Testing for the Presence of Measurement Error

Daniel Wilhelm
University College London & CeMMAP, IFS, CReAM

Two Types of Hypothesis Testing Problems

Null hypotheses of interest:

1. **No measurement error** (ME) in X:

$$H_0 : \quad X^* = X \quad a.s.$$

2. ME in X **does not distort** relationship to outcome Y, e.g.

$$H_0^{func} : \quad E[Y|X^*] = E[Y|X] \quad a.s.$$

where

- X possibly mismeasured observation of "truth" X^*
- Y an outcome that depends on X^*

How do we test such hypotheses based on observables?

Example: Technology of Skill Production

Production of **adulthood output** Y from **skill inputs** X^* (Cunha, Heckman, and Schennach (2010)):

$$Y = g(X^*, \varepsilon)$$

where:

- X is a **noisy measure** of skill inputs X^* (e.g. test scores)
- Z is **another measure** related to X^* (e.g. a noisy measure of parental skills)

Main idea:

- Do not attempt to identify and estimate g or distribution of X^*
- Instead perform inference by exploiting the restriction

$$Y \text{ independent of } Z \text{ given } X^*$$

Other Examples

1. **reliability of admin data**
   - e.g. Kapteyn and Ypma (2007), Groen (2011), Abowd and Stinson (2007). Fitzenberger, Osikominu, and Völter (2006)
   - test for ME in admin data or corrected variables

2. **optimization frictions**
   - e.g. Chetty (2012)
   - test for absence of frictions (e.g. adjustment cost, inattention, status quo biases)

3. **investment and Q**
   - Erickson and Whited (2000, 2012)
   - test for economic assumptions (perfect competition, . . .)

4. **consumption and income inequality**
   - Aguiar and Bils (2015)
   - test for ME in total expenditure

5. **effect of language proficiency on immigrants' wages**
   - e.g. Dustmann and van Soest (2001)
   - test for ME in self-reported language proficiency

6. **unemployment dynamics**
   - e.g. Feng and Hu (2013)
   - testing for misclassification in unemployment

Literature

Tests for Presence of ME:

- binary misclassification: Mahajan (2006)
- parametric model with "small" ME: Chesher (1990), Chesher, Dumangane, and Smith (2002)
- Hausman tests: e.g. Hausman (1978), Hahn and Hausman (2002), Hu (2008)

Related Testing Problems:

- testing for classical ME: Wilhelm (2017)
- powerful t-tests: Kim and Wilhelm (2017)
- omitted variable tests: Gozalo (1993), Fan and Li (1996), Lavergne and Vuong (1996), Delgado and Gonzalez Manteiga (2001), . . .

Overview

1. The Null Hypothesis of No ME
2. The Null of Equal Functionals
3. The Test
4. Is the Skill Production Function Distorted?
5. Is There ME in Admin Data?

The Null Hypothesis of No ME

The Null Hypothesis of Interest

We want to test:

$$H_0 \colon P(X^* = X) = 1 \qquad \text{vs.} \qquad H_1 \colon P(X^* = X) < 1$$

- Y outcome (observed)
- X^* "true" explanatory variable (unobserved?)
- X measurement of X^* (observed)
- Z second indicator of X^* (observed)

H_0 depends on distribution of unobservables!

How do we test H_0 based on observables (Y, X, Z)?

Observable Implications of H_0

Assumption 1 (exclusion)

$Y \perp (X, Z) \mid X^*$

Theorem 1

For any P satisfying Assumption 1, H_0 implies

$$Y \perp Z \mid X \tag{1}$$

- (1) depends only on observables
- a valid test of (1) is a valid test of H_0!
- alternatively allow for nondifferential ME in X: $Y \perp Z \mid (X^*, X)$

Equivalence

Under what conditions is the reverse also true?

Equivalence

What type of ME models can be recovered?

Equivalence

Assumption 2 (monotonicity)

There is a function μ s.t. $x^* \mapsto E_P[\mu(Y)|X^* = x^*]$ is monotone and not constant.

