7 – ASSESSING REGRESSION-BASED STUDIES

University of Massachusetts Amherst BE REVOLUTIONARY™

SECTION 7 – ASSESSING REGRESSION-BASED STUDIES
THE PLAN

1. Internal & External Validity.

2. Threats to Validity in Causal Inference Analysis.

3. Threats to Validity in Prediction.

4. Application to Test Scores & Class Size.

OVERVIEW

- What makes an econometric analysis solid/convincing?
- Need to identify & evaluate assumptions!

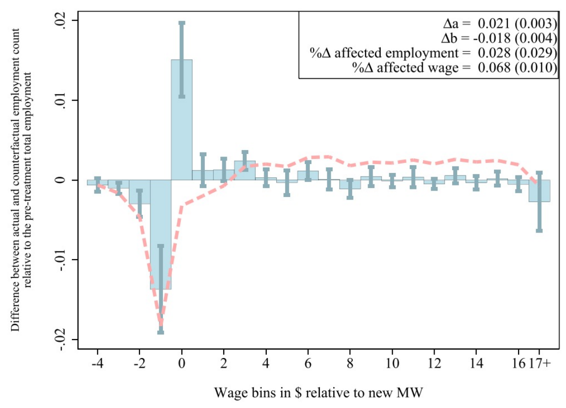

EFFECT OF MINIMUM WAGES ON LOW-WAGE JOBS 1423

FIGURE II

Impact of Minimum Wages on the Wage Distribution

The figure shows the main results from our event study analysis (see equation (1)) exploiting 138 state-level minimum wage changes between 1979 and 2016. The blue bars show for each dollar bin (relative to the minimum wage) the estimated average employment changes in that bin during the five-year post-treatment relative to the total employment in the state one year before the treatment. The error bars show the 95% confidence interval using standard errors that are clustered at the state level shown using the error bar. The dashed red line (color version available online) shows the running sum of employment changes up to the wage bin it corresponds to.

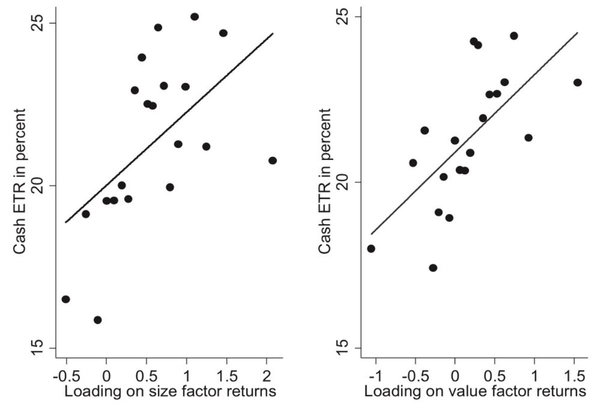

8 A.F. Wagner et al. / Journal of Financial Economics 000 (2018) 1–24

Fig. 2. Binned scatter plots of loading on size (left panel) and value (right panel) factor returns against cash ETR. The factor loadings are computed by regressing daily firm excess returns on daily market excess return, size, and value factor returns (obtained from Ken French's website) from October 1, 2015 to September 30, 2016. The sample includes Russell 3000 firms.

7.1 INTERNAL & EXTERNAL VALIDITY

INTERNAL & EXTERNAL VALIDITY

Two main criteria:

1. Internal Validity
2. External Validity.

THREATS TO EXTERNAL VALIDITY

1. Differences in populations.
   - Population sampled might be fundamentally different from population of interest.

2. Differences in settings.
   - Populations may be very similar but operate under fundamentally different (legal, environmental, social…) settings.

7.2 THREATS TO INTERNAL VALIDITY IN CAUSAL INFERENCE ANALYSIS

THREATS TO INTERNAL VALIDITY IN CAUSAL INFERENCE ANALYSIS

1. Omitted Variables Bias.
2. Wrong functional form.
3. Errors-in-variables (or *measurement error*) bias.
4. Sample selection bias.
5. Simultaneous (or *reverse*) causality bias.

…we already know about 1,2 & 5, so let's discuss 3 & 4 now…

→ *They all result in correlation between regressor X and error term u.*

ERRORS-IN-VARIABLES BIAS

- *Measurement error* in the independent variable
  - we observe $\tilde{X}$ rather than the true X.
  - $\tilde{X} - X$ = measurement error.
- Some algebra to understand what we are estimating:
  - $Y_i = \beta_0 + \beta_1 X_i + u_i$
  - $Y_i = \beta_0 + \beta_1 \tilde{X}_i + \beta_1 (X_i - \tilde{X}_i) + u_i$
  - $Y_i = \beta_0 + \beta_1 \tilde{X}_i + v_i$ with $v_i = \beta_1 (X_i - \tilde{X}_i) + u_i$

ERRORS-IN-VARIABLES BIAS

- What we estimate with measurement error in X is:

$Y_i = \beta_0 + \beta_1 \tilde{X}_i + v_i$ with $v_i = \beta_1 \left(X_i - \tilde{X}_i \right) + u_i$

- → $\hat{\beta}_1$ biased if $corr(\tilde{X}_i, X_i - \tilde{X}_i) \neq 0$ → $corr\left(\tilde{X}_i, v_i\right) \neq 0$

Case (1): Classical measurement error

$$\tilde{X}_i = X_i + w_i;$$

w_i is random: $corr(X_i, w_i) = corr(\tilde{X}_i, w_i) = corr(w_i, u_i) = 0$

→ Then $\hat{\beta}_1$ is biased towards zero (*attenuation bias*).

