

# Mostly Harmless Econometrics

Alex Young

January 17, 2012

## Contents

|          |   |          |
|----------|---|----------|
| <b>1</b> | <b>The Experimental Ideal</b>                       | <b>2</b> |
| 1.1      | Selection Bias . . . . .                            | 2        |
| <b>2</b> | <b>Making Regression Make Sense</b>                 | <b>3</b> |
| 2.1      | Regression versus Matching . . . . .                | 3        |
| <b>3</b> | <b>Instrumental Variables in Action</b>             | <b>4</b> |
| 3.1      | Example: Parker et al. [2011] . . . . .             | 5        |
| <b>4</b> | <b>Parallel Worlds</b>                              | <b>7</b> |
| 4.1      | Differences-in-Differences . . . . .                | 7        |
| 4.1.1    | Example: Bertrand and Mullainathan [2003] . . . . . | 8        |

# 1 The Experimental Ideal

## 1.1 Selection Bias

Think about treatment as described by a binary random variable:  $D_i = \{0, 1\}$ . The outcome of interest is denoted by  $Y_i$ . The question is whether  $Y_i$  is *affected* by the treatment. To address this question, assume we can imagine what might have happened to someone who was treated if that person had not been (treated), and vice versa.

Hence, for any individual, there are two potential outcomes:  $1 \cdot Y_{1i}[D_i = 1]$ . We would like to know the difference between  $Y_{1i}$  and  $Y_{0i}$ , which can be said to be the causal effect of the treatment for individual  $i$ .

The observed outcome,  $Y_i$ , can be written in terms of potential outcomes as

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) \cdot D_i$$

A naïve comparison of averages by treatment tells something about potential outcomes, though not necessarily what we want to know, which is the causal effect (of treatment). The observed difference in average outcome is formally linked to the average causal effect by the equation

$$\begin{aligned} E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0) &= \{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)\} \\ &\quad + \{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)\} \end{aligned}$$

The first term in braces is the average treatment effect on the treated. The observed difference in treatment, however, adds to this causal effect *selection bias*, which is the difference

in average  $Y_{0i}$  between those who were and were not treated.

## 2 Making Regression Make Sense

### 2.1 Regression versus Matching

Consider the following equation:

$$Y_i = \sum_x d_{ix} \alpha_x + \delta_R D_i + e_i$$

where  $d_{ix} = 1[X_i = x]$  is a dummy variable that indicates  $X_i = x$ ,  $\alpha_x$  is a regression effect for  $X_i = x$  and  $\delta_R$  is the regression estimand.

To see the difference between regression and matching, consider the following:

$$\begin{aligned} \delta_R &= \frac{Cov(Y_i, \tilde{D}_i)}{V(\tilde{D}_i)} \\ &= \frac{E[Y_i \tilde{D}_i] - E[Y_i]E[\tilde{D}_i]}{E[(D_i - E[D_i|X_i])^2]} \end{aligned}$$

After that the equalities should be clear. What about “similarly, the second term simplifies to”?

$$\begin{aligned} E\{(D_i - E[D_i|X_i])D_i\delta_X\} &= E\{(D_i - E(D_i|X_i))D_i\delta_X - (D_i - E(D_i|X_i))E(D_i|X_i)\delta_X\} \\ &= E\{(D_i - E(D_i|X_i))^2\delta_X\} \end{aligned}$$

The key “trick,” which is kind of cheap, is that you’re subtracting something that upon expectation is zero.

### 3 Instrumental Variables in Action

Suppose that potential outcomes can be written as follows:

$$\begin{aligned} Y_{si} &\equiv f_i(s) \\ &= \alpha + \rho s_i + \eta_i \\ &= \alpha + \rho s_i + (A_i' \gamma + \nu_i) \end{aligned}$$

where  $\gamma$  is again a vector of population regression coefficients, so that  $\nu_i$  and  $A_i$  are uncorrelated by construction.  $A_i$  (i.e. “ability”) is assumed to be the only reason why  $\eta_i$  and  $s_i$  are correlated so that  $\mathbb{E}[s_i \cdot \nu_i] = 0$ . But we cannot observe, or at least it is very difficult to estimate, “ability.” Hence, we cannot estimate  $\rho$  from the following equation due to omitted variables bias:

$$Y_{si} = \alpha + \rho s_i + A_i' \gamma + \nu_i$$

But suppose that there exists a variable  $z_i$ , an instrument, that is correlated with  $s_i$  but not correlated with either  $A_i$  or  $\nu_i$ . Following Hayashi [2000], we can estimate  $\rho$  with this variable:

$$\begin{aligned} Cov(y_i, z_i) &= Cov(\alpha + \rho s_i + A_i' \gamma + \nu_i, z_i) \\ &= \rho Cov(s_i, z_i) \\ \Rightarrow \rho &= \frac{Cov(y_i, z_i)}{Cov(s_i, z_i)} \end{aligned}$$