Assumption 3 (relevance – FOSD)

If $P(X^* = X) < 1$, then there exist x, z_1, z_2 such that the distribution of $X^*|X = x, Z = z_2$ under P first-order stochastically dominates that of $X^*|X = x, Z = z_1$ under P, and they are not equal.

Theorem 2

If P satisfies Assumptions 1–3, then H_0 holds if, and only if,

$$Y \perp Z \mid X.$$

▸ intuition

When Does the Distribution of $X^*|X, Z$ Satisfy the FOSD Condition?

Consider **outcome equation**

$$Y = g(X^*, \varepsilon)$$

with different types of measurement systems:

1. **binary misclassification:** X^*, X, Z binary
   - directly verify the FOSD condition ▸ details
2. Z **repeated measurement** of continuous X^*:
   - single-crossing condition on density of $Z|X^*$ for two values of Z
3. Z **instrumental variable** for continuous X^*:
   - single-crossing condition on density of $X^*|Z$ for two values of Z ▸ details

Discussion

Identification/estimation of the latent structure:

- requires strong assumptions
- implicitly defines "true" latent variable through statistical assumptions
- claims to recover the latent "truth"

This paper:

- gives up on identification of the latent structure
- exploits one weak restriction that can often be motivated through knowledge of the economic setting
- avoids recovering the latent truth

The Null of Equal Functionals

The Null of Equal Functionals

A similar equivalence result can be shown for the null of equal functions:

We want to test:

$$H_0^{func}: P\Big(E_P[\rho(Y)|X] = E_P[\rho(Y)|X^*]\Big) = 1 \qquad \text{vs.} \qquad H_1^{func}: \text{not } H_0^{func}$$

Theorem 3

If P satisfies assumptions similar to those of Theorem 2, then H_0^{func} holds if, and only if,

$$E_P[\rho(Y) \mid X, Z] = E_P[\rho(Y) \mid X] \quad a.s.$$

The Test

Test for the Presence of ME

Want to test:

$$P(Y \leq y \mid X, Z) = P(Y \leq y \mid X)$$

Can use variety of existing tests:

- Gozalo (1993), Fan and Li (1996), Delgado and Gonzalez Manteiga (2001), Mahajan (2006), and Huang, Sun, and White (2016), ...
- Delgado and Gonzalez Manteiga (2001) implemented in R and STATA:
  `http://github.com/danielwilhelm/R-ME-test`
  `http://github.com/danielwilhelm/STATA-ME-test`

▸ Delgado and Gonzalez Manteiga (2001) details ▸ simulations

Is the Skill Production Function Distorted?

Production function (Cunha, Heckman, and Schennach (2010)):

$$Y = g(X^*) + \varepsilon, \qquad E[\varepsilon | X^*] = 0$$

CNLSY/79 data on

- Y: various adulthood outcomes
- X: contains a measure of cognitive and of noncognitive skills
- Z: measure of parental cognitive and noncognitive skills

Test for presence of **ME distortion in production function**!

Test Results

Y	Z	p-value	bandwidths		sample size
logsalary	both	1.00	2.69	0.44	254
	se1	1.00	2.69	0.44	254
	asvab2	1.00	2.69	0.44	254
education	both	1.00	0.28	0.70	630
	se1	0.21	0.28	0.70	630
	asvab2	1.00	0.28	0.70	630
convict	both	0.04	1.31	0.60	854
	se1	0.02	1.31	0.60	854
	asvab2	0.03	1.31	0.60	854

Table 1: testing the null of no ME distortions in production function; cross-validated bandwidths; Cramér-von Mises version of Delgado-Manteiga test; Mammen bootstrap multiplier

Is There ME in Admin Data?

Data

Matched CPS and employer-reported social security earnings:

- **survey data**: CPS 1978
- **admin data**: social security earnings

▸ sample selection

Model:

- Y: survey earnings in 1977
- X: admin earnings in 1977
- Z: admin earnings in 1976

▸ assumptions

Test for presence of **ME in admin earnings**!

