end{aligned}$$

with σ_1 and σ_0 fixed.

- So our test becomes

$$\begin{aligned} H_0 &: \text{ Generalized Roy model with } \rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0 \\ H_1 &: \text{ Generalized Roy model with } \rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k \end{aligned}$$

  with σ_1 and σ_0 fixed.

- We use this bootstrap procedure and a grid of alternative values for k to trace out the power function for each of our tests in the one dimension of this index.

- So our test becomes

$$
\begin{aligned}
H_0 &: \text{ Generalized Roy model with } \rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0 \\
H_1 &: \text{ Generalized Roy model with } \rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k
\end{aligned}
$$

  with σ_1 and σ_0 fixed.

- We use this bootstrap procedure and a grid of alternative values for k to trace out the power function for each of our tests in the one dimension of this index.

- We calculate the power functions using both the exact distribution of the test statistic under the null and the asymptotic (χ^2) distribution of the test statistic under the null.

Polynomial Test

The procedure for doing this is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.

The procedure for doing this is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.
2. Sample N observations with replacement from the empirical distribution of the data.

The procedure for doing this is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.
2. Sample N observations with replacement from the empirical distribution of the data.
3. Estimate $\widehat{P}(Z_i^*)$ using a probit.

The procedure for doing this is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.
2. Sample N observations with replacement from the empirical distribution of the data.
3. Estimate $\widehat{P}(Z_i^*)$ using a probit.
4. Run OLS on $Y_i^* = X_i^*\beta_0 + X_i^*(\beta_1 - \beta_0)\widehat{P}(Z_i^*) + \sum_{j=1}^{J} \phi_j \widehat{P}(Z_i^*)^j + \varepsilon_i$.

5. Form V^*, a Huber-White robust estimator of the covariance matrix of the parameters.

5. Form V^*, a Huber-White robust estimator of the covariance matrix of the parameters.

6. Form the test statistic $W^* = \phi' \left[R V^* R' \right]^{-1} \phi$, where ϕ is the vector of coefficients on the nonlinear term of $\widehat{P}(Z_i^*)$ and R is the $(J-1) \times k$ restriction matrix that picks out the coefficients on nonlinear terms of $\widehat{P}(Z_i^*)$.

5. Form V^*, a Huber-White robust estimator of the covariance matrix of the parameters.
6. Form the test statistic $W^* = \phi' \left[RV^*R'\right]^{-1} \phi$, where ϕ is the vector of coefficients on the nonlinear term of $\widehat{P}(Z_i^*)$ and R is the $(J-1) \times k$ restriction matrix that picks out the coefficients on nonlinear terms of $\widehat{P}(Z_i^*)$.
7. Repeat steps two through six 1,000 times.

5. Form V^*, a Huber-White robust estimator of the covariance matrix of the parameters.

6. Form the test statistic $W^* = \phi' \left[RV^*R'\right]^{-1} \phi$, where ϕ is the vector of coefficients on the nonlinear term of $\widehat{P}(Z_i^*)$ and R is the $(J-1) \times k$ restriction matrix that picks out the coefficients on nonlinear terms of $\widehat{P}(Z_i^*)$.

7. Repeat steps two through six 1,000 times.

8. Find the 0.95 quantile of the distribution of W^* from the bootstrap samples, call this critical value $c^*_{0.95}$.

Then, for a given alternative hypothesis (k), the procedure to calculate the power of the test is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k$.

Then, for a given alternative hypothesis (k), the procedure to calculate the power of the test is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k$.
2. Repeat steps two through six from above 500 times – in each iteration calculating the test statistic W^*_{alt}.

Then, for a given alternative hypothesis (k), the procedure to calculate the power of the test is:

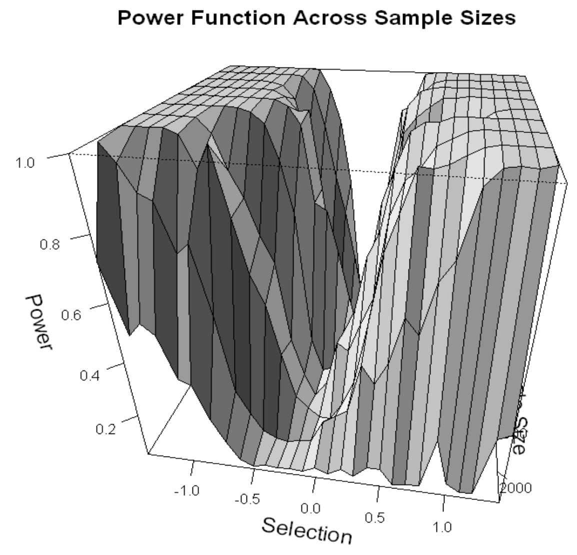
1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k$.
2. Repeat steps two through six from above 500 times – in each iteration calculating the test statistic W^*_{alt}.
3. For the power using the exact distribution of under the null, calculate the proportion of bootstrap samples in which $W^*_{alt} > c^*_{0.95}$.

Then, for a given alternative hypothesis (k), the procedure to calculate the power of the test is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k$.
2. Repeat steps two through six from above 500 times – in each iteration calculating the test statistic W^*_{alt}.
3. For the power using the exact distribution of under the null, calculate the proportion of bootstrap samples in which $W^*_{alt} > c^*_{0.95}$.
4. For the power using the asymptotic χ^2 distribution under the null, calculate the proportion of the bootstrap samples in which $W^*_{alt} > Q_{\chi^2_{J-1}}(0.95)$ where $Q_{\chi^2_k}(\tau)$ is the τ-quantile of a χ^2 distribution with k degrees of freedom.

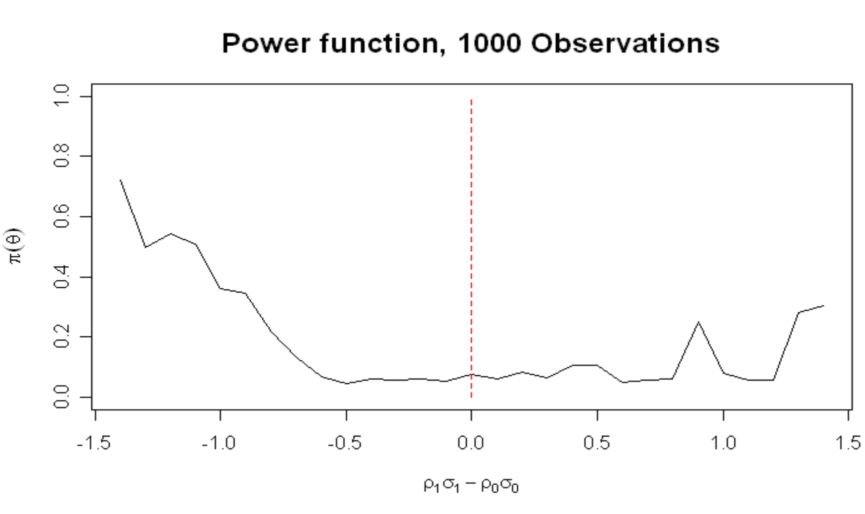
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



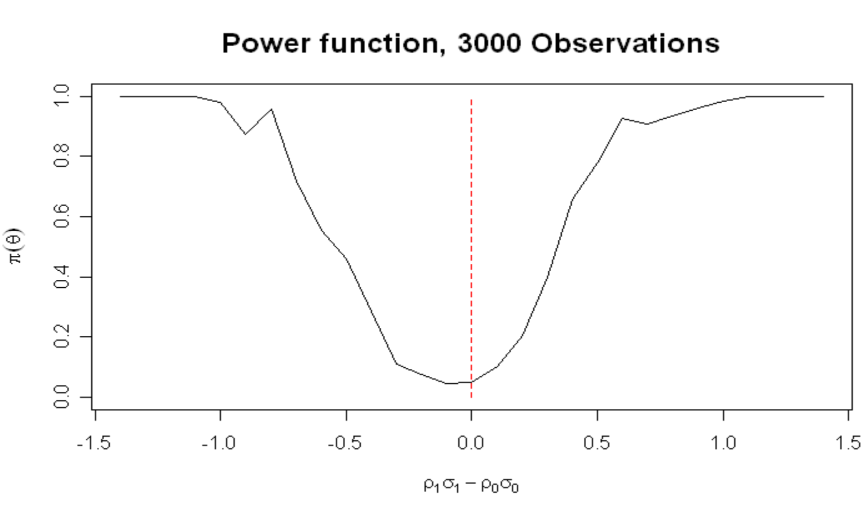
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



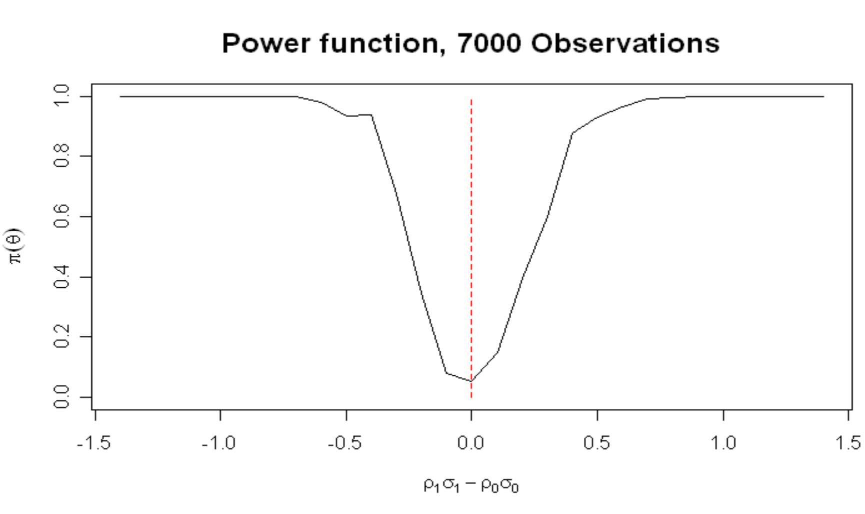
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



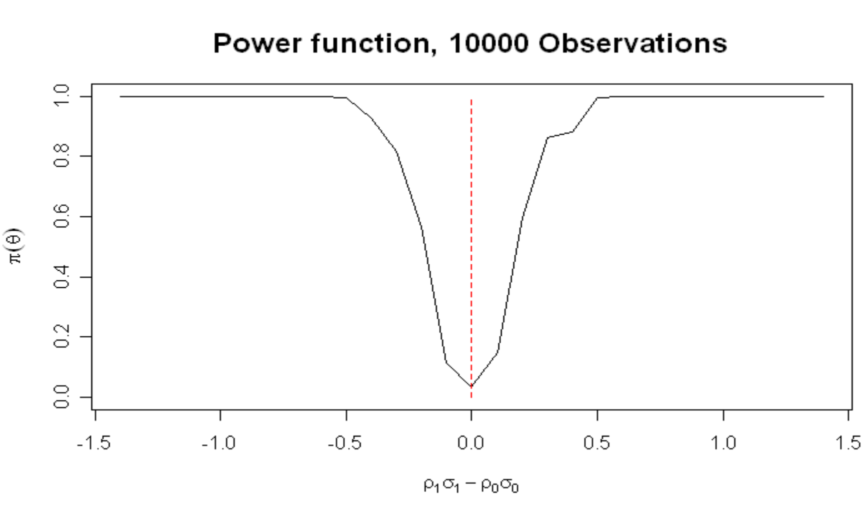
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)

- The variance of the instrument is 10.

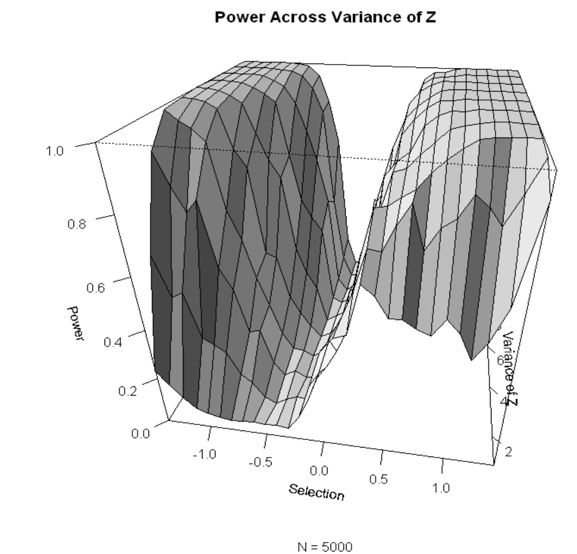
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)

- The variance of the instrument is 10.
- For each alternative hypothesis (each value of $\rho_{1v}\sigma_1 - \rho_{0v}\sigma_0$) we bootstrap the Wald statistic for the test that the coefficient on the P^2 term is zero 500 times and calculate what proportion of those test statistics lie outside the 95th percentile of a χ_2 distribution with 1 degree of freedom (we are testing one coefficient).

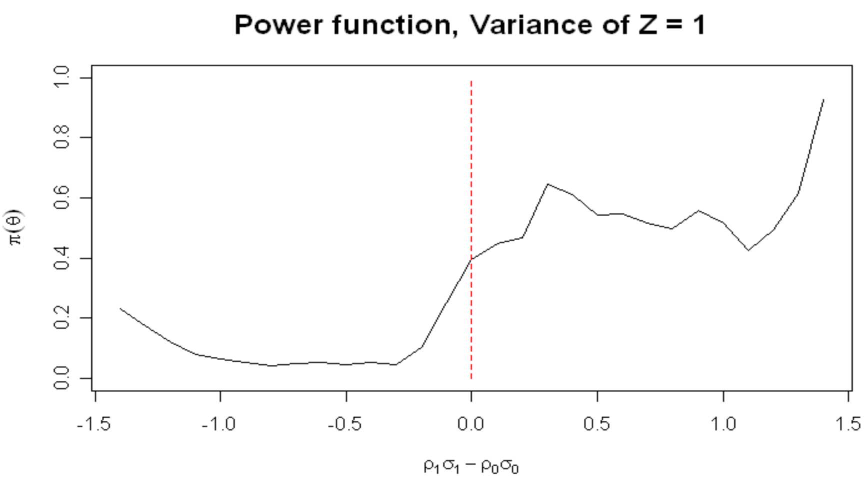
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



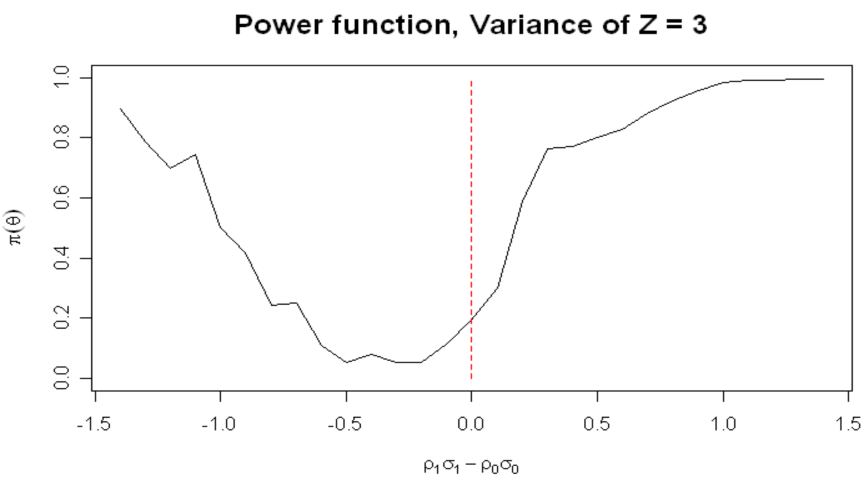
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