ERRORS-IN-VARIABLES BIAS

- What we estimate with measurement error in X is:

$$Y_i = \beta_0 + \beta_1 \tilde{X}_i + v_i \qquad \text{with } v_i = \beta_1 (X_i - \tilde{X}_i) + u_i$$

- → $\hat{\beta}_1$ biased if $corr(\tilde{X}_i\,, X_i - \tilde{X}_i) \neq 0$ → $corr(\tilde{X}_i\,, v_i) \neq 0$

Case (2): "Best guess" measurement error

$$\tilde{X}_i = E(X_i | info_i);$$

→ $corr(\tilde{X}_i\,, X_i - \tilde{X}_i) = 0$ → $\hat{\beta}_1$ is not biased

ERRORS-IN-VARIABLES BIAS

What if *the dependent variable Y* is measured with error?

- $\hat{\beta}_1$ remains unbiased both with classical & "best guess" measurement error.
- Random measurement error in Y = just another component of error term u_i.

EXAMPLE OF MEASUREMENT ERROR

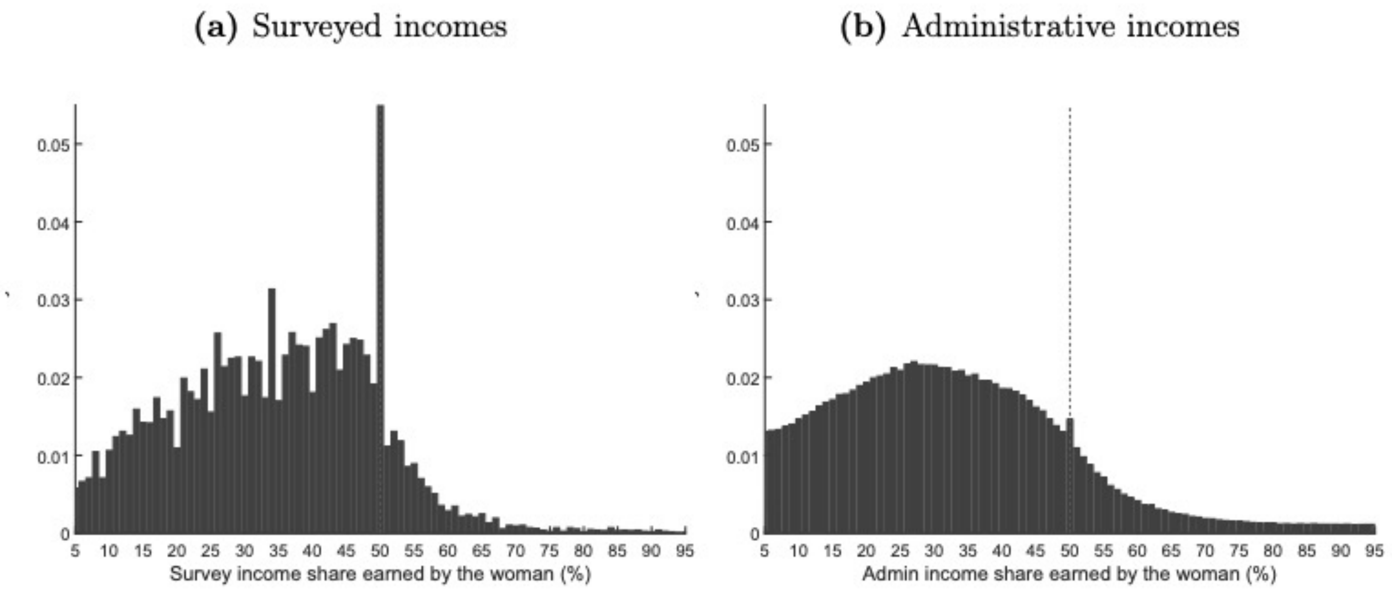


Figure 1: Overall distribution of female income shares in the couple. The shaded area represents the histogram of the underlying data in 1 percent bins. The figure on the left visualizes the distribution observed in survey data (based on SAKE survey years 2002, 2005, 2008, 2012, and 2015). The figure on the right shows the same distribution based on administrative income data for married couples (this data are described in detail in Section A.2 in the appendix). The corresponding density discontinuity estimates can be found in row (1) and row (4) of Table A.1 in Appendix A.4.