Distribution of Survey and Admin Earnings

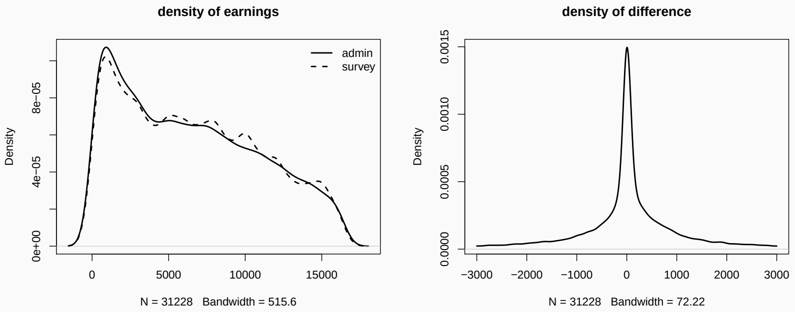


Figure 1: Nonparametric density estimates, using cross-validated bandwidth

Nonparametric Estimate of the Conditional Mean Function

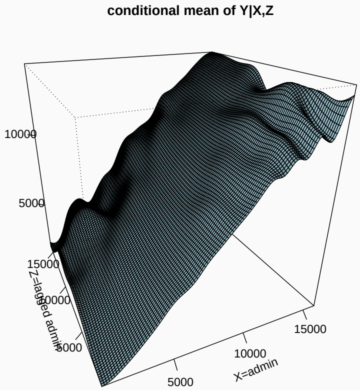


Figure 2: Nonparametric estimate of $E[Y|X, Z]$, where Y is survey earnings in 1977, X is admin earnings in 1977, and Z is admin earnings in 1976; bandwidths chosen by cross-validation

Test Results

	p-value	bandwidth	sample size
full sample	0.000	287.6	31,228
earnings in IQR	0.000	251.8	15,614
white males	0.000	127.1	12,591
+ singles	0.000	340.5	5,043
+ age in [25,65]	0.012	870.0	1,669
+ full-time, full-year (*)	0.017	914.2	972
+ at least highschool	0.030	851.3	867
+ earnings in IQR	0.012	787.4	342

Table 2: testing the null of no ME in admin earnings 1977; cross-validated bandwidths; Cramér-von Mises version of Delgado-Manteiga test; Mammen bootstrap multiplier

Conclusions

Testing for presence of ME and ME distortions:

- under weak assumptions
- using simple techniques
- consistent against large class of nonclassical ME models

More in the paper:

- more detailed discussion of sufficient conditions for FOSD
- including sufficient conditions in well-studied ME models
- simulations

Software:

- R code at `http://github.com/danielwilhelm/R-ME-test`
- STATA code at `http://github.com/danielwilhelm/STATA-ME-test`

Thank you!

Under What Conditions Does $Y \perp Z \mid X$ Imply H_0?

$Y \perp Z \mid X$ implies that, for any z_1, z_2,

$$P_{Y|X,Z=z_1} = P_{Y|X,Z=z_2}$$

back

Under What Conditions Does $Y \perp Z \mid X$ Imply H_0?

$Y \perp Z \mid X$ implies that, for any z_1, z_2,

$$P_{Y|X,Z=z_1} = P_{Y|X,Z=z_2}$$

and thus, by Assumption 1, for any function μ,

$$E_P[\mu(Y)|X, Z = z_1] = E_P[\mu(Y)|X, Z = z_2] \tag{2}$$

back

Under What Conditions Does $Y \perp Z \mid X$ Imply H_0?

$Y \perp Z \mid X$ implies that, for any z_1, z_2,

$$P_{Y|X,Z=z_1} = P_{Y|X,Z=z_2}$$

and thus, by Assumption 1, for any function μ,

$$E_P[\mu(Y)|X, Z = z_1] = E_P[\mu(Y)|X, Z = z_2]$$

$$\Rightarrow \quad \int E_P[\mu(Y)|X^*, X, Z]\, dP_{X^*|X,Z=z_1} = \int E_P[\mu(Y)|X^*, X, Z]\, dP_{X^*|X,Z=z_2} \quad (2)$$

back

Under What Conditions Does $Y \perp Z \mid X$ Imply H_0?