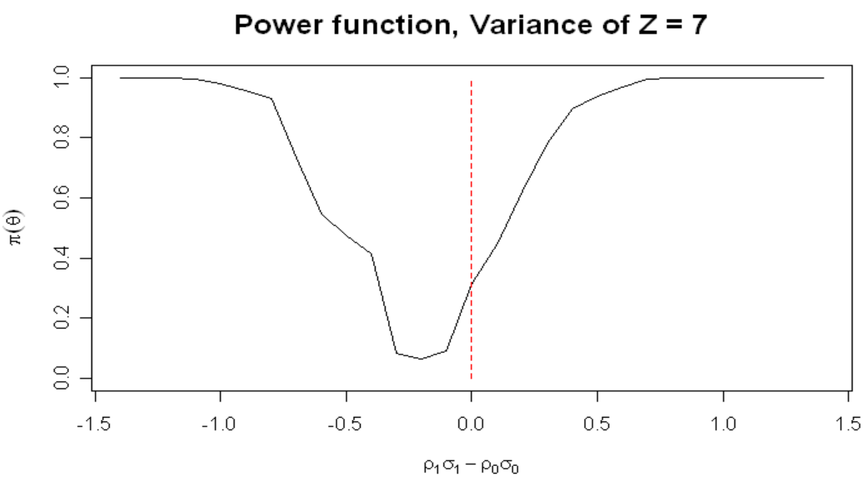
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



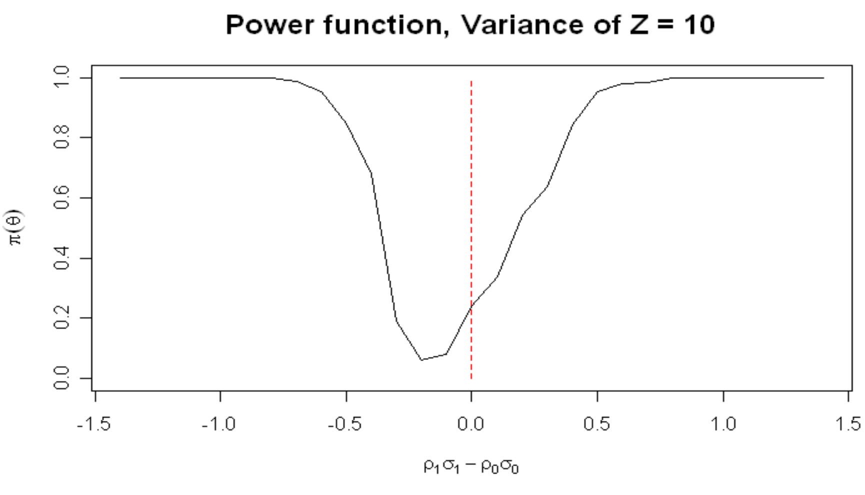
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)

- The sample size is 5,000.

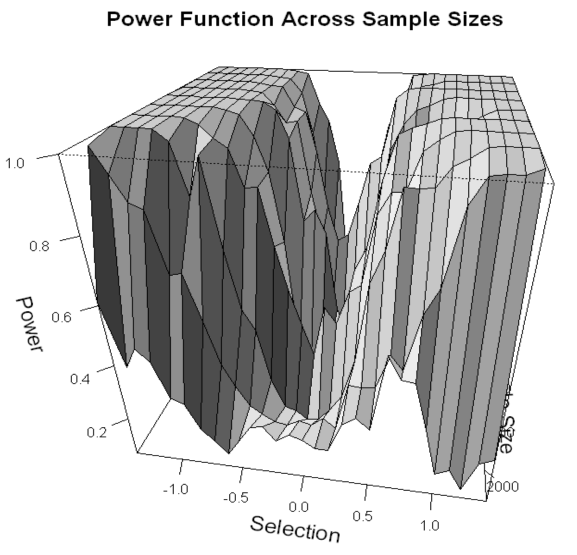
Power of the Test of Linearity of $E(Y|P)$

Using Quadratic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)

- The sample size is 5,000.
- For each alternative hypothesis (each value of $\rho_{1v}\sigma_1 - \rho_{0v}\sigma_0$) we bootstrap the Wald statistic for the test that the coefficient on the P^2 term is zero 500 times and calculate what proportion of those test statistics lie outside the 95th percentile of a χ^2 distribution with 1 degree of freedom (we are testing one coefficient).

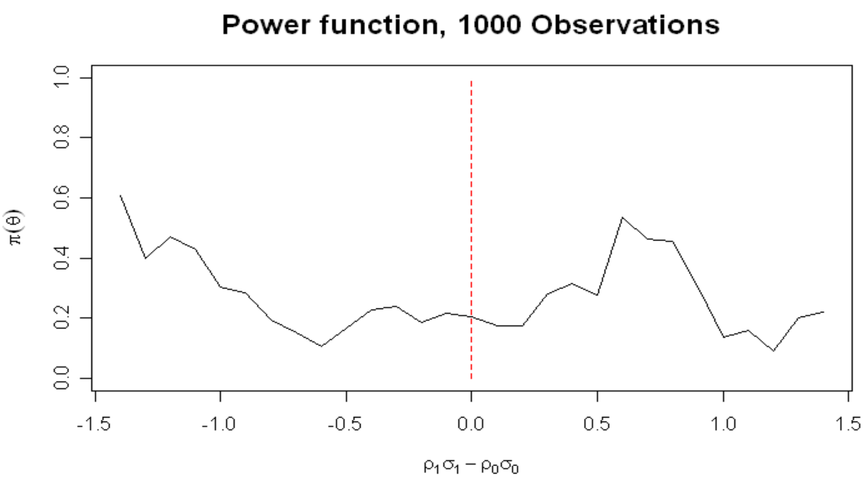
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



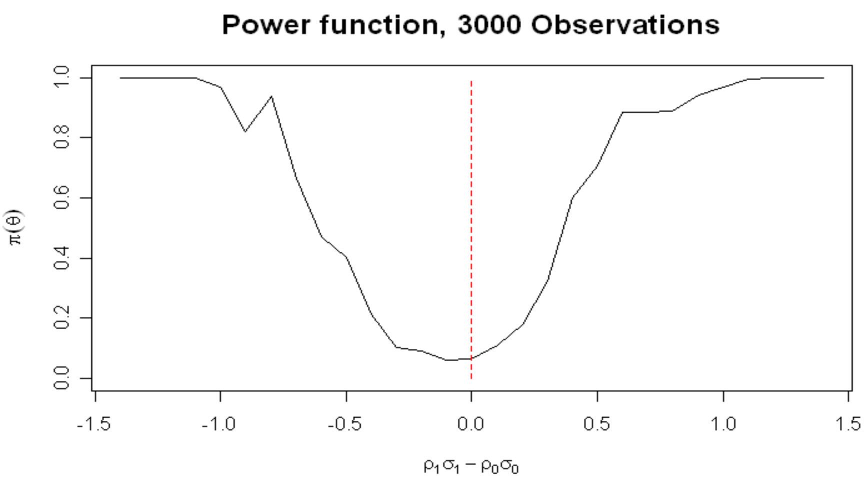
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



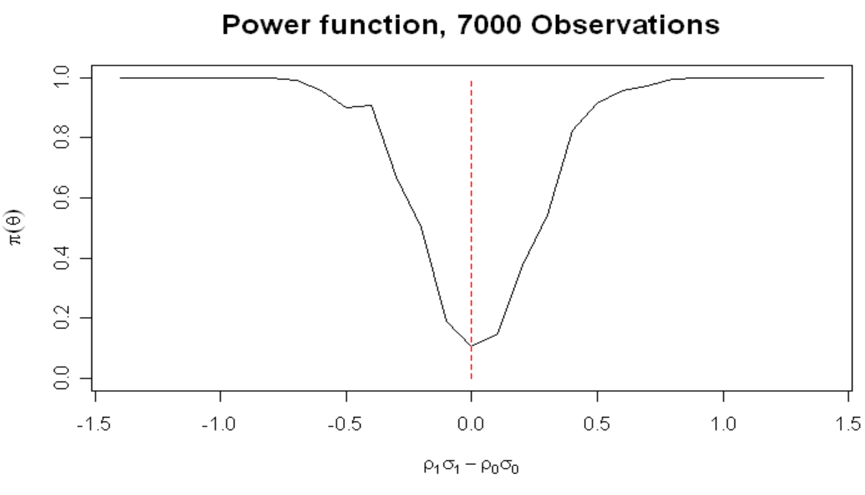
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



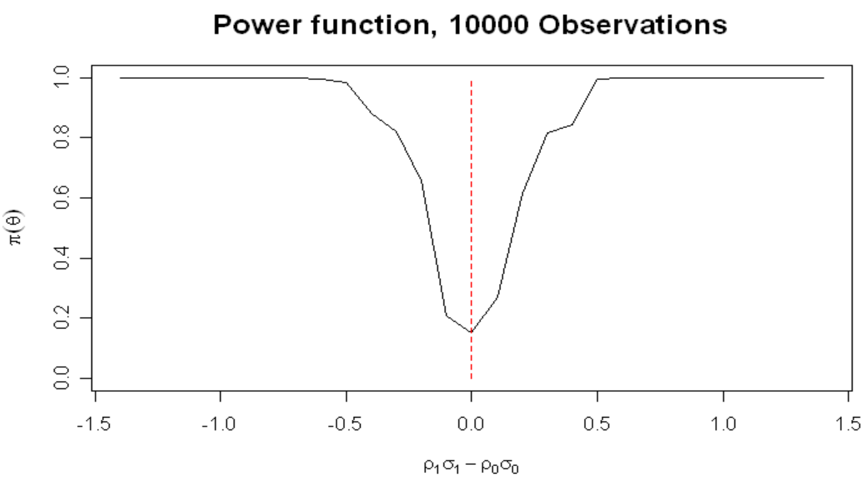
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)



Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)

- The variance of the instrument is 10.

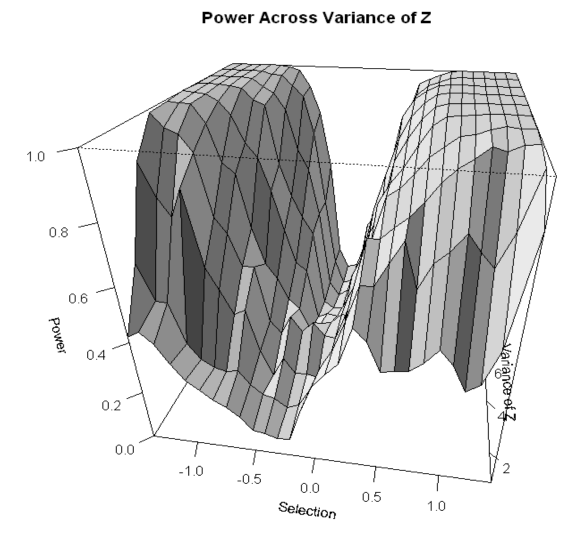
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying Sample Size (Using Chi-Square Distribution under the Null)

- The variance of the instrument is 10.
- For each alternative hypothesis (each value of $\rho_{1v}\sigma_1 - \rho_{0v}\sigma_0$) we bootstrap the Wald statistic for the joint test that the coefficients on the P^2 and P^3 terms are zero 500 times and calculate what proportion of those test statistics lie outside the 95th percentile of a χ^2 distribution with 2 degrees of freedom (we are testing two coefficients).

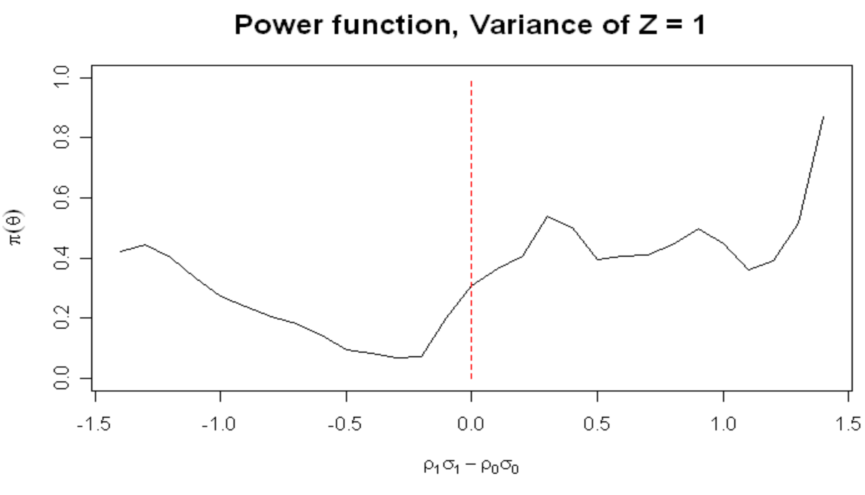
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



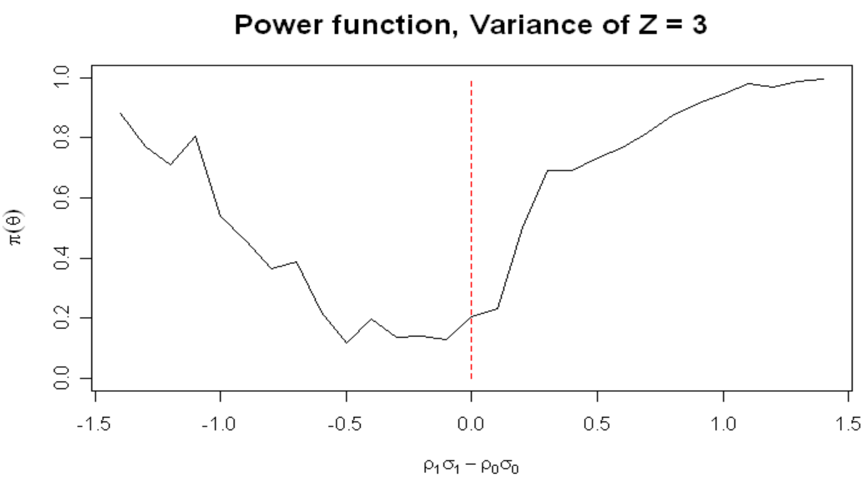
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



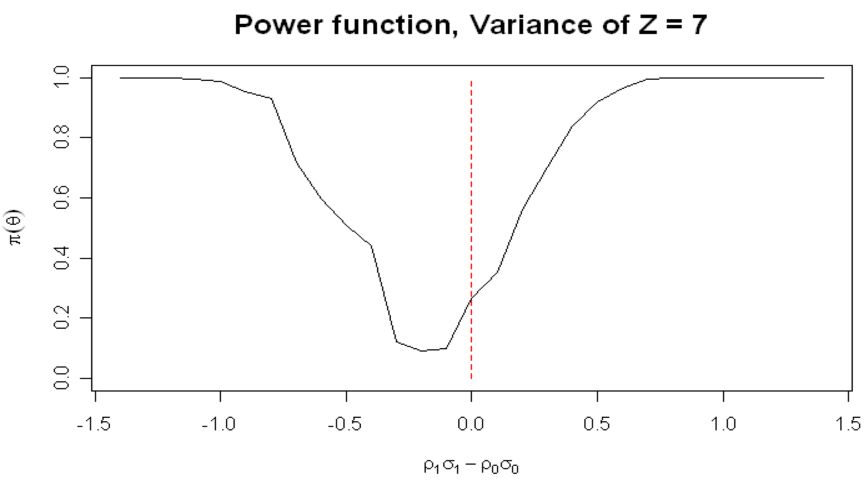
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



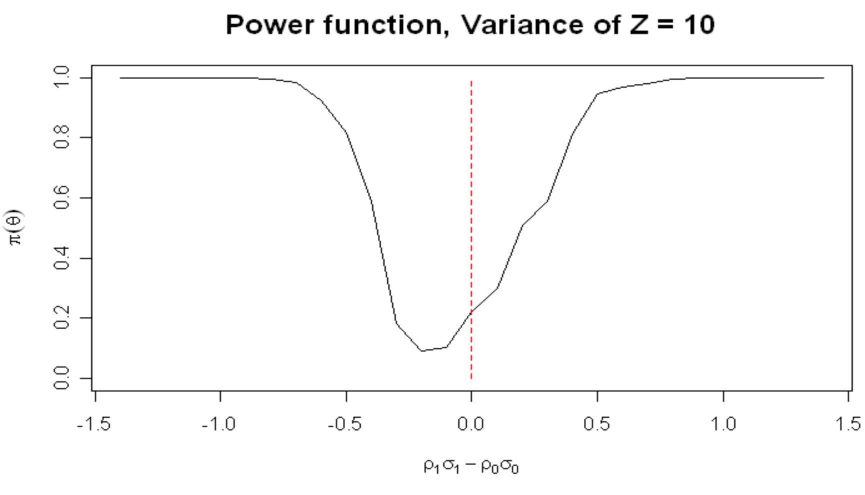
Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)



Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)

- The sample size is 5,000.

