Empirical Methods in Applied Economics

Jörn-Steffen Pischke
LSE

October 2007

1 Fixed Effects, Differences-in-differences, and Panel Data

The key strategy to identify causal effects with regression was to control for confounding factors. With IV, the confounding factor remains unobserved, and we solved the identification problem by finding an instrument correlated with the regressor of interest but not with the confounder. These are the basic strategies to deal with confounding factors, and they form the core of the toolkit of the empirical economist. But there are variations on these themes, and strategies which place some particular structure on the confounder or the variable of interest. We will deal with these in this chapter and the next.

We start in this chapter with strategies which exploit the time structure of the data, and which put restrictions on the way the confounding factor evolves over time. These strategies are fixed effects, and its cousin differences-in-differences. We also discuss the idea of controlling for lagged dependent variables, another strategy which exploits timing. We will point out that the lagged dependent variable strategy is simply regression control, and hence distinct from fixed effects and differences-in-differences.

1.1 Individual Fixed Effects

Consider the problem of estimating the union wage differential. Let Y_{it} equal log earnings of worker i at time t, and D_{it} the individual's union status. We are interested in the causal effect of union membership (or coverage by a union contract) on earnings. However, we are concerned that union firms may hire different types of workers. For example, union firms may hire workers who are more able or more motivated on average. We are willing to make the assumption that $E(\text{Y}_{0it}|A_i,\text{X}_{it},t,\text{D}_{it}) = E(\text{Y}_{0it}|A_i,\text{X}_{it},t)$, i.e that union status is as good as randomly assigned conditional on worker ability A_i, maybe some other covariates X_{it}, like age, and an aggregate time effect (for example, the propensity of workers to join unions may change over time).

The first important assumption we have made is that the effect of ability on earnings is time invariant, i.e. ability does not evolve over time. As a result, A_i appears without a time subscript. Moreover, we postulate

$$E(\text{Y}_{0it}|A_i, \text{X}_{it}, t) = \alpha + \lambda_t + A_i'\gamma + \text{X}_{it}\delta, \tag{1}$$

i.e. the effects of ability and time enter earnings in a linear and additive fashion. These are much more structural assumptions than those necessary for regression and IV in the previous chapters. We are saying that (counterfactual) non-union earnings evolve in parallel for union and non-union members, and there is no interaction between union status and the time trend in earnings. Identification here relies a lot on these functional form restrictions but this also buys us a lot.

Finally, we assume that the causal effect of union membership is additive as well

$$E(\text{Y}_{1it}|A_i, \text{X}_{it}, t) = E(\text{Y}_{0it}|A_i, \text{X}_{it}, t) + \beta = \alpha + \lambda_t + \beta + A_i'\gamma + \text{X}_{it}\delta.$$

This implies

$$E(\text{Y}_{it}|A_i, \text{X}_{it}, t) = \alpha + \lambda_t + \beta\text{D}_{it} + A_i'\gamma + \text{X}_{it}\delta.$$

It is now obvious that we like to run the regression

$$\text{Y}_{it} = \alpha + \lambda_t + \beta\text{D}_{it} + A_i'\gamma + \text{X}_{it}\delta + \varepsilon_{it} \tag{2}$$

but we do not observe ability.[1] This regression is very similar to the one we were interested in to estimate the returns to schooling. The main difference between this regression and our earlier one, and the key to identification, is that we now observe earnings and the causal variable of interest over time (and hence both of them are indexed by i and t), the regressor of interest

[1] It is possible to allow for some heterogeneity in the fixed effects model. If the treatment effect is heterogeneous so that $\text{Y}_{it} = \alpha_i + \lambda_t + \beta_i\text{D}_{it} + \varepsilon_i$, using standard arguments as in section **??**, it is easy to see that the fixed effects estimator estimates the average treatment effect on the treated $E(\text{Y}_{1i} - \text{Y}_{0i}|\text{D}_{it} = 1)$. The important restriction of the model is that α_i and λ_t are additive and there is no heterogeneity in λ_t.

D_{it} varies over time, while the unobserved confounder A_i is time invariant. Define

$$\alpha_i = \alpha + A_i'\gamma.$$

We can then rewrite (2) as

$$\text{Y}_{it} = \alpha_i + \lambda_t + \beta\text{D}_{it} + \text{X}_{it}\delta + \varepsilon_{it}. \tag{3}$$

The model in equation (3) can be estimated as a *fixed effects model*, i.e. treating α_i as a parameter to be estimated. We have estimated the regression we want without observing A_i directly!