Thus, the correlation between the causal variable of interest,  $s_i$ , and the instrument,  $z_i$ , cannot be zero for the **instrumental variables** procedure to work. That is, the instrument must have a clear effect on  $s_i$  in the first stage. Moreover, the first stage is the *only* reason for the relationship between  $y_i$  and  $z_i$  (i.e. the exclusion restriction).

After (4.1.11) in the text, Angrist and Pischke claim that they can “easily show” something. How, exactly?

$$\begin{aligned}
Cov(y_i, z_i) &= E(y_i z_i) - E(y_i)E(z_i) \\
&= E(E(y_i z_i | z_i)) - E(E(y_i | z_i))p \\
&= E(z_i f(z_i)) - E(f(z_i))p \\
&= \{1 \cdot f(1) \cdot p + 0 \cdot f(0) \cdot (1 - p)\} - \{f(1)p + f(0)(1 - p)\}p \\
&= \{E(y_i | z_i = 1) - E(y_i | z_i = 0)\}p(1 - p)
\end{aligned}$$

Several “tricks” are used here:

- (ii) The second equality uses the Law of Iterated Expectations.
- (iii) The third equality treats an expectation conditional on  $z_i$  as a function of  $z_i$ .
- (iv) The fourth equality uses the definition of (discrete) expected value.

### 3.1 Example: Parker et al. [2011]

Parker et al. are interested in measuring the average response of household expenditure to the arrival of a stimulus payment,  $\beta_2$ , in

$$\Delta C_i \equiv C_{i,t+1} - C_{i,t} = \sum_s \beta_{0s} \text{Month}_{s,i} + \beta'_1 \mathbf{X}_{i,t} + \beta_2 \text{ESP}_{i,t+1} + u_{i,t+1} \quad (1)$$

where  $i$  indexes households and  $t$  indexes time;  $C$  is either household consumption expenditures or their log;  $Month$  represents time fixed effects at the monthly level, “used to absorb the seasonal variation in consumption expenditures as well as the average of all other concurrent aggregate factors”; and  $X$  represents control variables to absorb some of the

preference-driven differences in the growth rate of  $C$  across  $i$ .

When Eq. (1) is estimated with OLS using all available  $i$  and the total dollar amount of payments received by  $i$  in  $t + 1$  ( $ESP_{i,t+1}$ ),  $\beta_2$  measures “the average fraction of the payment spent on the different expenditure aggregates in each column, within the three-month reference-period in which the payment was received.” The effect of a payment is identified from variation in both the *timing* and *amount* of the payment. While timing is random because it is based on Social Security Numbers, amount is not since it depends upon  $i$ ’s characteristics.

To use only variation in whether a payment was received at all in a given period (i.e. timing), Parker et al. use  $1 \cdot [ESP_{t+1} > 0]$  in lieu of  $ESP_{i,t+1}$ .  $\beta_2$  then measures “the average dollar increase in expenditures *caused* by receipt of a payment.”

To estimate a value interpretable as a “marginal propensity to spend upon the payment’s arrival without using variation in ESP amount,” Parker et al. estimate Eq. (1) using 2SLS, instrumenting  $ESP_{i,t+1}$  with  $1 \cdot [ESP_{i,t+1} > 0]$  and the other independent variables<sup>1</sup>.  $\beta_2$  then measures the “fraction of the payment that is spent within the three-month period of receipt,” as before. The possibility endogeneity of  $ESP_{i,t+1}$  may have motivated the IV approach, since  $ESP_{i,t+1}$  depends upon  $i$ ’s characteristics, some of which are not exogenous (e.g. family size).

The effect of spending is identified “by comparing the behavior of households that received payments at different times *to* the behavior of households that did not receive payments at those times” (i.e. counterfactual), controlling for age and changes in family size ( $\mathbf{X}_{i,t}$ ).

---

<sup>1</sup>Can you always instrument this way?

Since some households did not receive any payment in any period, the results still use some information that comes from comparing households that received payments to households that never did. What happens when Parker et al. exclude all households that did not report a payment in any reference quarter? When variation in ESP amount is not used, the response of spending is identified using **only** the variation in timing, conditional on receipt. Identification comes from comparing the spending of households that received payments in a given period to the spending of households that also received payments but in other periods (counterfactual).