Power of the Test of Linearity of $E(Y|P)$

Using Cubic Polynomial in P, Varying σ_Z (Using Chi-Square Distribution under the Null)

- The sample size is 5,000.
- For each alternative hypothesis (each value of $\rho_{1v}\sigma_1 - \rho_{0v}\sigma_0$) we bootstrap the Wald statistic for the joint test that the coefficients on the P^2 and P^3 terms are zero 500 times and calculate what proportion of those test statistics lie outside the 95th percentile of a χ^2 distribution with 2 degrees of freedom (we are testing two coefficients).

LATE/IV Test

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.
2. Calculate the median of the propensity scores in the simulated data $p_{med} = Q_{\widehat{F}}(0.5)$.

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.
2. Calculate the median of the propensity scores in the simulated data $p_{med} = Q_{\widehat{F}}(0.5)$.
3. Sample N observations with replacement from the empirical distribution of the data, $\widehat{F}$.

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = 0$.
2. Calculate the median of the propensity scores in the simulated data $p_{med} = Q_{\widehat{F}}(0.5)$.
3. Sample N observations with replacement from the empirical distribution of the data, $\widehat{F}$.
4. Estimate $\widehat{P}(Z_i^*)$ using a probit.

5. Regress $X * D$ on $X * P(Z)$ and $X * P(Z) * \mathbf{1}(P(Z) > p_{med})$ and calculate the fitted values, call them $\widehat{XD}$. Regress $X * D * \mathbf{1}(P(Z) > p_{med})$ on $X * P(Z)$ and $X * P(Z) * \mathbf{1}(P(Z) > p_{med})$ and calculate the fitted values, call them $\widehat{XD}^{+}$.

5. Regress $X * D$ on $X * P(Z)$ and $X * P(Z) * \mathbf{1}(P(Z) > p_{med})$ and calculate the fitted values, call them $\widehat{XD}$. Regress $X * D * \mathbf{1}(P(Z) > p_{med})$ on $X * P(Z)$ and $X * P(Z) * \mathbf{1}(P(Z) > p_{med})$ and calculate the fitted values, call them $\widehat{XD}^{+}$.

6. Regress Y on X, $X * \mathbf{1}(P(Z) > p_{med})$ and the fitted values $\widehat{XD}$ and $\widehat{XD}^{+}$.

5. Regress $X * D$ on $X * P(Z)$ and $X * P(Z) * \mathbf{1}(P(Z) > p_{med})$ and calculate the fitted values, call them $\widehat{XD}$. Regress $X * D * \mathbf{1}(P(Z) > p_{med})$ on $X * P(Z)$ and $X * P(Z) * \mathbf{1}(P(Z) > p_{med})$ and calculate the fitted values, call them $\widehat{XD}^+$.
6. Regress Y on X, $X * \mathbf{1}(P(Z) > p_{med})$ and the fitted values $\widehat{XD}$ and $\widehat{XD}^+$.
7. Call $\underline{\gamma}$ the vector of coefficients on $\widehat{XD}$, which are the IV estimate using observations below the median, and $\overline{\gamma}$ the vector of coefficients on $\widehat{XD}$ plus the vector of coefficients on $\widehat{XD}^+$, this will be the IV estimate using observations above the median.

8. Form V^*, a Huber-White robust estimator of the covariance matrix of the parameters.

8. Form V^*, a Huber-White robust estimator of the covariance matrix of the parameters.
9. Form the test statistic $W^* = (\underline{\gamma} - \overline{\gamma})' \left[RV^*R'\right]^{-1} (\underline{\gamma} - \overline{\gamma})$ where R is restriction matrix that selects the relevant terms of the covariance matrix.

8. Form V^*, a Huber-White robust estimator of the covariance matrix of the parameters.
9. Form the test statistic $W^* = (\underline{\gamma} - \overline{\gamma})' \left[RV^*R'\right]^{-1} (\underline{\gamma} - \overline{\gamma})$ where R is restriction matrix that selects the relevant terms of the covariance matrix.
10. Repeat steps two through nine 1,000 times.

8. Form V^*, a Huber-White robust estimator of the covariance matrix of the parameters.
9. Form the test statistic $W^* = (\underline{\gamma} - \overline{\gamma})' \left[R V^* R' \right]^{-1} (\underline{\gamma} - \overline{\gamma})$ where R is restriction matrix that selects the relevant terms of the covariance matrix.
10. Repeat steps two through nine 1,000 times.
11. Find the 0.95 quantile of the distribution of W^* from the bootstrap samples, call this critical value $c^*_{0.95}$.

Then, for each alternative hypothesis, the procedure to calculate the power of the test is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k$.

Then, for each alternative hypothesis, the procedure to calculate the power of the test is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k$.
2. Repeat steps two through nine above 500 times – in each iteration calculating the test statistic W^*_{alt}.

Then, for each alternative hypothesis, the procedure to calculate the power of the test is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k$.
2. Repeat steps two through nine above 500 times – in each iteration calculating the test statistic W^*_{alt}.
3. For the power using the exact distribution of under the null, calculate the proportion of bootstrap samples in which $W^*_{alt} > c^*_{0.95}$.

Then, for each alternative hypothesis, the procedure to calculate the power of the test is:

1. Generate data under the parameterization such that $\rho_{1V}\sigma_1 - \rho_{0V}\sigma_0 = k$.
2. Repeat steps two through nine above 500 times – in each iteration calculating the test statistic W^*_{alt}.
3. For the power using the exact distribution of under the null, calculate the proportion of bootstrap samples in which $W^*_{alt} > c^*_{0.95}$.
4. For the power using the asymptotic χ^2 distribution under the null, calculate the proportion of the bootstrap samples in which $W^*_{alt} > Q_{\chi^2_{\dim(X)}}(0.95)$ where $Q_{\chi^2_k}(\tau)$ is the τ-quantile of a χ^2 distribution with k degrees of freedom.

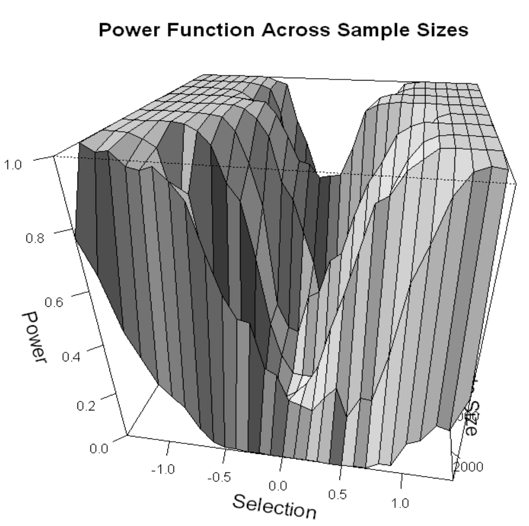
- Finally, we also calculate the power of a test for whether the IV estimates above and below the median differ, but based on a specification which does not include all of the $X * D$ interactions.

- Finally, we also calculate the power of a test for whether the IV estimates above and below the median differ, but based on a specification which does not include all of the $X * D$ interactions.
- This specification just regresses Y on X and D and instruments D with $P(Z)$.

Power of the Test of Equality of IV Estimates

Using Propensity Scores Above and Below the Median, Varying Sample Size

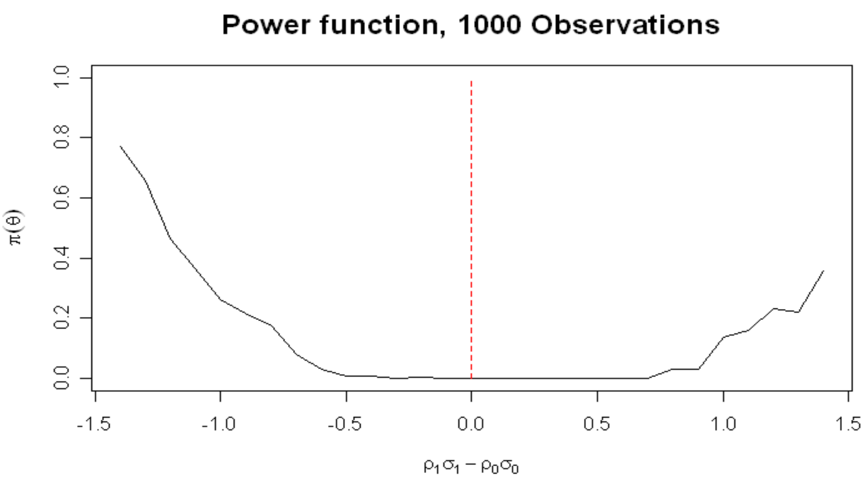
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates

Using Propensity Scores Above and Below the Median, Varying Sample Size

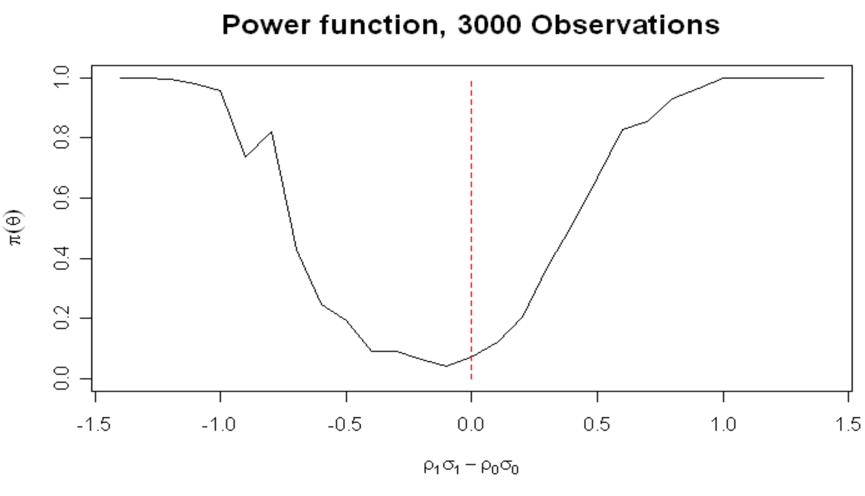
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates

Using Propensity Scores Above and Below the Median, Varying Sample Size

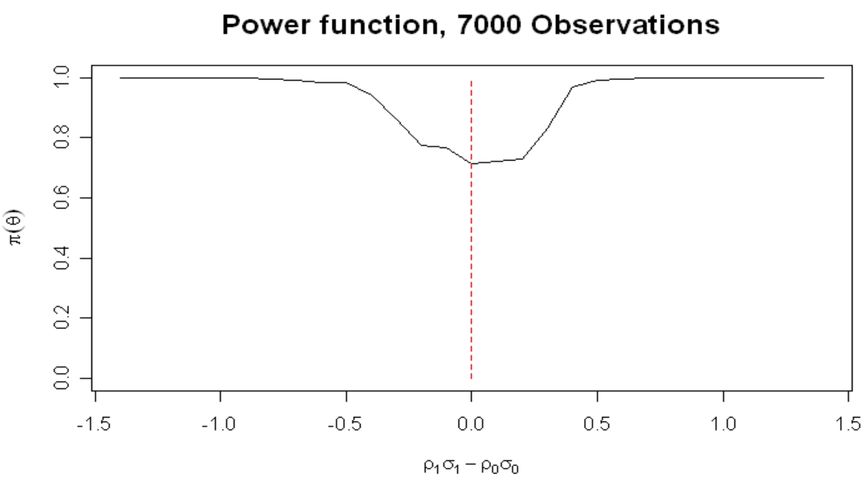
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates

Using Propensity Scores Above and Below the Median, Varying Sample Size

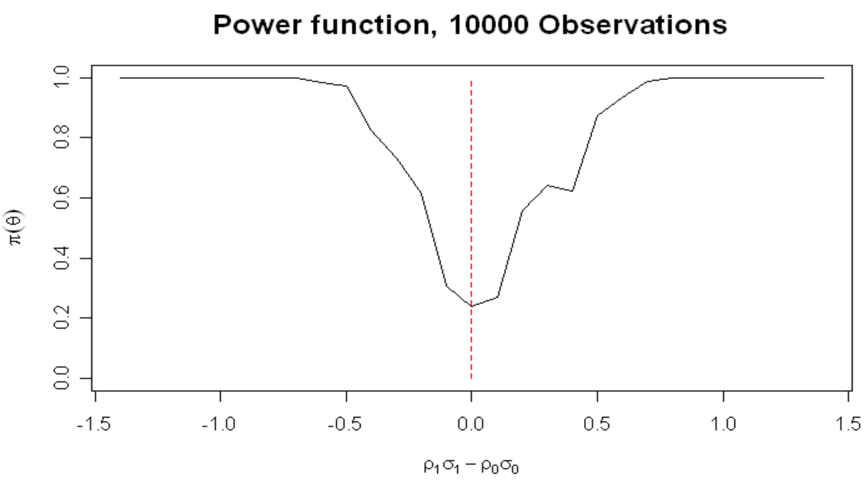
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates

Using Propensity Scores Above and Below the Median, Varying Sample Size

(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates

Using Propensity Scores Above and Below the Median, Varying Sample Size

(Using Chi-Square Distribution under the Null)

- The variance of the instrument is 10.

Power of the Test of Equality of IV Estimates

Using Propensity Scores Above and Below the Median, Varying Sample Size

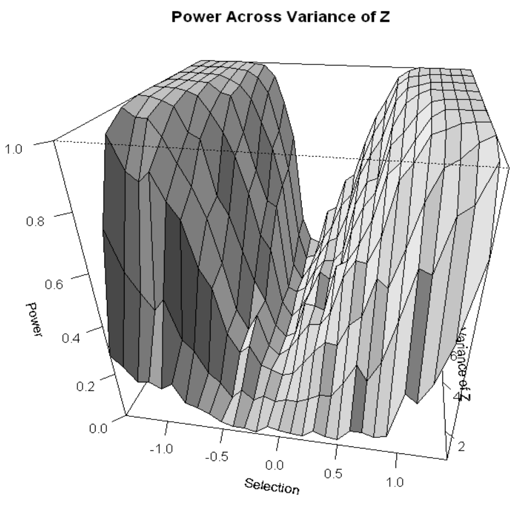
(Using Chi-Square Distribution under the Null)

- The variance of the instrument is 10.
- The power is calculated for each alternative hypothesis (each value of $\rho_{1v}\sigma_1 - \rho_{0v}\sigma_0$) by bootstrapping the Wald statistic 500 times calculating what proportion of the test statistics lie outside the 95th percentile of a χ^2 distribution with 3 degrees of freedom (we are testing 3 coefficients).