"This misreporting is best explained by the role of gender norms in individuals' self-portrayals and self perception"

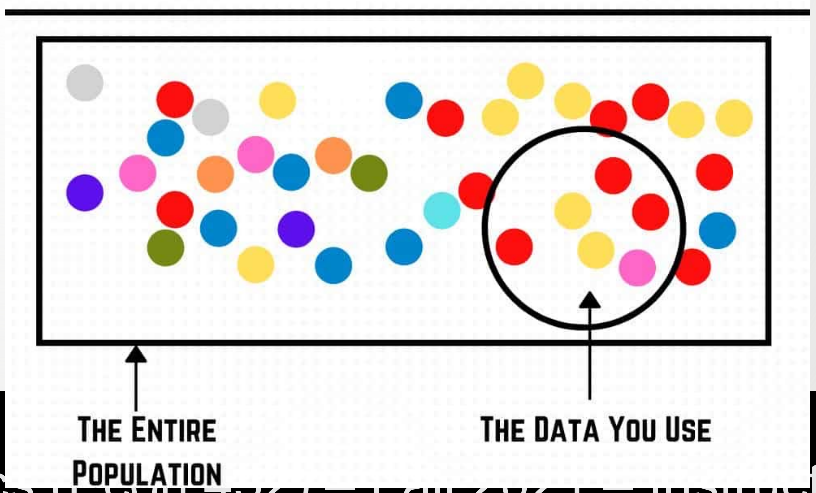
[Source: Anja Roth and Michaela Slotwinski "Gender norms and income misreporting within households"]

MISSING DATA & SAMPLE SELECTION BIAS

3 cases:

1. Data missing at random
2. Data missing based on value of X
3. Data missing based on the value of Y → sample selection bias

Just a smaller sample but no bias.



MISSING DATA & SAMPLE SELECTION BIAS

Sample selection bias - some examples:

1. Estimate effect of a policy on firm performance, but firms that go bankrupt are unobserved.

2. Electoral polls, when nonresponse rate varies across parties.

3. Estimate the effect of height on basketball performance using a sample of NBA players.

7.3 VALIDITY IN PREDICTION/FORECASTING

University of Massachusetts Amherst

VALIDITY IN PREDICTION/FORECASTING

- What if we are interested in prediction/forecasting rather than causal inference?
- What matters in forecasting is that X is a reliable predictor of Y.
  - Omitted Variables Bias not a problem.
  - We don't need $corr(Xi, u_i) = 0$.
  - (Adjusted) R2 matters a lot.
- External validity is key: the model must hold *out of sample*.

7.4 APPLICATION TO TEST SCORES & CLASS SIZE

University of Massachusetts Amherst

APPLICATION: TEST SCORES & CLASS SIZE

External Validity: do results generalize beyond CA schools in 1999?

- Analyze a dataset of MA school districts in 1998 and compare results.

TABLE 9.1 Summary Statistics for California and Massachusetts Test Score Data Sets

	California		Massachusetts	
	Average	Standard Deviation	Average	Standard Deviation
Test scores	654.1	19.1	709.8	15.1
Student–teacher ratio	19.6	1.9	17.3	2.3
% English learners	15.8%	18.3%	1.1%	2.9%
% receiving subsidized lunch	44.7%	27.1%	15.3%	15.1%
Average district income ($)	$15,317	$7226	$18,747	$5808
Number of observations	420		220	
Year	1999		1998	

ESTIMATED CLASS SIZE EFFECTS: CA VS MA

TABLE 9.3 Student-Teacher Ratios and Test Scores: Comparing the Estimates from California and Massachusetts

	OLS Estimate $\hat{\beta}_{STR}$	Standard Deviation of Test Scores Across Districts	Estimated Effect of Two Fewer Students per Teacher, in Units of: Points on the Test	Standard Deviations
California				
Linear: Table 8.3(2)	−0.73 (0.26)	19.1	1.46 (0.52) [0.46, 2.48]	0.076 (0.027) [0.024, 0.130]
Massachusetts				
Linear: Table 9.2(3)	−0.64 (0.27)	15.1	1.28 (0.54) [0.22, 2.34]	0.085 (0.036) [0.015, 0.154]

Standard errors are given in parentheses. 95% confidence intervals for the effect of a two-student reduction are given in brackets.

Complication: one test point does not mean the same in MA and CA.

→ Standardization: express results in terms of standard deviations.

1. Similar effects ✅
2. Statistically significant in both cases ✅
3. Effect of STR doesn't depend on PctEL (no interaction) ✅
4. Nonlinearities ❌

CLASS SIZE & TEST SCORES: INTERNAL VALIDITY

1. **Omitted variables bias?**
   - We addressed it, but might always remain
   - Schools/parents' attention to teaching quality?
2. **Functional form?**
   - Doesn't seem to matter.
3. **Measurement error?**
   - Possible. Might bias results towards zero
4. **Sample selection?**
   - No, we have data on all CA school districts
5. **Simultaneous causality?**
   - No reason to think so.