In practice, with many individuals it is typically not feasible to estimate the coefficients on the individual dummy variables in equation (3) directly. Instead, the model is either estimated by differencing out the fixed effect

$$\Delta\text{Y}_{ist} = \Delta\lambda_t + \beta\Delta\text{D}_{it} + \Delta\text{X}_{it}\delta + \Delta\varepsilon_{it} \tag{4}$$

or by taking deviations from means. Note that taking means across time for each individual yields

$$\overline{\text{Y}}_i = \alpha_i + \overline{\lambda} + \beta\overline{\text{D}}_i + \overline{\text{X}}_i\delta + \overline{\varepsilon}_i$$

so that

$$\text{Y}_{it} - \overline{\text{Y}}_i = \lambda_t - \overline{\lambda} + \beta\left(\text{D}_{it} - \overline{\text{D}}_i\right) + \left(\text{X}_{it} - \overline{\text{X}}_i\right)\delta + \varepsilon_{it} - \overline{\varepsilon}_i,$$

which also sweeps out the individual effect. The deviations from means estimator for β and δ is identical to the dummy variables estimator, but it is only identical to the first difference estimator if there are only two periods. For more periods, if ε_{it} is iid, then the difference estimator introduces serial correlation in the new error $\Delta\varepsilon_{it}$. This would have to be accounted for in calculating the covariance matrix. Regression packages will typically implement the deviations from means estimator, with an appropriate adjustment for the degrees of freedoms lost in estimating the N individual level means. This estimator will often be referred to as the within estimator, and the procedure as absorbing the α_is.[2]

Our discussion highlights that controlling for A_i is central here, since we believe that A_i and D_{it} are correlated. We know from the omitted variables bias formula how the regression coefficient from the short regression without

[2] Notice that the individual fixed effects (estimates of which could be recovered from the residuals after applying the deviations from means estimator) are not being estimated consistently in a panel where the number of periods T is fixed while $N \to \infty$.

A_i or α_i differs. Nevertheless, the short regression has a life of its own in the econometrics literature. The literature realizes that the omission of a time invariant factor from the regression will lead to correlated residuals across observations, i.e.

$$\begin{aligned} \text{Y}_{it} &= \lambda_t + \beta \text{D}_{it} + \text{X}_{it}\delta + u_{it} \\ u_{it} &= \alpha_i + \varepsilon_{it}. \end{aligned} \tag{5}$$

This case is called the *random effects model*, where α_i is now considered part of the residual, not a parameter to be estimated. Interpreting estimates from (5) causally requries the stronger assumption $E(\text{Y}_{0it}|\text{X}_{it}, \lambda_t, \text{D}_{it}) = E(\text{Y}_{0it}|\text{X}_{it}, \lambda_t)$, where we are no longer conditioning on A_i. The random effects model involves estimating the short regression by GLS to deal with the correlation in the residuals. While there may be good reasons why we want to run the short regression, we prefer to apply OLS in this case, and fix the standard errors as discussed later. The reasons are the same as those for not weighting regressions to address heteroskedasticity . The CEF may not be linear and OLS remains the BLP. [3]

Freeman (1984) estimates union wage effects under the assumption that selection into union status is based on fixed individual characteristics. His Table 6 displays estimates for the union wage effects from four data sets used by Freeman using both the cross section and the fixed effects estimator. The cross section estimates are typically higher (in the range of 0.15 to 0.25) than the fixed effects estimates (in the range 0.10 to 0.20). This may indicate positive selection of union workers.