Power of the Test of Equality of IV Estimates Using

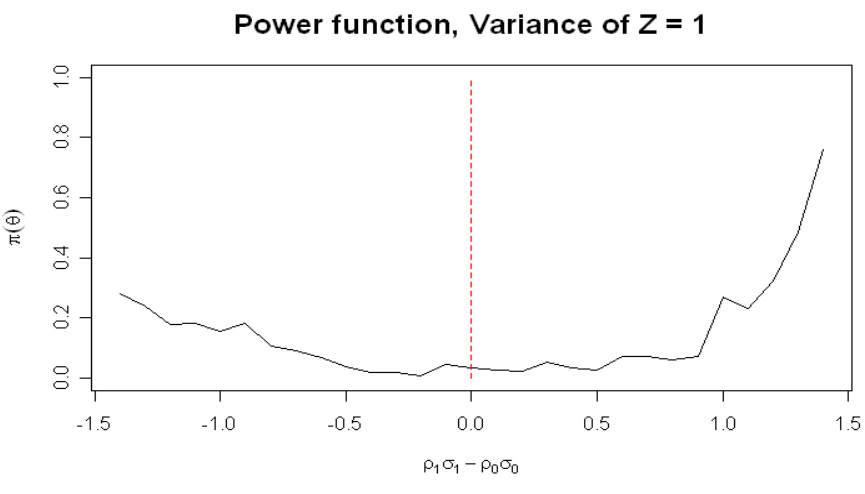
Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)



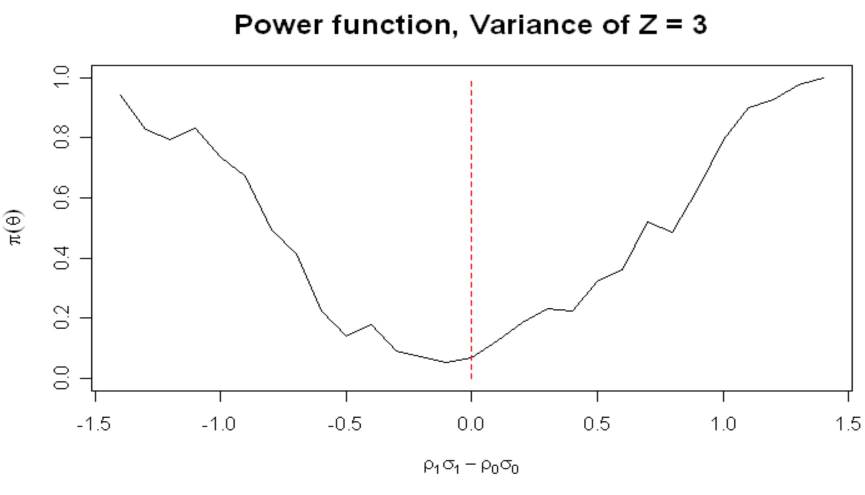
Power of the Test of Equality of IV Estimates Using Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates Using Propensity Scores Above and Below the Median, Varying σ_Z

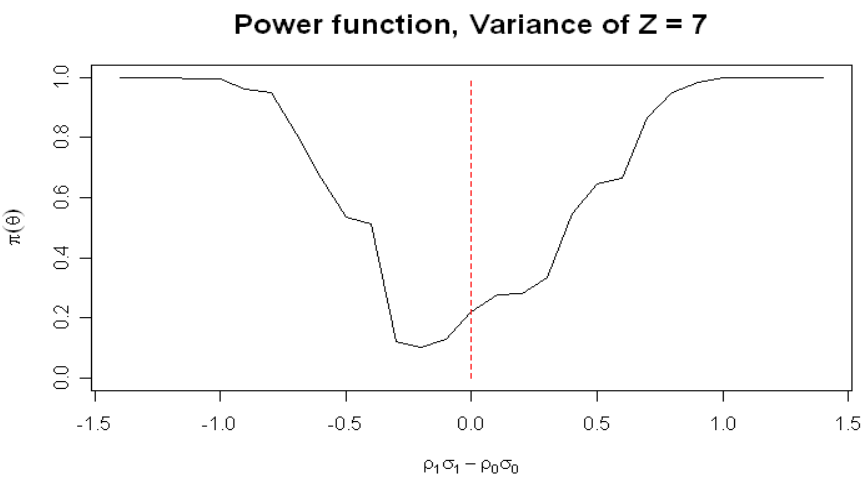
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates Using

Propensity Scores Above and Below the Median, Varying σ_Z

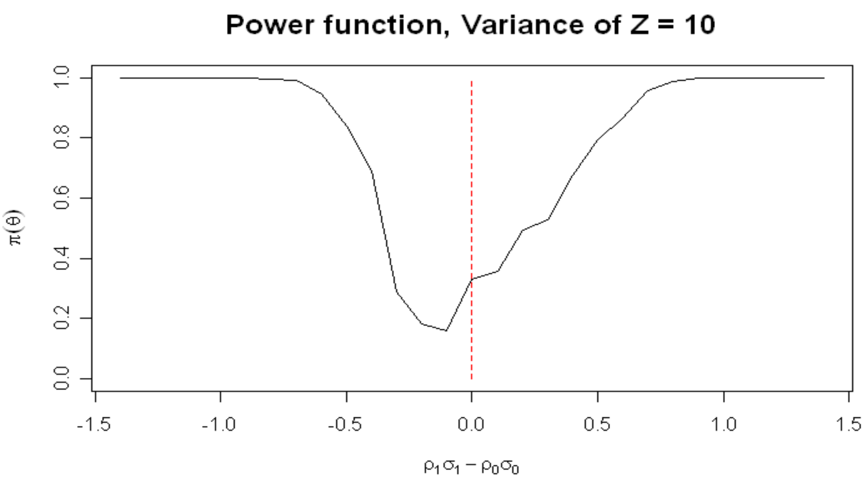
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates Using

Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of IV Estimates Using Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)

- The sample size is 5,000.

Power of the Test of Equality of IV Estimates Using

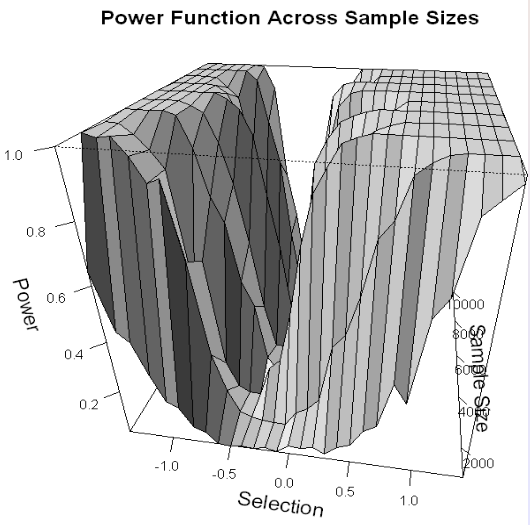
Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)

- The sample size is 5,000.
- The power is calculated for each alternative hypothesis (each value of $\rho_{1v}\sigma_1 - \rho_{0v}\sigma_0$) by bootstrapping the Wald statistic 500 times calculating what proportion of the test statistics lie outside the 95th percentile of a χ^2 distribution with 3 degrees of freedom (we are testing 3 coefficients).

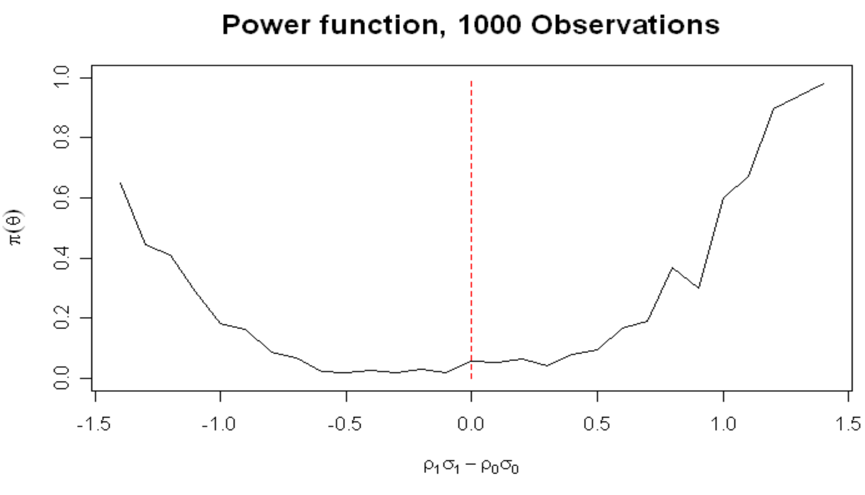
Power of the Test of Equality of Simple IV Estimates Using Propensity Scores Above and Below the Median, Varying Sample Size

(Using Chi-Square Distribution under the Null)



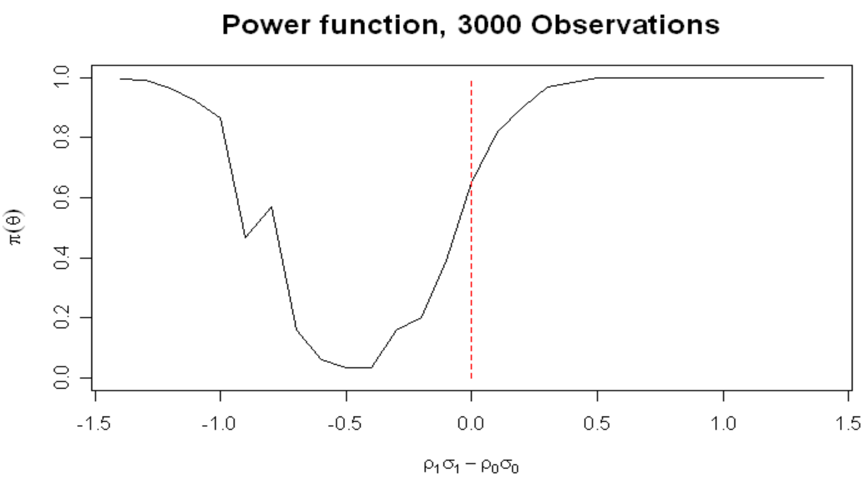
Power of the Test of Equality of Simple IV Estimates Using Propensity Scores Above and Below the Median, Varying Sample Size

(Using Chi-Square Distribution under the Null)



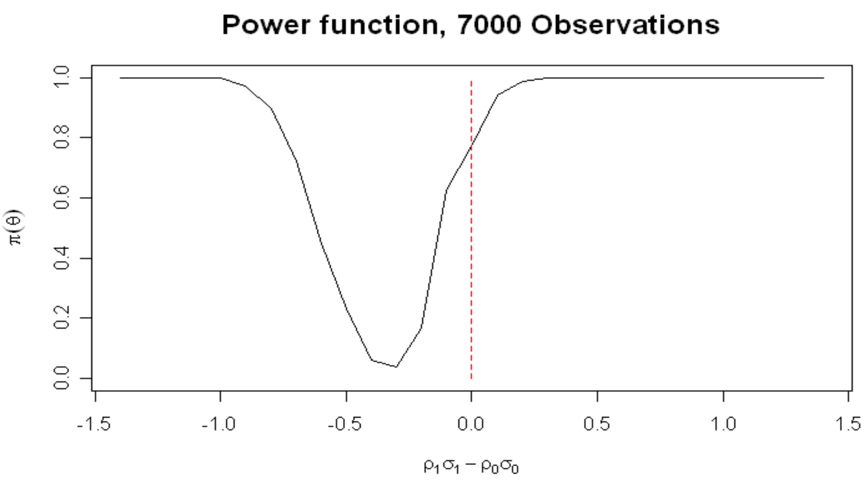
Power of the Test of Equality of Simple IV Estimates Using
Propensity Scores Above and Below the Median, Varying Sample Size

(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of Simple IV Estimates Using Propensity Scores Above and Below the Median, Varying Sample Size

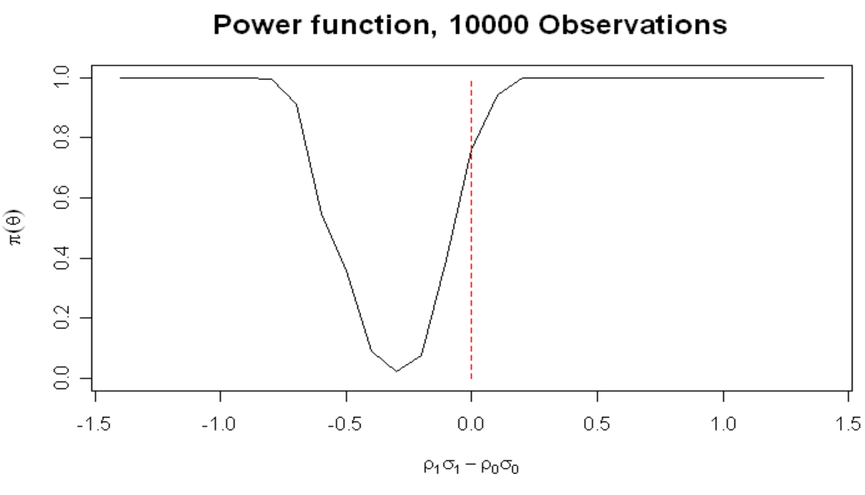
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of Simple IV Estimates Using

Propensity Scores Above and Below the Median, Varying Sample Size

(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of Simple IV Estimates Using Propensity Scores Above and Below the Median, Varying Sample Size

(Using Chi-Square Distribution under the Null)

- The variance of the instrument is 10.

Power of the Test of Equality of Simple IV Estimates Using

Propensity Scores Above and Below the Median, Varying Sample Size

(Using Chi-Square Distribution under the Null)

- The variance of the instrument is 10.
- The power is calculated for each alternative hypothesis (each value of $\rho_{1v}\sigma_1 - \rho_{0v}\sigma_0$) by bootstrapping the Wald statistic 500 times calculating what proportion of the test statistics lie outside the 95th percentile of a χ^2 distribution with 1 degree of freedom (we are testing 1 coefficient). These IV estimates are the coefficient on D and contain no interactions with X (so they are misspecified).