## 4 Parallel Worlds

### 4.1 Differences-in-Differences

Differences-in-differences (DD) is a version of fixed effects estimation using aggregate data. Consider the effect of a minimum wage on unemployment. Let  $Y_{1ist}$  be fast food employment at restaurant  $i$  in state  $s$  at period  $t$  if there is a high (state) minimum wage and  $Y_{0ist}$  be fast food employment at restaurant  $i$  in state  $s$  at period  $t$  if there is a low (state) minimum wage. The DD setup relies on an additive structure for potential outcomes in the **no**-treatment state:

$$\mathbb{E}(Y_{0ist} | s, t) = \gamma_s + \lambda_t$$

That is, in the absence of a minimum wage change, employment is determined by the sum of a time-invariant state (fixed) effect and a year effect that is common across states (unit-invariant). Let  $D_{st}$  be a dummy for high-minimum-wage states, where states are indexed by

$s$  and observed in period  $t$ . Assuming that  $\mathbb{E}(Y_{1ist} - Y_{0ist} | s, t) = \beta$ , we have

$$Y_{ist} = \gamma_s + \lambda_t + \beta D_{st} + \varepsilon_{ist}$$

where  $\mathbb{E}(\varepsilon_{ist} | s, t) = 0$ . Consider two time periods, February and November, and two states, New Jersey and Pennsylvania, where New Jersey is the high minimum wage state in November.

$$\begin{aligned} \mathbb{E}[Y_{ist} | s = PA, t = 11] - \mathbb{E}[Y_{ist} | s = PA, t = 2] &= (\gamma_{PA} + \lambda_{11}) - (\gamma_{PA} + \lambda_2) \\ &= \lambda_{11} - \lambda_2 \\ \mathbb{E}[Y_{ist} | s = NJ, t = 11] - \mathbb{E}[Y_{ist} | s = NJ, t = 2] &= (\gamma_{NJ} + \lambda_{11} + \beta) - (\gamma_{NJ} + \lambda_2) \\ &= \lambda_{11} - \lambda_2 + \beta \end{aligned}$$

We have two differences. As the name suggest, we take the difference *of* these differences:

$$\begin{aligned} &(\mathbb{E}[Y_{ist} | s = NJ, t = 11] - \mathbb{E}[Y_{ist} | s = NJ, t = 2]) - (\mathbb{E}[Y_{ist} | s = PA, t = 11] - \mathbb{E}[Y_{ist} | s = PA, t = 2]) \\ &= (\lambda_{11} - \lambda_2 + \beta) - (\lambda_{11} - \lambda_2) \\ &= \beta \end{aligned}$$

which gives us the causal effect of interest.

#### 4.1.1 Example: Bertrand and Mullainathan [2003]

Bertrand and Mullainathan study managers' preferences by asking what goals managers would pursue if they were not closely monitored. States' passing of takeover legislation

constitutes treatment, and they estimate

$$y_{jkl t} = \alpha_t + \alpha_j + \gamma X_{jkl t} + \delta BC_{kt} + \varepsilon_{jkl t}$$

where  $j$  indexes firms,  $k$  indexes state of incorporation,  $l$  indexes state of location,  $t$  indexes time,  $y_{jkl t}$  is the dependent variable of interest,  $\alpha_t$  and  $\alpha_j$  are year and firm fixed effects,  $X_{jkl t}$  are control variables,  $BC_{kt}$  is a dummy variable that equals one if an antitakeover law has been passed by time  $t$  in state  $k$ , and  $\varepsilon_{jkl t}$  is an error term.  $\delta$  is the estimate of the law's effect and is the primary coefficient of interest.

The specification implicitly takes as the control group all firms incorporated in states not passing a law at time  $t$ , even if they have already passed a law or will pass one later on.

## References

- J. Angrist and J.-S. Pischke. *Mostly Harmless Econometrics*. Princeton University Press, Princeton, New Jersey, 2008.
- M. Bertrand and S. Mullainathan. Enjoying the quiet life? corporate governance and managerial preferences. *Journal of Political Economy*, 2003.
- F. Hayashi. *Econometrics*. Princeton University Press, 2000.
- J. A. Parker, N. S. Souleles, D. S. Johnson, and R. McClelland. Consumer spending and the economic stimulus payments of 2008. *Working paper*, 2011.