$Y \perp Z \mid X$ implies that, for any z_1, z_2,

$$P_{Y|X,Z=z_1} = P_{Y|X,Z=z_2}$$

and thus, by Assumption 1, for any function μ,

$$\begin{aligned} & E_P[\mu(Y)|X, Z = z_1] = E_P[\mu(Y)|X, Z = z_2] \\ \Rightarrow \quad & \int E_P[\mu(Y)|X^*] \, dP_{X^*|X,Z=z_1} = \int E_P[\mu(Y)|X^*] \, dP_{X^*|X,Z=z_2} \end{aligned} \tag{2}$$

◂ back

Under What Conditions Does $Y \perp Z \mid X$ Imply H_0?

$Y \perp Z \mid X$ implies that, for any z_1, z_2,

$$P_{Y|X,Z=z_1} = P_{Y|X,Z=z_2}$$

and thus, by Assumption 1, for any function μ,

$$E_P[\mu(Y)|X, Z = z_1] = E_P[\mu(Y)|X, Z = z_2]$$

$$\Rightarrow \quad \int E_P[\mu(Y)|X^*]\, dP_{X^*|X,Z=z_1} = \int E_P[\mu(Y)|X^*]\, dP_{X^*|X,Z=z_2} \tag{2}$$

Suppose, under the alternative H_1,

1. there exists μ s.t. $x^* \mapsto E_P[\mu(Y)|X^* = x^*]$ **monotone**
2. $P_{X^*|X,Z=z_2}$ **first-order stochastically dominates** $P_{X^*|X,Z=z_1}$

then (2) cannot hold under H_1, but only under H_0.

back

Test by Delgado and Gonzalez Manteiga (2001)

For a random sample $\{(Y_i, X_i, Z_i)\}_{i=1}^n$:

1. nonparametrically estimate $E[Y|X = x]$
2. form **Test statistic**:

$$K_n := \sup_{x,z} |\sqrt{n} T_n(x, z)|$$

   where T_n is sample analogue of

$$T(x, z) := E\left[f_X(X)\left\{Y - E(Y|X)\right\} \mathbb{1}\{X \leq x\}\mathbb{1}\{Z \leq z\}\right]$$

3. bootstrap critical value

▸ simulations

Motivation for Test by Delgado and Gonzalez Manteiga (2001)

Null hypothesis:

$$H_0' : \quad E[Y|X,Z] = E[Y|X] \qquad a.s.$$

Notice that H_0' is equivalent to

$$f_X(X)E\left[Y - E(Y|X) \mid X, Z\right] = 0$$

or

$$T(X,Z) = 0$$

with

$$T(x,z) := E\left[f_X(X)\left\{Y - E(Y|X)\right\}\mathbb{1}\{X \leq x\}\mathbb{1}\{Z \leq z\}\right]$$

back

Simulations

Simulation Setup I

Outcome equation:

$$Y = X^{*2} + \frac{1}{2}X^* + N(0, \sigma_\varepsilon^2)$$

Measurement system:

1. **Kotlarski**:

$$X = X^* + D \cdot N(0, \sigma_{ME}^2)$$
$$Z = X^* + N(0, 0.5^2)$$

   where $D \sim Bernoulli(1 - \lambda)$, $X^* \sim U[0, 1]$, and $\sigma_\varepsilon = 0.5$.

2. **Quadratic**: Kotlarski with

$$Z = -(X^* - 1)^2 + N(0, 0.2^2)$$

   and $\sigma_\varepsilon = 0.2$.

Simulation Setup II

1. **Nonclassical ME I**: Kotlarski with heteroskedastic ME:

$$X = X^* + D \cdot N(0, \sigma_{ME}^2) e^{-|X^* - 0.5|}$$
$$Z = X^* + N(0, 0.5^2) e^{-|X^* - 0.5|}$$

2. **Nonclassical ME II**: Nonclassical ME I with correlated ME:

$$X = X^* + D \cdot \left[\frac{1}{2}\xi + N(0, \sigma_{ME}^2) \frac{e^{-|X^* - 0.5|}}{2} \right]$$
$$Z = X^* + \left[\frac{1}{2}\xi + N(0, 0.5^2) \frac{e^{-|X^* - 0.5|}}{2} \right]$$

with $\xi \sim N(0, 0.5^2)$.