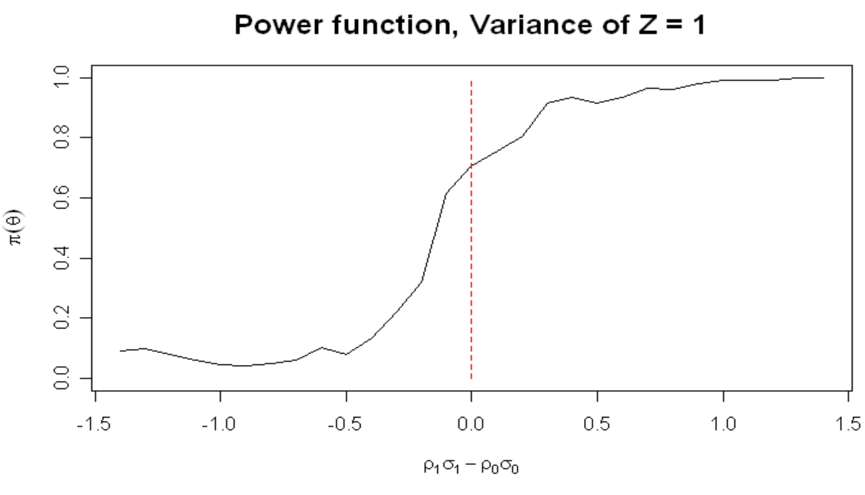
Power of the Test of Equality of Simple IV Estimates Using

Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)

Power of the Test of Equality of Simple IV Estimates Using
Propensity Scores Above and Below the Median, Varying σ_Z

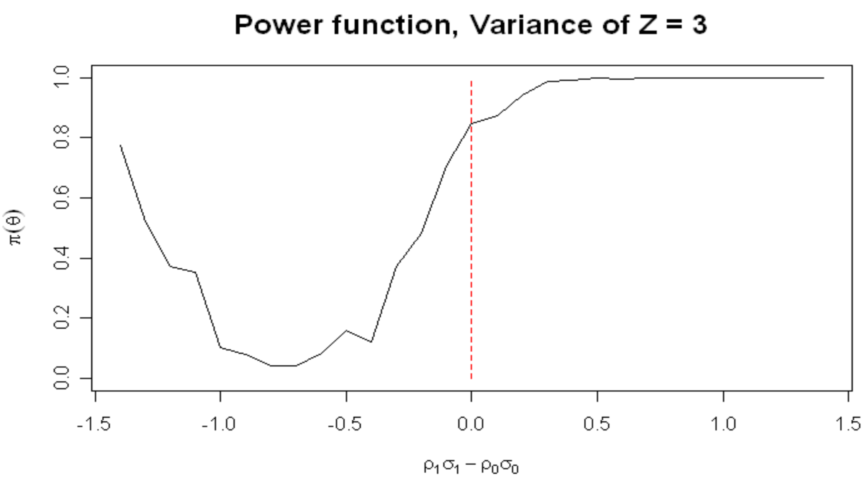
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of Simple IV Estimates Using

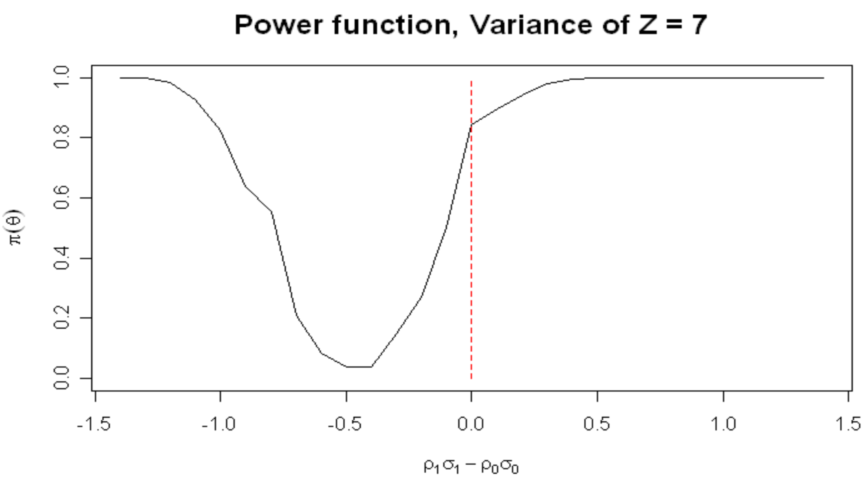
Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of Simple IV Estimates Using
Propensity Scores Above and Below the Median, Varying σ_Z

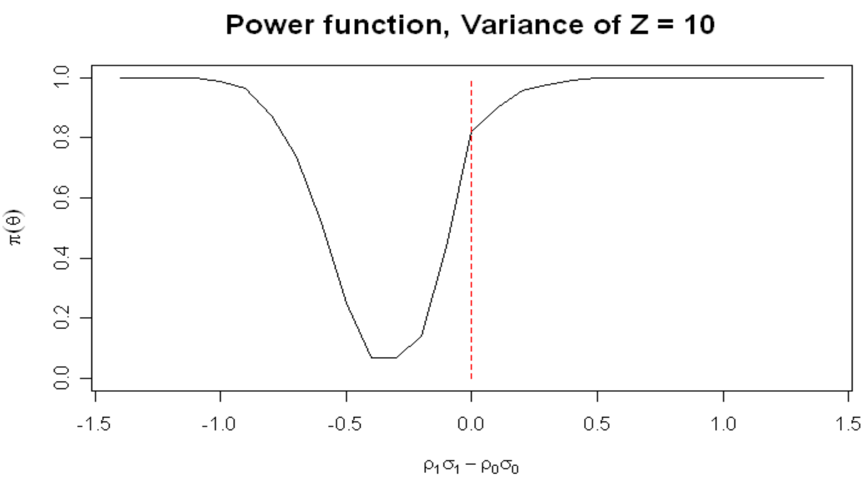
(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of Simple IV Estimates Using

Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)



Power of the Test of Equality of Simple IV Estimates Using

Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)

- The sample size is 5,000.

Power of the Test of Equality of Simple IV Estimates Using

Propensity Scores Above and Below the Median, Varying σ_Z

(Using Chi-Square Distribution under the Null)

- The sample size is 5,000.
- The power is calculated for each alternative hypothesis (each value of $\rho_{1v}\sigma_1 - \rho_{0v}\sigma_0$) by bootstrapping the Wald statistic 500 times calculating what proportion of the test statistics lie outside the 95th percentile of a χ^2 distribution with 1 degree of freedom (we are testing 1 coefficient). These IV estimates are the coefficient on D and contain no interactions with X (so they are misspecified).

Applying the Tests to the Data

- We implement our method of testing for essential heterogeneity in a wide variety of settings to show how ubiquitous the concept is.

- We implement our method of testing for essential heterogeneity in a wide variety of settings to show how ubiquitous the concept is.
- We consider some examples from labor economics, including the choices of college graduation, high school graduation, GED certification, and union membership, as well as an example from education, namely the effect of school vouchers on test scores.

Table 1: Specification of the Generalized Roy Model Used To Calculate the Power of the Tests

Outcomes	Decision Rule:
$Y_0 = \alpha_0 + \beta_{10}X_1 + \beta_{20}X_2 + U_0$	$D = \mathbf{1}(\alpha_d + \gamma_d Z \geq V)$
$Y_1 = \alpha_1 + \beta_{11}X_1 + \beta_{21}X_2 + U_1$	

Observed $Y = DY_1 + (1 - D)Y_0$

with parameters:	with parameters:
$\alpha_0 = 0,\ \beta_{10} = 0.1,\ \beta_{20} = 0.3$	$\alpha_d = 0.2$
$\alpha_1 = 0.2,\ \beta_{11} = 0.2,\ \beta_{21} = 0.4$	$\gamma_d = 0.3$

Table 1: Specification of the Generalized Roy Model Used To Calculate the Power of the Tests

Distribution of Unobservables:

$$\begin{pmatrix} U_1 \\ U_0 \\ V \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \\ 0 \end{matrix}, \begin{pmatrix} 1 & 0 & \rho_{1V} \\ 0 & 1 & -\rho_{1V} \\ \rho_{1V} & -\rho_{1V} & 1 \end{pmatrix} \right)$$

The power function is traced out by varying ρ_{1V} from -0.7 to 0.7. Values outside this interval lead to a covariance matrix which is not positive definite.

Table 1: Specification of the Generalized Roy Model Used To Calculate the Power of the Tests

Distribution of Observables:

$$\begin{pmatrix} X_1 \\ X_2 \\ Z \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \\ 0 \end{matrix}, \begin{pmatrix} 1 & 0.5 * \sqrt{10} & 0.5 * \sigma_Z \\ 0.5 * \sqrt{10} & 10 & 0.5 * \sigma_Z \\ 0.5 * \sigma_Z & 0.5 * \sigma_Z & \sigma_Z^2 \end{pmatrix} \right)$$

The power function calculated for values of σ_Z^2 between 1 and 10.

School Vouchers

- The 1981 reforms decentralized the administration of public schools, and established certain privately-run schools which received a fixed per-pupil payment from the government.

School Vouchers

- The 1981 reforms decentralized the administration of public schools, and established certain privately-run schools which received a fixed per-pupil payment from the government.
- The data comes from the Sistema de Medición de la Calidad de la Educación (SIMCE), which is a national standardized test administered once a year to students entering the 4^{th} grade, the 8^{th} grade and the 10^{th} grade.

School Vouchers

- The 1981 reforms decentralized the administration of public schools, and established certain privately-run schools which received a fixed per-pupil payment from the government.
- The data comes from the Sistema de Medición de la Calidad de la Educación (SIMCE), which is a national standardized test administered once a year to students entering the 4^{th} grade, the 8^{th} grade and the 10^{th} grade.
- We have data on 56,213 individual students.

- The binary choice in this setting is whether a student attends a voucher school ($D = 1$) or a public school ($D = 0$).

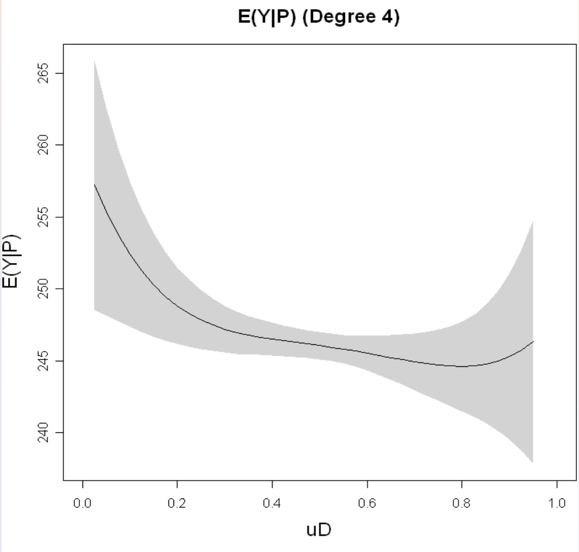
- The binary choice in this setting is whether a student attends a voucher school ($D = 1$) or a public school ($D = 0$).
- Instruments: (Z): number of family members in the household, the quality of the infrastructure of the school, indicators for various household income categories, mother's highest grade completed, father's highest grade completed, and region indicators.

- The binary choice in this setting is whether a student attends a voucher school ($D = 1$) or a public school ($D = 0$).
- Instruments: (Z): number of family members in the household, the quality of the infrastructure of the school, indicators for various household income categories, mother's highest grade completed, father's highest grade completed, and region indicators.
- The test we are using as our outcome measure has two components – a math score and a verbal score.

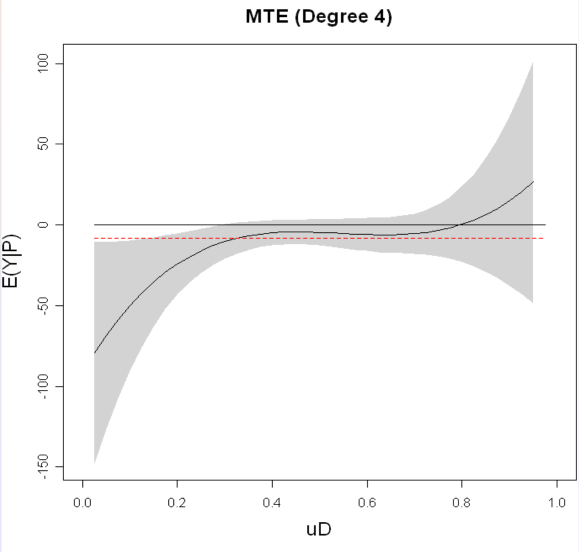
- The binary choice in this setting is whether a student attends a voucher school ($D = 1$) or a public school ($D = 0$).
- Instruments: (Z): number of family members in the household, the quality of the infrastructure of the school, indicators for various household income categories, mother's highest grade completed, father's highest grade completed, and region indicators.
- The test we are using as our outcome measure has two components – a math score and a verbal score.
- We define as our outcome the average of the two scores.

- We regress the outcome on the following controls (X) in addition to polynomial terms in the propensity score P: gender, mother's age, father's age, indicators for various household income categories, mother's highest grade completed, father's highest grade completed, indicators for the number of books in the household, whether the child attended a preschool, how hard the child studies, whether the child has a job, whether the parents attend meetings with the child's teachers, whether the parents regularly communicate with the child's teachers, whether the parents participate in the schooling of the child, whether the child's school helps economically disadvantaged students, whether the child's math teacher adequately prepared the child, whether the child's language teacher adequately prepared the child, an indicator for urban, and region indicators.

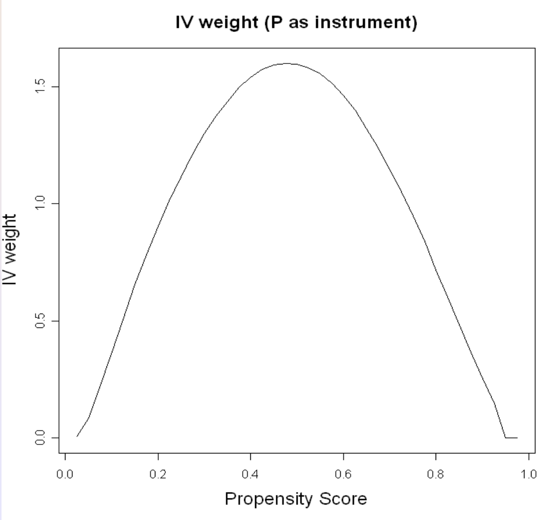
Chile School Vouchers – Math Score



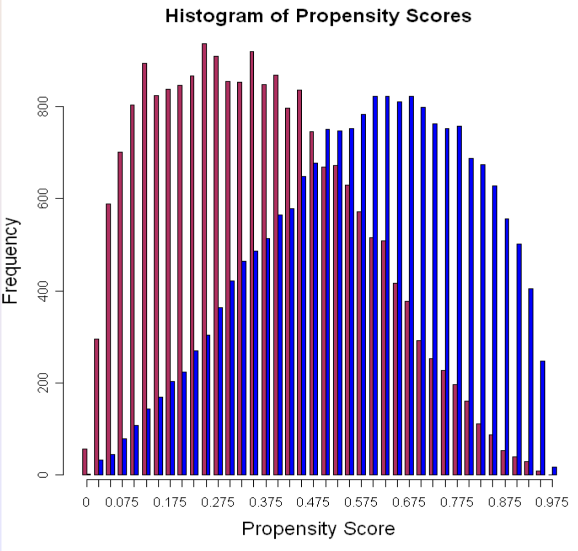
Chile School Vouchers – Math Score



Chile School Vouchers – Math Score



Chile School Vouchers – Math Score



Chile School Vouchers – Math Score

- The covariates in the outcome equations are: gender, mother's highest grade completed, father's highest grade completed, number of family members, an indicator for urban residence, household income categories and region indicators.

Chile School Vouchers – Math Score

- The covariates in the outcome equations are: gender, mother's highest grade completed, father's highest grade completed, number of family members, an indicator for urban residence, household income categories and region indicators.
- The instruments are: the proportion of schools in one's municipality that were voucher schools in 2002, the difference in average test scores between the voucher schools and the public schools in one's municipality in 2002, in addition to all of the X variables.

Chile School Vouchers – Math Score

- The covariates in the outcome equations are: gender, mother's highest grade completed, father's highest grade completed, number of family members, an indicator for urban residence, household income categories and region indicators.
- The instruments are: the proportion of schools in one's municipality that were voucher schools in 2002, the difference in average test scores between the voucher schools and the public schools in one's municipality in 2002, in addition to all of the X variables.
- The dependent variable in the probit is 1 if the individual is enrolled in a voucher school, and 0 if the individual is enrolled in a public school.

Chile School Vouchers – Math Score

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.

Chile School Vouchers – Math Score

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.

Chile School Vouchers – Math Score

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.

Chile School Vouchers – Math Score

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.

Chile School Vouchers – Math Score

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.
- The sample size is 40,501.

Chile Vouchers – Math Score

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5

Chile Vouchers – Math Score

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0000	0.0000	0.0014	0.0696

Chile Vouchers – Math Score

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0000	0.0000	0.0014	0.0696
P^2	0.3550	0.0890	0.0395	0.3885

Chile Vouchers – Math Score

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0000	0.0000	0.0014	0.0696
P^2	0.3550	0.0890	0.0395	0.3885
P^3		0.1593	0.0767	0.5795

Chile Vouchers – Math Score

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0000	0.0000	0.0014	0.0696
P^2	0.3550	0.0890	0.0395	0.3885
P^3		0.1593	0.0767	0.5795
P^4			0.1208	0.7376

Chile Vouchers – Math Score

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0000	0.0000	0.0014	0.0696
P^2	0.3550	0.0890	0.0395	0.3885
P^3		0.1593	0.0767	0.5795
P^4			0.1208	0.7376
P^5				0.8667

Chile Vouchers – Math Score

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0000	0.0000	0.0014	0.0696
P^2	0.3550	0.0890	0.0395	0.3885
P^3		0.1593	0.0767	0.5795
P^4			0.1208	0.7376
P^5				0.8667
Joint test of nonlinear terms	0.3550	0.0875	0.0205	0.8796

Chile Vouchers – Math Score

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.

Chile Vouchers – Math Score

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-7.2307	-12.6136	-10.9667	-11.7628	-4.7516	-7.7067

Chile Vouchers – Math Score

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-7.2307	-12.6136	-10.9667	-11.7628	-4.7516	-7.7067
TT	-12.9179	-18.6105	-28.4809	-29.4113	-8.5286	-12.924

Chile Vouchers – Math Score

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-7.2307	-12.6136	-10.9667	-11.7628	-4.7516	-7.7067
TT	-12.9179	-18.6105	-28.4809	-29.4113	-8.5286	-12.924
TUT	-0.7015	-9.5799	5.5403	4.1173	-0.6851	-5.9742

Chile Vouchers – Math Score

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-7.2307	-12.6136	-10.9667	-11.7628	-4.7516	-7.7067
TT	-12.9179	-18.6105	-28.4809	-29.4113	-8.5286	-12.924
TUT	-0.7015	-9.5799	5.5403	4.1173	-0.6851	-5.9742
IV	-8.0559	-8.0559	-8.0559	-8.0559	-8.0559	-8.0559

Chile Vouchers – Math Score

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-7.2307	-12.6136	-10.9667	-11.7628	-4.7516	-7.7067
TT	-12.9179	-18.6105	-28.4809	-29.4113	-8.5286	-12.924
TUT	-0.7015	-9.5799	5.5403	4.1173	-0.6851	-5.9742
IV	-8.0559	-8.0559	-8.0559	-8.0559	-8.0559	-8.0559
IV (using weights)	-8.3914	-9.0786	-10.9789	-11.2349	-5.4911	-4.2007

Chile Vouchers – Math Score

Treatment Effects

- The P-values in the first panel are from t-tests in the case of the individual coefficients and Wald tests for the joint tests.

Chile Vouchers – Math Score

Treatment Effects

- The P-values in the first panel are from t-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.

Chile Vouchers – Math Score

Treatment Effects

- The P-values in the first panel are from t-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.
- The treatment effects in the second panel are calculated by weighting the estimated MTE by the weights from Heckman and Vytlacil (2005).

Chile Vouchers – Math Score

Treatment Effects

- The P-values in the first panel are from t-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.
- The treatment effects in the second panel are calculated by weighting the estimated MTE by the weights from Heckman and Vytlacil (2005).
- Therefore, they vary depending on the degree of the polynomial used to approximate $E(Y \mid P)$ (and hence the polynomial used to approximate the MTE).

Chile Vouchers – Math Score

Treatment Effects

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.

Chile Vouchers – Math Score

Treatment Effects

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.

Chile Vouchers – Math Score

Treatment Effects

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.
- The estimates differ not only because the estimate of the MTE is inexact, but also because the weights are estimated.

Chile Vouchers – Math Score

Treatment Effects

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.
- The estimates differ not only because the estimate of the MTE is inexact, but also because the weights are estimated.
- In both panels the degree of the polynomial refers to the degree used to approximate $E(Y \mid P)$ (the degree of the approximation to the MTE is one less).

Union Membership

- We focus on replicating the analysis of Lee(1978).

- We focus on replicating the analysis of Lee(1978).
- We use the Panel Study on Income Dynamics (PSID) at a cross-section in 1988.

- We focus on replicating the analysis of Lee(1978).
- We use the Panel Study on Income Dynamics (PSID) at a cross-section in 1988.
- The outcome variable is (log) weekly wages.

- We focus on replicating the analysis of Lee(1978).
- We use the Panel Study on Income Dynamics (PSID) at a cross-section in 1988.
- The outcome variable is (log) weekly wages.
- The binary choice that agents face is whether to be a union member or not.

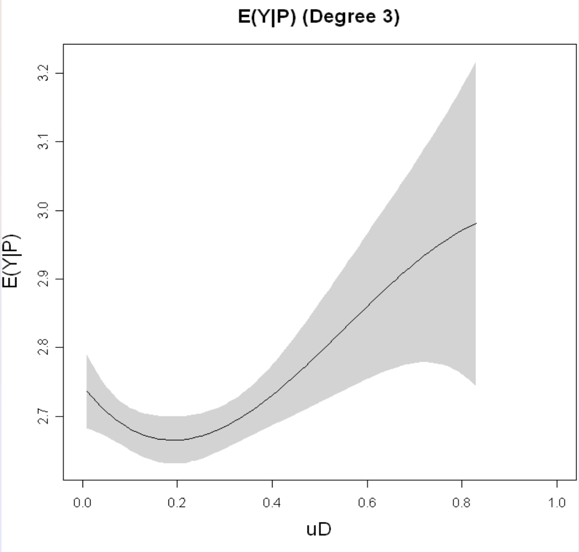
- We focus on replicating the analysis of Lee(1978).
- We use the Panel Study on Income Dynamics (PSID) at a cross-section in 1988.
- The outcome variable is (log) weekly wages.
- The binary choice that agents face is whether to be a union member or not.
- We include only men between the ages of 18 and 65 who are not enrolled in school and who worked at least one week in the previous year.

- We focus on replicating the analysis of Lee(1978).
- We use the Panel Study on Income Dynamics (PSID) at a cross-section in 1988.
- The outcome variable is (log) weekly wages.
- The binary choice that agents face is whether to be a union member or not.
- We include only men between the ages of 18 and 65 who are not enrolled in school and who worked at least one week in the previous year.
- Sample size of $N = 4,081$.

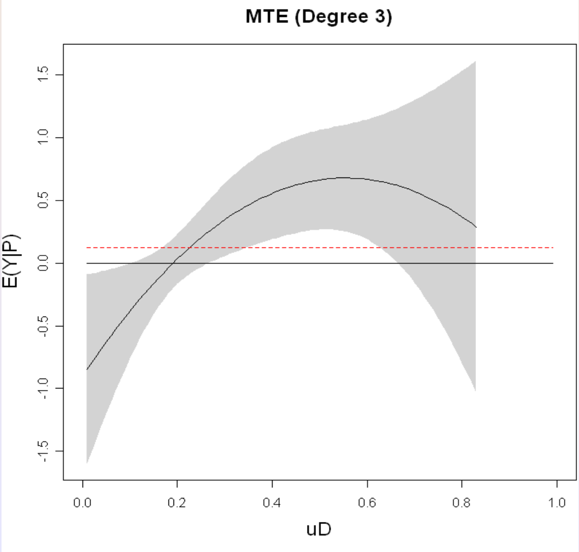
- The intruments (Z) are indicators for residence in the northeast, midwest, south, and a metropolitan area of at least 250,000; indicators for years of education categories: 1 to 7 years, 9 to 11 years, 12 years, and 13 or more years; experience, experience squared, and indicator for white; indicators for having worked 1 to 26 weeks in the previous years and 48 to 52 weeks in the previous year; and indicators for the occupations: mining, construction, manufacturing durable goods, and manufacturing non-durable goods.

- We regress log weekly wages on controls (X), plus polynomial terms in P: indicators for residence in the northeast, midwest, south, and a metropolitan area of at least 250,000; indicators for years of education categories: 1 to 7 years, 9 to 11 years, 12 years, and 13 or more years; experience, experience squared, and indicator for white; and indicators for having worked 1 to 26 weeks in the previous years and 48 to 52 weeks in the previous year.

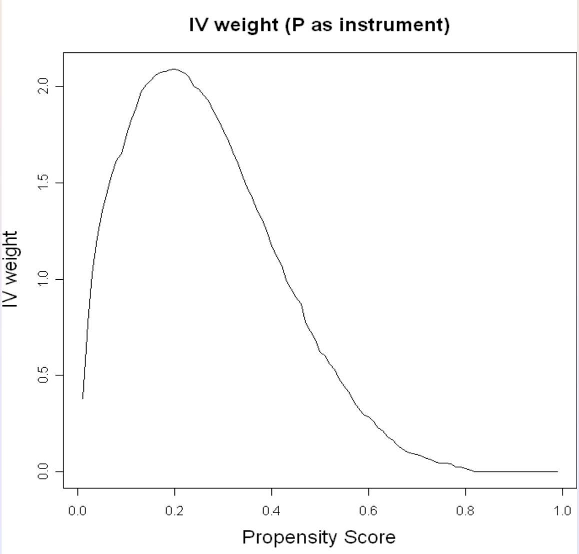
Union Wages



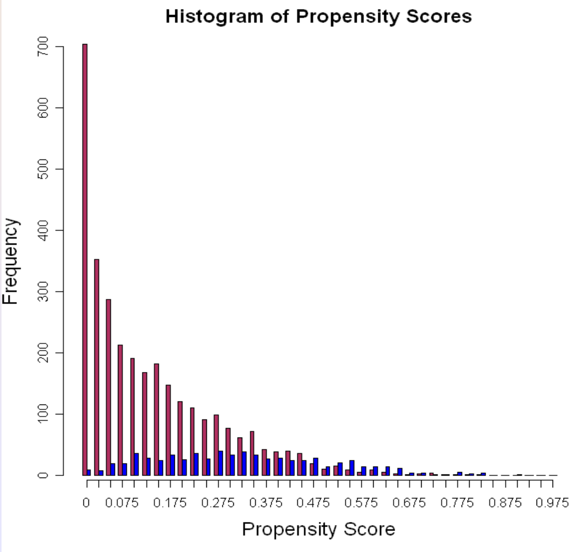
Union Wages



Union Wages


IV weight (P as instrument)
IV weight
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Propensity Score

Union Wages


Histogram of Propensity Scores
Frequency
0
100
200
300
400
500
600
700
0
0.075
0.175
0.275
0.375
0.475
0.575
0.675
0.775
0.875
0.975
Propensity Score

Union Wages

- The covariates in the outcome equations are: experience, experience squared, various education categories, indicators for region of the country, indicator for urban, indicator for white, indicator for weeks worked between 1 and 26, indicator for weeks worked between 48 and 52 weeks.

Union Wages

- The covariates in the outcome equations are: experience, experience squared, various education categories, indicators for region of the country, indicator for urban, indicator for white, indicator for weeks worked between 1 and 26, indicator for weeks worked between 48 and 52 weeks.
- The instruments are: all of the X variables in addition to indicators for two-digit occupation codes.

Union Wages

- The covariates in the outcome equations are: experience, experience squared, various education categories, indicators for region of the country, indicator for urban, indicator for white, indicator for weeks worked between 1 and 26, indicator for weeks worked between 48 and 52 weeks.
- The instruments are: all of the X variables in addition to indicators for two-digit occupation codes.
- The dependent variable in the probit is 1 if the individual is a union member, and 0 if the individual is not a union member.

Union Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.

Union Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.
- The confidence intervals are found using 100 bootstraps.

Union Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.

Union Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.

Union Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.
- The sample size is 3815.

Union Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5

Union Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0096	0.1606	0.2803	0.4140

Union Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0096	0.1606	0.2803	0.4140
P^2	0.0041	0.0302	0.5305	0.9094

Union Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0096	0.1606	0.2803	0.4140
P^2	0.0041	0.0302	0.5305	0.9094
P^3		0.1065	0.8824	0.9146

Union Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0096	0.1606	0.2803	0.4140
P^2	0.0041	0.0302	0.5305	0.9094
P^3		0.1065	0.8824	0.9146
P^4			0.9412	0.8728

Union Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0096	0.1606	0.2803	0.4140
P^2	0.0041	0.0302	0.5305	0.9094
P^3		0.1065	0.8824	0.9146
P^4			0.9412	0.8728
P^5				0.8724

Union Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.0096	0.1606	0.2803	0.4140
P^2	0.0041	0.0302	0.5305	0.9094
P^3		0.1065	0.8824	0.9146
P^4			0.9412	0.8728
P^5				0.8724
Joint test of nonlinear terms	0.0041	0.0144	0.0294	0.0311

Union Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.

Union Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.6215	0.2437	0.3192	1.8325	0.2149	0.1510

Union Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.6215	0.2437	0.3192	1.8325	0.2149	0.1510
TT	-0.1187	-0.2225	-0.2323	-0.1458	-0.0959	-0.0413

Union Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.6215	0.2437	0.3192	1.8325	0.2149	0.1510
TT	-0.1187	-0.2225	-0.2323	-0.1458	-0.0959	-0.0413
TUT	0.8512	0.3645	0.4683	2.4845	0.3083	0.2153

Union Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.6215	0.2437	0.3192	1.8325	0.2149	0.1510
TT	-0.1187	-0.2225	-0.2323	-0.1458	-0.0959	-0.0413
TUT	0.8512	0.3645	0.4683	2.4845	0.3083	0.2153
IV	0.1249	0.1249	0.1249	0.1249	0.1249	0.1249

Union Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.6215	0.2437	0.3192	1.8325	0.2149	0.1510
TT	-0.1187	-0.2225	-0.2323	-0.1458	-0.0959	-0.0413
TUT	0.8512	0.3645	0.4683	2.4845	0.3083	0.2153
IV	0.1249	0.1249	0.1249	0.1249	0.1249	0.1249
IV (using weights)	0.0593	0.0498	0.0487	0.0566	0.0031	0.0064

Union Wages

P-values from sequentially adding polynomial terms

- The *p*-values in the first panel are from *t*-tests in the case of the individual coefficients and Wald tests for the joint tests.

Union Wages

P-values from sequentially adding polynomial terms

- The *p*-values in the first panel are from *t*-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.

Union Wages

P-values from sequentially adding polynomial terms

- The *p*-values in the first panel are from *t*-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.
- The treatment effects in the second panel are calculated by weighting the estimated MTE by the weights from Heckman and Vytlacil (2005).

Union Wages

P-values from sequentially adding polynomial terms

- The p-values in the first panel are from t-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.
- The treatment effects in the second panel are calculated by weighting the estimated MTE by the weights from Heckman and Vytlacil (2005).
- Therefore, they vary depending on the degree of the polynomial used to approximate $E(Y \mid P)$ (and hence the polynomial used to approximate the MTE).

Union Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.

Union Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.

Union Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.
- The estimates differ not only because the estimate of the MTE is inexact, but also because the weights are estimated.

Union Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.
- The estimates differ not only because the estimate of the MTE is inexact, but also because the weights are estimated.
- In both panels the degree of the polynomial refers to the degree used to approximate $E(Y \mid P)$ (the degree of the approximation to the MTE is one less).

- The data comes from the National Longitudinal Survey of Youth 1979 (NLSY79).

- The data comes from the National Longitudinal Survey of Youth 1979 (NLSY79).
- We include only 30-year-old men from the “core” sample.

GED Recipient vs. High School Dropout

- $D = 1$ if the individual is a GED recipient and $D = 0$ if the individual is a high school dropout.