Simulation Setup III

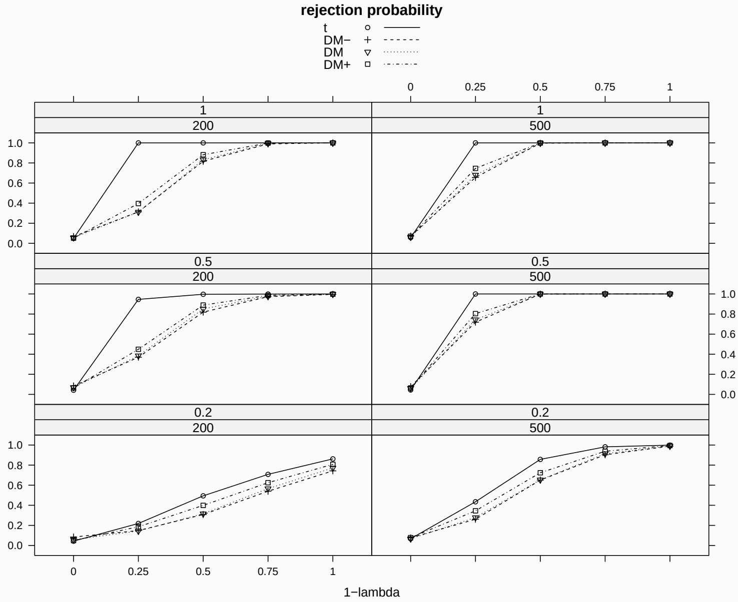
Tests:

1. t-test: test $\gamma = 0$ in
$$Y = \alpha + \beta X + \gamma Z + \varepsilon$$
2. Delgado and Manteiga's test:
    - "DM": DM test using rule-of-thumb bandwidth $h = 0.2n^{-1/3}$
    - "DM-": DM test using smaller bandwidth $h = 0.1n^{-1/3}$
    - "DM+": DM test using larger bandwidth $h = 0.5n^{-1/3}$

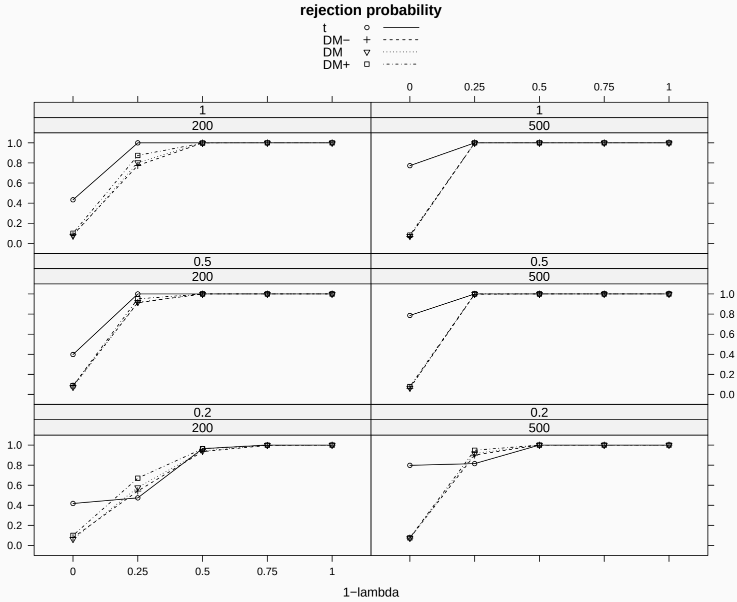
Simulation parameters:

- vary probability of ME $\lambda = P(X^* = X)$
    - no ME (i.e. null hypothesis): $1 - \lambda = 0$
    - ME everywhere: $1 - \lambda = 1$
- vary ME variance $\sigma^2_{ME} \in \{0.2^2, 0.5^2, 1\}$
- vary sample size $n \in \{200, 500\}$
- 1,000 MC samples with 100 bootstrap replications for DM test