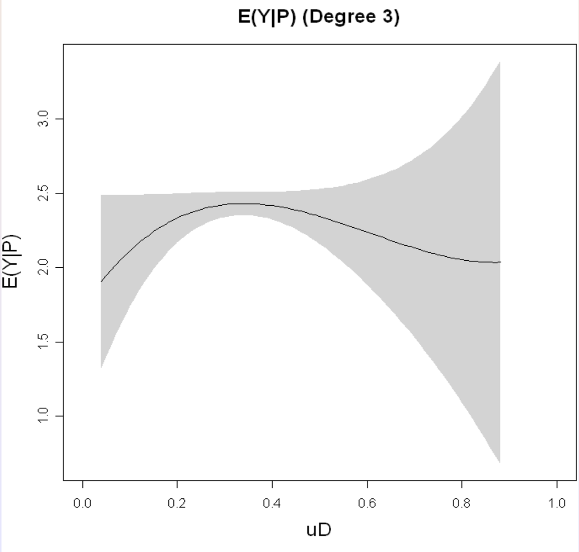
- $D = 1$ if the individual is a GED recipient and $D = 0$ if the individual is a high school dropout.
- This leads to a sample size of 409.

- $D = 1$ if the individual is a GED recipient and $D = 0$ if the individual is a high school dropout.
- This leads to a sample size of 409.
- Outcome: average of log weekly wages at ages 29, 30 and 31.

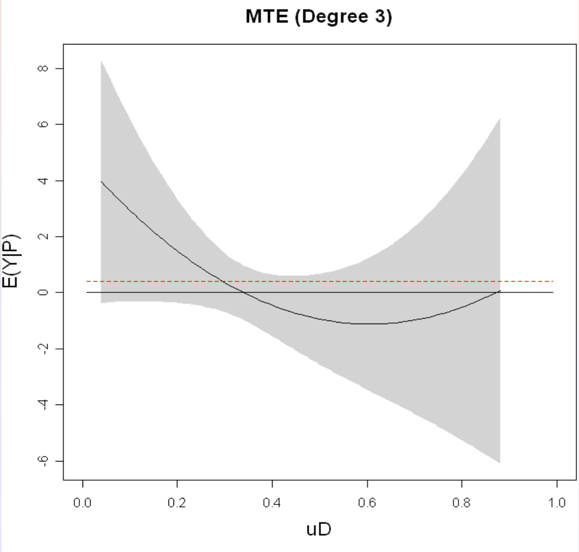
- Instruments (Z): standardized AFQT score, father's highest grade completed, mother's highest grade completed, number of siblings, family income in 1979, cost of GED, wages of local high school dropouts, unemployment of local high school graduates, indicators for black, hispanic, residence in the south at age 14, residence in an urban area at age 14 and year of birth indicators.

- We regress the outcome variable on polynomial terms in P, in addition to the controls (X): job tenure, job tenure squared, experience, standardized AFQT score, standardized noncognitive test scores, highest grade completed, and indicators for black, hispanic and being married.

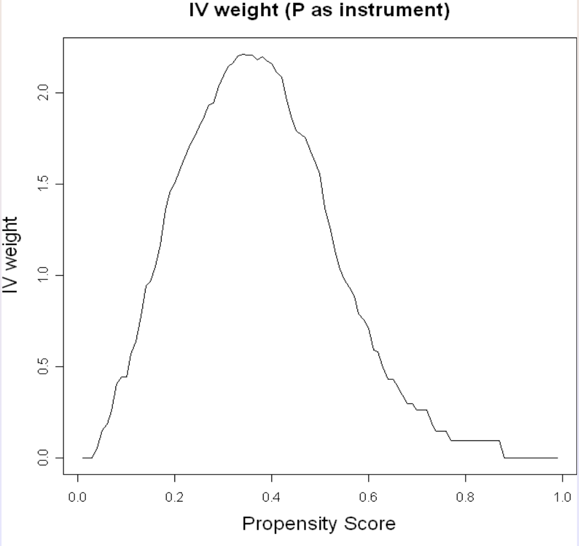
GED vs. Dropout Wages



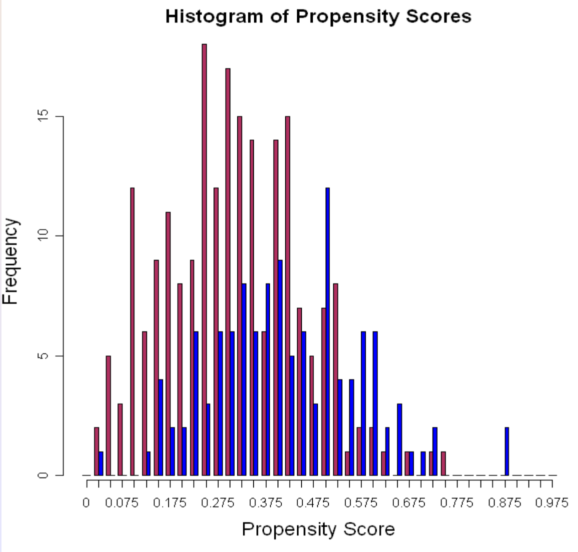
GED vs. Dropout Wages



GED vs. Dropout Wages



GED vs. Dropout Wages



GED vs. Dropout Wages

- The covariates in the outcome equations are: job tenure, job tenure squared, AFQT score, noncognitive score, marital status, indicators for black and hispanic, and year of birth indicators.

GED vs. Dropout Wages

- The covariates in the outcome equations are: job tenure, job tenure squared, AFQT score, noncognitive score, marital status, indicators for black and hispanic, and year of birth indicators.
- The instruments are: AFQT score, noncognitive score, father's highest grade completed, mother's highest grade completed, number of siblings, family income in 1979, local cost of the GED, wages of local dropouts, unemployment rates of local high school graduates, indicators for black and hispanic, indicators for south residence and urban residence at age 14, and year of birth indicators.

GED vs. Dropout Wages

- The covariates in the outcome equations are: job tenure, job tenure squared, AFQT score, noncognitive score, marital status, indicators for black and hispanic, and year of birth indicators.
- The instruments are: AFQT score, noncognitive score, father's highest grade completed, mother's highest grade completed, number of siblings, family income in 1979, local cost of the GED, wages of local dropouts, unemployment rates of local high school graduates, indicators for black and hispanic, indicators for south residence and urban residence at age 14, and year of birth indicators.
- The dependent variable in the probit is 1 if the individual's highest education is a GED, and 0 if the individual is a high school dropout.

GED vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.

GED vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.
- The confidence intervals are found using 100 bootstraps.

GED vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.

GED vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.

GED vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2 and P^3.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.
- The sample size is 331.

GED vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5

GED vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.4640	0.4123	0.6893	0.9723

GED vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.4640	0.4123	0.6893	0.9723
P^2	0.6947	0.5529	0.9544	0.8231

GED vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.4640	0.4123	0.6893	0.9723
P^2	0.6947	0.5529	0.9544	0.8231
P^3		0.6469	0.9052	0.7851

GED vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.4640	0.4123	0.6893	0.9723
P^2	0.6947	0.5529	0.9544	0.8231
P^3		0.6469	0.9052	0.7851
P^4			0.8311	0.7887

GED vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.4640	0.4123	0.6893	0.9723
P^2	0.6947	0.5529	0.9544	0.8231
P^3		0.6469	0.9052	0.7851
P^4			0.8311	0.7887
P^5				0.8061

GED vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.4640	0.4123	0.6893	0.9723
P^2	0.6947	0.5529	0.9544	0.8231
P^3		0.6469	0.9052	0.7851
P^4			0.8311	0.7887
P^5				0.8061
Joint test of nonlinear terms	0.6947	0.8205	0.9096	0.9631

GED vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.

GED vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.5278	0.3563	0.3088	-2.1581	0.0637	-0.0939

GED vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.5278	0.3563	0.3088	-2.1581	0.0637	-0.0939
TT	0.7162	1.7196	1.7843	0.2074	-0.0239	0.0960

GED vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.5278	0.3563	0.3088	-2.1581	0.0637	-0.0939
TT	0.7162	1.7196	1.7843	0.2074	-0.0239	0.0960
TUT	-1.4192	-0.4096	-0.5354	-4.1226	0.1216	-0.2121

GED vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.5278	0.3563	0.3088	-2.1581	0.0637	-0.0939
TT	0.7162	1.7196	1.7843	0.2074	-0.0239	0.0960
TUT	-1.4192	-0.4096	-0.5354	-4.1226	0.1216	-0.2121
IV	0.3934	0.3934	0.3934	0.3934	0.3934	0.3934

GED vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.5278	0.3563	0.3088	-2.1581	0.0637	-0.0939
TT	0.7162	1.7196	1.7843	0.2074	-0.0239	0.0960
TUT	-1.4192	-0.4096	-0.5354	-4.1226	0.1216	-0.2121
IV	0.3934	0.3934	0.3934	0.3934	0.3934	0.3934
IV (using weights)	0.5152	1.5113	1.5689	0.7140	-0.0089	0.0968

High School Diploma vs. High School Dropout

- $D = 1$ if an individual's highest level of education is a high school diploma and $D = 0$ if the individual is a high school dropout (not a GED recipient).

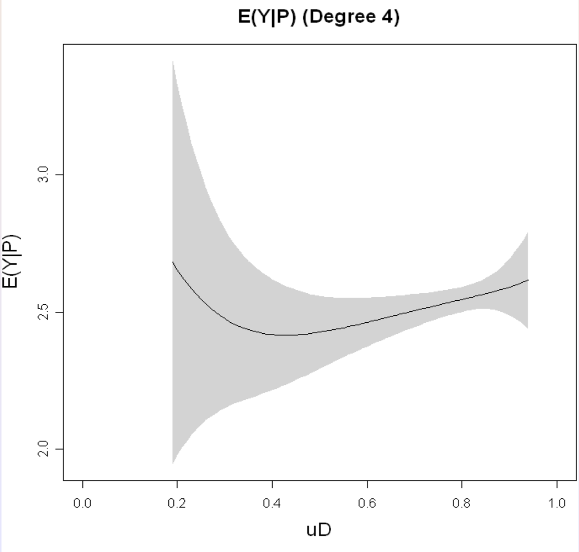
- $D = 1$ if an individual's highest level of education is a high school diploma and $D = 0$ if the individual is a high school dropout (not a GED recipient).
- This gives a sample size of 1083.

- $D = 1$ if an individual's highest level of education is a high school diploma and $D = 0$ if the individual is a high school dropout (not a GED recipient).
- This gives a sample size of 1083.
- The outcome variable is the average of log hourly wages at ages 29, 30, and 31.

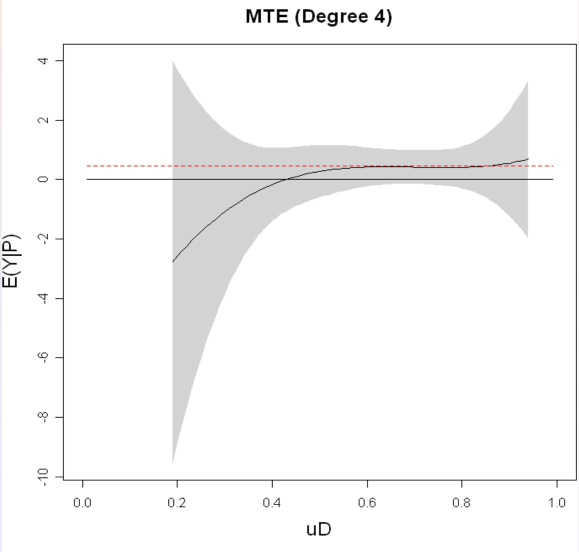
- Instrument (Z): standardized AFQT score, father's highest grade completed, mother's highest grade completed, number of siblings, family income in 1979, wages of local high school dropouts, wages of local high school graduates, unemployment of local high school dropouts, unemployment of local high school graduates, indicators for black, hispanic, residence in the south at age 14, residence in an urban area at age 14 and year of birth indicators.

- We regress the outcome variable on polynomials in P plus the regressors (X): job tenure, job tenure squared, experience, standardized AFQT score, standardized noncognitive test scores, highest grade completed, and indicators for black, hispanic and being married.

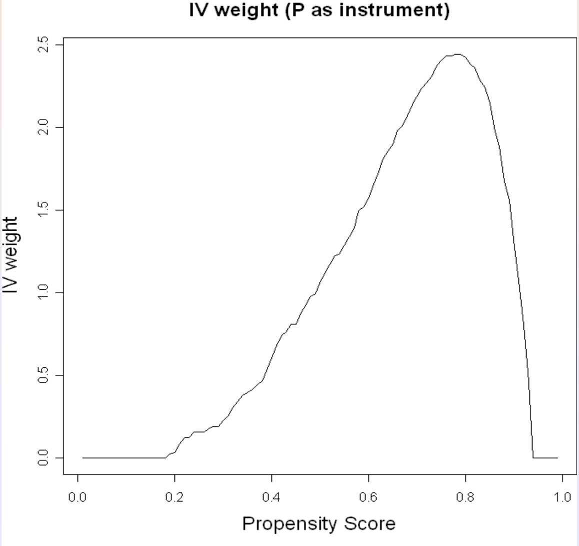
High School vs. Dropout Wages



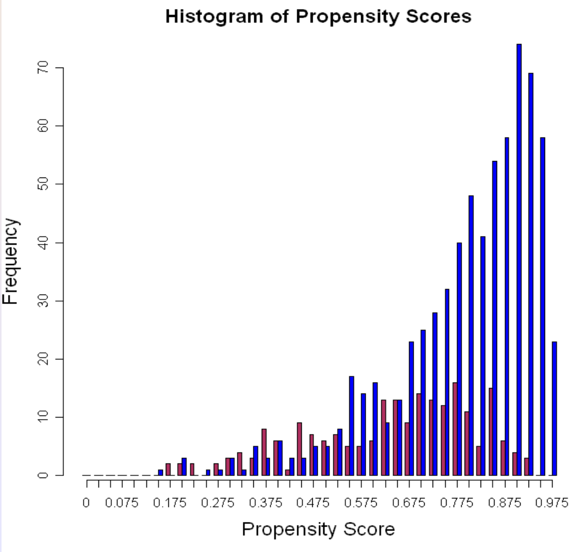
High School vs. Dropout Wages



High School vs. Dropout Wages



High School vs. Dropout Wages



High School vs. Dropout Wages

- The covariates in the outcome equations are: job tenure, job tenure squared, experience, experience squared, AFQT score, noncognitive score, marital status, indicators for black and hispanic, and year of birth indicators.

High School vs. Dropout Wages

- The covariates in the outcome equations are: job tenure, job tenure squared, experience, experience squared, AFQT score, noncognitive score, marital status, indicators for black and hispanic, and year of birth indicators.
- The instruments are: AFQT score, noncognitive score, father's highest grade completed, mother's highest grade completed, number of siblings, family income in 1979, wages and unemployment rates of local dropouts, wages and unemployment rates of local high school graduates, indicators for black and hispanic, indicators for south residence and urban residence at age 14, and year of birth indicators.

High School vs. Dropout Wages

- The covariates in the outcome equations are: job tenure, job tenure squared, experience, experience squared, AFQT score, noncognitive score, marital status, indicators for black and hispanic, and year of birth indicators.
- The instruments are: AFQT score, noncognitive score, father's highest grade completed, mother's highest grade completed, number of siblings, family income in 1979, wages and unemployment rates of local dropouts, wages and unemployment rates of local high school graduates, indicators for black and hispanic, indicators for south residence and urban residence at age 14, and year of birth indicators.
- The dependent variable in the probit is 1 if the individual's highest education is a high school diploma, and 0 if the individual is a high school dropout (GEDs are excluded).

High School vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.

High School vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.

High School vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.

High School vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.

High School vs. Dropout Wages

- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.
- The sample size is 1,144.

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.9766	0.5533	0.6785	0.8442

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.9766	0.5533	0.6785	0.8442
P^2	0.3789	0.4515	0.6739	0.8629

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.9766	0.5533	0.6785	0.8442
P^2	0.3789	0.4515	0.6739	0.8629
P^3		0.5520	0.7251	0.8877

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.9766	0.5533	0.6785	0.8442
P^2	0.3789	0.4515	0.6739	0.8629
P^3		0.5520	0.7251	0.8877
P^4			0.7626	0.9020

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.9766	0.5533	0.6785	0.8442
P^2	0.3789	0.4515	0.6739	0.8629
P^3		0.5520	0.7251	0.8877
P^4			0.7626	0.9020
P^5				0.9108

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.9766	0.5533	0.6785	0.8442
P^2	0.3789	0.4515	0.6739	0.8629
P^3		0.5520	0.7251	0.8877
P^4			0.7626	0.9020
P^5				0.9108
Joint test of nonlinear terms	0.3789	0.5188	0.8640	0.9384

High School Graduate vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.

High School Graduate vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.0017	-0.4598	-0.9649	-1.5510	0.2420	0.2398

High School Graduate vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.0017	-0.4598	-0.9649	-1.5510	0.2420	0.2398
TT	-0.2118	-0.7159	-1.4474	-2.1723	0.0915	0.2411

High School Graduate vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.0017	-0.4598	-0.9649	-1.5510	0.2420	0.2398
TT	-0.2118	-0.7159	-1.4474	-2.1723	0.0915	0.2411
TUT	0.8954	0.3059	0.7462	0.4588	1.0078	0.1933

High School Graduate vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.0017	-0.4598	-0.9649	-1.5510	0.2420	0.2398
TT	-0.2118	-0.7159	-1.4474	-2.1723	0.0915	0.2411
TUT	0.8954	0.3059	0.7462	0.4588	1.0078	0.1933
IV	0.4641	0.4641	0.4641	0.4641	0.4641	0.4641

High School Graduate vs. Dropout Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	-0.0017	-0.4598	-0.9649	-1.5510	0.2420	0.2398
TT	-0.2118	-0.7159	-1.4474	-2.1723	0.0915	0.2411
TUT	0.8954	0.3059	0.7462	0.4588	1.0078	0.1933
IV	0.4641	0.4641	0.4641	0.4641	0.4641	0.4641
IV (using weights)	0.2371	0.2605	0.2237	0.2236	0.3547	0.2636

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

- The *p*-values in the first panel are from *t*-tests in the case of the individual coefficients and Wald tests for the joint tests.

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

- The *p*-values in the first panel are from *t*-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

- The *p*-values in the first panel are from *t*-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.
- The treatment effects in the second panel are calculated by weighting the estimated MTE by the weights from Heckman and Vytlacil (2005).

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

- The p-values in the first panel are from t-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.
- The treatment effects in the second panel are calculated by weighting the estimated MTE by the weights from Heckman and Vytlacil (2005).
- Therefore, they vary depending on the degree of the polynomial used to approximate $E(Y \mid P)$ (and hence the polynomial used to approximate the MTE).

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.
- The estimates differ not only because the estimate of the MTE is inexact, but also because the weights are estimated.

High School Graduate vs. Dropout Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.
- The estimates differ not only because the estimate of the MTE is inexact, but also because the weights are estimated.
- In both panels the degree of the polynomial refers to the degree used to approximate $E(Y \mid P)$ (the degree of the approximation to the MTE is one less).

College Degree vs. High School Diploma

- We consider $D = 1$ if the individual is a college graduate and $D = 0$ if the individual's highest educational attainment is a high school diploma.

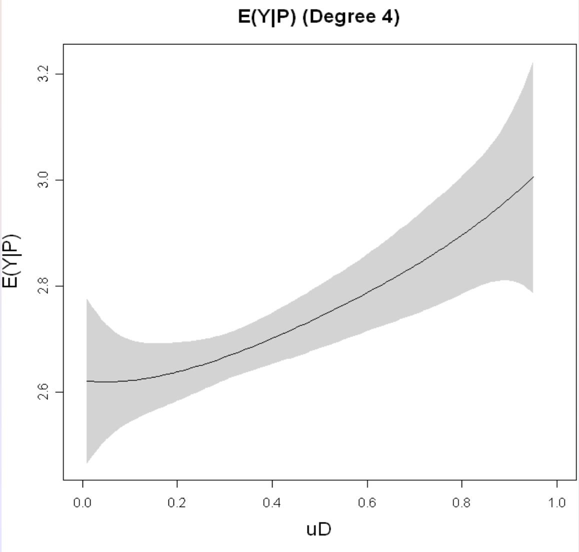
- We consider $D = 1$ if the individual is a college graduate and $D = 0$ if the individual's highest educational attainment is a high school diploma.
- This leads to a sample with 1335 observations.

- We consider $D = 1$ if the individual is a college graduate and $D = 0$ if the individual's highest educational attainment is a high school diploma.
- This leads to a sample with 1335 observations.
- The outcome variable is the average of log wages at ages 29, 30 and 31.

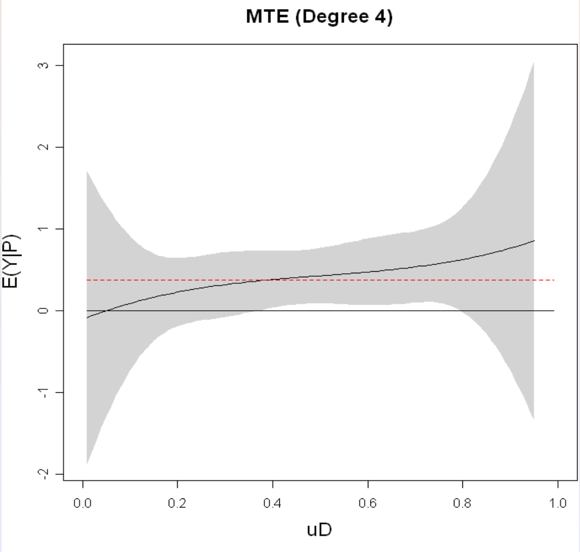
- Instruments (Z): standardized AFQT score, father's highest grade completed, mother's highest grade completed, number of siblings, family income in 1979, wages of local high school graduates, wages of local some college, wages of local college graduates, unemployment of local high school graduates, unemployment of local some college, unemployment of local college graduates, indicators for black, hispanic, residence in the south at age 14, residence in an urban area at age 14 and year of birth indicators.

- We regress the outcome variable on polynomials in the propensity score in addition to the control variables (X): job tenure, job tenure squared, experience, standardized AFQT score, standardized noncognitive test scores, highest grade completed, and indicators for black, hispanic and being married.

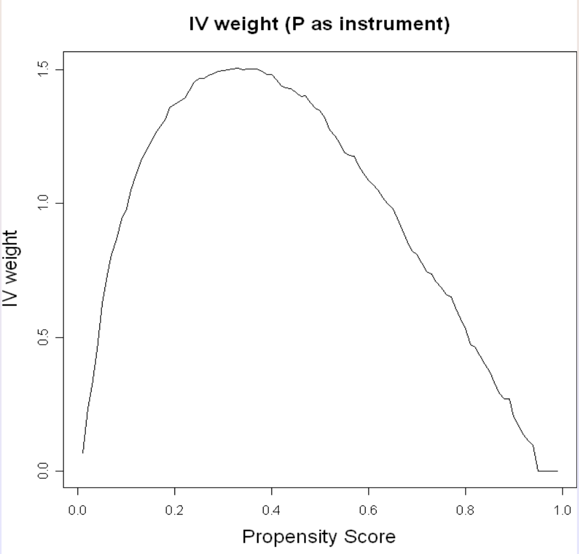
4-Year College Graduate vs. High School Wages



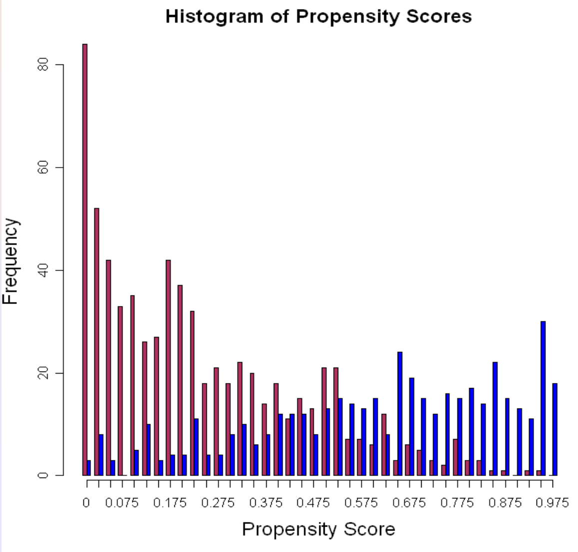
4-Year College Graduate vs. High School Wages



4-Year College Graduate vs. High School Wages



4-Year College Graduate vs. High School Wages



4-Year College Graduate vs. High School Wages

- The covariates in the outcome equations are: job tenure, job tenure squared, experience, experience squared, AFQT score, noncognitive score, marital status, indicators for black and hispanic, and year of birth indicators.

4-Year College Graduate vs. High School Wages

- The covariates in the outcome equations are: job tenure, job tenure squared, experience, experience squared, AFQT score, noncognitive score, marital status, indicators for black and hispanic, and year of birth indicators.
- The instruments are: AFQT score, noncognitive score, father's highest grade completed, mother's highest grade completed, number of siblings, family income in 1979, wages and unemployment rates of local high school graduates, wages and unemployment rates of local some college, wages and unemployment rates of local college graduates, indicators for black and hispanic, indicators for south residence and urban residence at age 14, and year of birth indicators.

4-Year College Graduate vs. High School Wages

- The dependent variable in the probit is 1 if the individual graduated from a 4-year college, and 0 if the individual's highest education is a high school diploma (GEDs are excluded).

4-Year College Graduate vs. High School Wages

- The dependent variable in the probit is 1 if the individual graduated from a 4-year college, and 0 if the individual's highest education is a high school diploma (GEDs are excluded).
- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.

4-Year College Graduate vs. High School Wages

- The dependent variable in the probit is 1 if the individual graduated from a 4-year college, and 0 if the individual's highest education is a high school diploma (GEDs are excluded).
- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.

4-Year College Graduate vs. High School Wages

- The dependent variable in the probit is 1 if the individual graduated from a 4-year college, and 0 if the individual's highest education is a high school diploma (GEDs are excluded).
- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.

4-Year College Graduate vs. High School Wages

- The dependent variable in the probit is 1 if the individual graduated from a 4-year college, and 0 if the individual's highest education is a high school diploma (GEDs are excluded).
- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.

4-Year College Graduate vs. High School Wages

- The dependent variable in the probit is 1 if the individual graduated from a 4-year college, and 0 if the individual's highest education is a high school diploma (GEDs are excluded).
- The $E(Y \mid P, X)$ curve is found by regressing log hourly wages on the X's, P, P^2, P^3, and P^4.
- The confidence intervals are found using 100 bootstraps.
- In the MTE graph, the horizontal red line indicates the IV estimate.
- In the histogram, the blue bars correspond to the $D = 1$ group and the red bars to the $D = 0$ group.
- The sample size is 1,144.

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.7101	0.7264	0.6813	0.5703

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.7101	0.7264	0.6813	0.5703
P^2	0.3324	0.7575	0.7448	0.3386

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.7101	0.7264	0.6813	0.5703
P^2	0.3324	0.7575	0.7448	0.3386
P^3		0.9467	0.8095	0.2804

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.7101	0.7264	0.6813	0.5703
P^2	0.3324	0.7575	0.7448	0.3386
P^3		0.9467	0.8095	0.2804
P^4			0.8172	0.2633

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.7101	0.7264	0.6813	0.5703
P^2	0.3324	0.7575	0.7448	0.3386
P^3		0.9467	0.8095	0.2804
P^4			0.8172	0.2633
P^5				0.2570

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

Degree of Polynomial	2	3	4	5
P	0.7101	0.7264	0.6813	0.5703
P^2	0.3324	0.7575	0.7448	0.3386
P^3		0.9467	0.8095	0.2804
P^4			0.8172	0.2633
P^5				0.2570
Joint test of nonlinear terms	0.3324	0.6047	0.7989	0.6972

4-Year College Graduate vs. High School Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.

4-Year College Graduate vs. High School Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.4275	0.4182	0.4296	0.5500	0.3369	0.3609

4-Year College Graduate vs. High School Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.4275	0.4182	0.4296	0.5500	0.3369	0.3609
TT	0.2351	0.2210	0.1940	0.3092	0.2617	0.2265

4-Year College Graduate vs. High School Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.4275	0.4182	0.4296	0.5500	0.3369	0.3609
TT	0.2351	0.2210	0.1940	0.3092	0.2617	0.2265
TUT	0.5763	0.5642	0.6028	0.8095	0.4002	0.4633

4-Year College Graduate vs. High School Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.4275	0.4182	0.4296	0.5500	0.3369	0.3609
TT	0.2351	0.2210	0.1940	0.3092	0.2617	0.2265
TUT	0.5763	0.5642	0.6028	0.8095	0.4002	0.4633
IV	0.3764	0.3764	0.3764	0.3764	0.3764	0.3764

4-Year College Graduate vs. High School Wages

Treatment Effects

Degree of Polynomial	2	3	4	5	Normal	Semipar.
ATE	0.4275	0.4182	0.4296	0.5500	0.3369	0.3609
TT	0.2351	0.2210	0.1940	0.3092	0.2617	0.2265
TUT	0.5763	0.5642	0.6028	0.8095	0.4002	0.4633
IV	0.3764	0.3764	0.3764	0.3764	0.3764	0.3764
IV (using weights)	0.3411	0.3421	0.3404	0.3468	0.3008	0.3128

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

- The *p*-values in the first panel are from *t*-tests in the case of the individual coefficients and Wald tests for the joint tests.

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

- The p-values in the first panel are from t-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

- The *p*-values in the first panel are from *t*-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.
- The treatment effects in the second panel are calculated by weighting the estimated MTE by the weights from Heckman and Vytlacil (2005).

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

- The p-values in the first panel are from t-tests in the case of the individual coefficients and Wald tests for the joint tests.
- The standard errors are calculated using 50 bootstrap samples.
- The treatment effects in the second panel are calculated by weighting the estimated MTE by the weights from Heckman and Vytlacil (2005).
- Therefore, they vary depending on the degree of the polynomial used to approximate $E(Y \mid P)$ (and hence the polynomial used to approximate the MTE).

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.
- The estimates differ not only because the estimate of the MTE is inexact, but also because the weights are estimated.

4-Year College Graduate vs. High School Wages

P-values from sequentially adding polynomial terms

- The IV estimate is using $P(Z)$, the propensity score, as the instrument.
- The IV estimate is calculated both using the weights from Heckman and Vytlacil (2005) and using the traditional ratio of covariances.
- The estimates differ not only because the estimate of the MTE is inexact, but also because the weights are estimated.
- In both panels the degree of the polynomial refers to the degree used to approximate $E(Y \mid P)$ (the degree of the approximation to the MTE is one less).

Summary and Conclusion

- This paper seeks to determine whether such concerns are important in practice.

- This paper seeks to determine whether such concerns are important in practice.
- Using data from four prototypical choice settings in labor economics, we have shown reasonable evidence that such heterogeneity is indeed present.

- This paper seeks to determine whether such concerns are important in practice.
- Using data from four prototypical choice settings in labor economics, we have shown reasonable evidence that such heterogeneity is indeed present.
- Our strongest results come from the data on union membership and college graduation, while the data on high school graduation and GED certification are less conclusive.