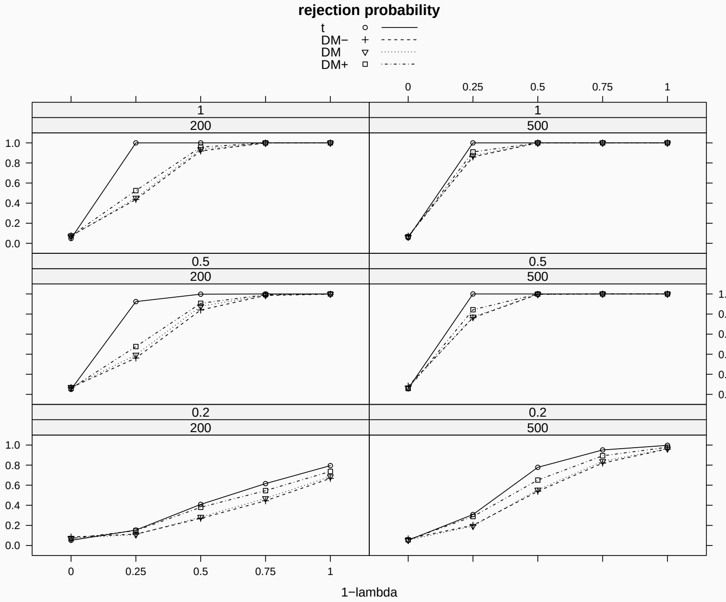
Power – Kotlarski



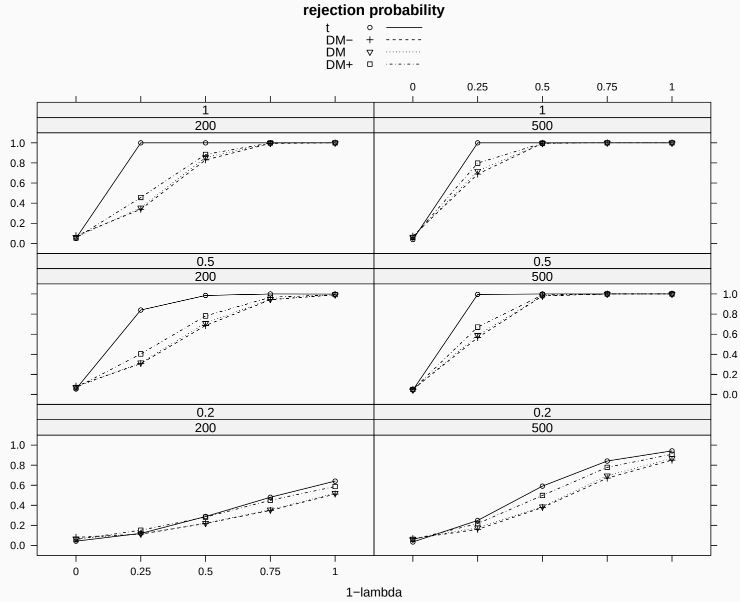
Power – Quadratic



Power – Nonclassical ME I



Power – Nonclassical ME II



Null Rejection Probabilities

n	σ_{ME}	test	Kotlarski	Nonclassical I	Nonclassical II	Quadratic
200	0.2	t	0.043	0.051	0.043	0.418
		DM-	0.084	0.084	0.081	0.078
		DM	0.058	0.073	0.068	0.062
		DM+	0.049	0.069	0.060	0.100
	0.5	t	0.042	0.050	0.054	0.397
		DM-	0.084	0.068	0.083	0.084
		DM	0.069	0.072	0.072	0.072
		DM+	0.064	0.060	0.067	0.086
	1	t	0.049	0.048	0.050	0.433
		DM-	0.067	0.075	0.075	0.084
		DM	0.057	0.069	0.060	0.075
		DM+	0.052	0.075	0.053	0.100
500	0.2	t	0.067	0.052	0.035	0.798
		DM-	0.076	0.066	0.069	0.074
		DM	0.067	0.053	0.064	0.073
		DM+	0.078	0.056	0.060	0.076
	0.5	t	0.044	0.058	0.043	0.786
		DM-	0.077	0.083	0.050	0.062
		DM	0.060	0.067	0.046	0.064
		DM+	0.064	0.058	0.049	0.077
	1	t	0.057	0.056	0.037	0.773
		DM-	0.072	0.072	0.072	0.068
		DM	0.062	0.064	0.064	0.071
		DM+	0.070	0.066	0.059	0.081

Table 3: Null rejection probabilities

Simulations

back

Sample Selection

Remove individuals

- with earnings from sources other than wages and salaries
- with zero admin earnings in 1977
- neither working full-time nor part-time
- whose earnings were top-coded (USD 16,500)
- that could not be matched to admin data

Summary Statistics

	before		after	
	mean	stand. dev.	mean	stand. dev.
survey earnings 1977	4,195.6	7,143.8	6,567.7	4,508.4
– fraction topcoded	0.1	0.4	0	0
administrative earnings 1977	4,703.6	5,867.8	6,422.6	4,539.6
– fraction topcoded	0.6	0.5	0	0
administrative earnings 1976	4,308.5	5,443.7	5,595.2	4,486.0
white	0.8	0.4	0.9	0.3
age	32.5	21.9	35.4	14.8
education	9.7	6.0	13.1	2.7
married	0.4	0.5	0.6	0.5
sample size	168,904		31,378	

Table 4: Summary statistics before and after sample selection.

back

Econometric Specification

- **true earnings** E_t^* over two time periods $t = 1, 2$:

$$E_2^* = h(E_1^*, U)$$

- **admin measurement** of earnings:

$$A_t = m_{At}(E_t^*, \underbrace{\eta_{At}}_{ME}), \qquad t = 1, 2$$

- **survey measurement** of earnings:

$$S_t = m_{St}(E_t^*, \underbrace{\eta_{St}}_{ME}), \qquad t = 1, 2$$

Econometric Specification

Rewriting slightly:

$$
\begin{aligned}
S_2 &= m_{S2}(E_2^*, \eta_{S2}) && \text{(“outcome } Y\text{”)} \\
A_2 &= m_{A2}(E_2^*, \eta_{A2}) && \text{(“1st measurement } X\text{”)} \\
A_1 &= m_{A1}(h^{-1}(E_2^*, U), \eta_{A1}) && \text{(“2nd measurement } Z\text{”)}
\end{aligned}
$$

Exclusion Restriction

Model:

$$A_1 = m_{A1}(h^{-1}(E_2^*, U), \eta_{A1}) \qquad (\text{“outcome } Y\text{”})$$

$$A_2 = m_{A2}(E_2^*, \eta_{A2}) \qquad (\text{“1st measurement } X\text{”})$$

$$S_2 = m_{S2}(E_2^*, \eta_{S2}) \qquad (\text{“2nd measurement } Z\text{”})$$

Sufficient condition for exclusion:

$$\eta_{S2} \quad \text{independent of} \quad (\eta_{A1}, U) \quad \text{given} \quad E_2^*$$

- survey ME independent of admin ME given true earnings
- plausible as admin and survey data collection independent
- nonclassical ME (allowed to depend on true earnings level)

Empirical Evidence for Relevance Condition

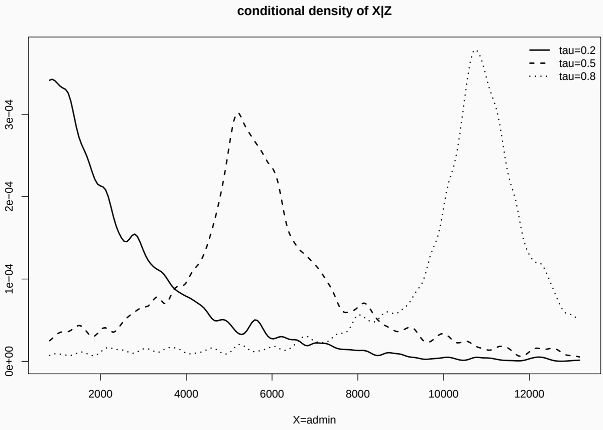


Figure 3: Conditional density of $A_2|A_1 = q$ for q being τ-quantile of A_1; here, A_2 and A_1 are admin earnings in 1977 and 1976; nonparametric estimates with bandwidths chosen by cross-validation

back