One problem with fixed effects estimation is that it tends to accentuate measurement error problems. The reason is that many economic variables like union status are fairly persistent (a union worker this year is most likely also a union worker next year) while measurement error is often much more transitory (union affiliation may be misreported or miscoded this year but not next year). In a cross-section, the union status may only be misreported for relatively few workers compared to the total number of unionized workers. But many of the observed transitions in and out of unions, which is the variation the fixed effects estimator exploits, will actually be spuriously

[3] The distinction between fixed and random effects is not really critical for identification of the treatment effect β. Key is whether we assume $E(\text{Y}_{0i}|\alpha_i,\text{X}_{it},\text{D}_{it}) = E(\text{Y}_{0i}|\alpha_i,\text{X}_{it})$ or $E(\text{Y}_{0i}|\text{X}_{it},\text{D}_{it}) = E(\text{Y}_{0i}|\text{X}_{it})$. Even with $E(\text{Y}_{0i}|\alpha_i,\text{X}_{it},\text{D}_{it}) = E(\text{Y}_{0i}|\alpha_i,\text{X}_{it})$ it is possible to treat α_i as a random effect for estimation purposes, albeit one correlated with D_{it}. See Chamberlain (1984) for details on how this can be done and Card (1996) for an application to union wage effects.

created by the measurement error, while there are few true transitions. In relative terms, the measurement error problem is much worse when using fixed effects, and this may also explain the difference between the cross-sectional and the fixed effects estimates reported by Freeman (1984).[4]

The insights from the discussion of measurement error are more general. Taking out fixed effects may remove a lot of the variance in the variable of interest D_{it} or S_{it}. Our assumption is that this variance is harmful in our exercise because (part of) it is correlated with the individual fixed effect. However, as we have seen in the measurement error case, it is quite possible that taking out fixed effects removes both good and bad variation. This is particularly troubling if the fixed effects strategy is imperfect, and some bad variation is left in the fixed effects estimates. Because much of the good variation has been filtered out, the consequences of the bad variation also get magnified. Hence, it easy to throw out the variation in the baby with the omitted variables bathwater.

An example for this type of concern is related to twin based estimates of returns to schooling. Ashenfelter and Krueger (1994) and Ashenfelter and Rouse (1998) present estimates of the returns to schooling among twins, controlling for twin pair fixed effects. Hence, these estimates compare the difference in earnings across twins to differences in their schooling. The idea is that ability is related to either genetics, family background, or school environment, which are all captured by the twin fixed effect. The authors argue that any remaining difference in schooling should therefore be unrelated to ability.

But how do differences in schooling come about between individuals who are otherwise so much alike? Bound and Solon (1999) point out that there are small differences between twins, with first borns (here differences in birth date are measured in minutes) typically having higher birthweight but also higher IQ scores. While these twin differences are not large, neither is the difference in schooling. Hence, a small amount of unobserved ability differences among twins could be responsible for a large amount of bias in the resulting estimates.

1.2 Differences-in-differences: Pre and Post, Treatment and Control

The fixed effects strategy requires panel data on the individual units of observation. Sometimes the regressor of interest only varies at a well defined

[4] See Griliches and Hausman (1986) for a more complete analysis of measurement error in panel data.

level of aggregation. In this particular case, we do not need panel data to apply the fixed effects strategy. Instead, we can simply rely on repeated cross-section samples from the same underlying population. This obviously extends the realm of problems we can apply the strategy to.

To take a concrete example again, say we are interested in the effect of the minimum wage on employment. A number of studies have exploited changes in minimum wages at the state level, and we will use the example of Card and Krueger (1994) here, who studied the increase in the minimum wage in New Jersey from \$4.25 to \$5.05. This change took effect on April 1, 1992. Card and Krueger collected data on employment at fast food restaurants in New Jersey in February and in November 1992. They also collected similar data on restaurants in eastern Pennsylvania, the neighboring state, for the same period. The minimum wage in Pennsylvania remained at \$4.25 throughout this period. While Card and Krueger actually have a panel of restaurants over time this is not really necessary to apply their basic strategy.

The fixed effects strategy applied at a group level is typically referred to as differences-in-differences (DD).[5] It is simple to restate the assumptions underlying DD identification as a special case of the fixed effects scenario. Let

$$\begin{aligned}
\mathrm{Y}_{1ist} &= \text{fast food employment at restaurant } i \text{ and period } t \\
& \text{for a high minimum wage} \\
\mathrm{Y}_{0ist} &= \text{fast food employment at restaurant } i \text{ and period } t \\
& \text{for a low minimum wage}
\end{aligned}$$

be the counterfactual outcomes. We have also indexed restaurants by the state s they are in. In the standard fixed effects case we made the assumption $E(\mathrm{Y}_{0ist}|\alpha_i, t, \mathrm{D}_{st}) = E(\mathrm{Y}_{0ist}|\alpha_i, t)$. This says that (counterfactual) employment at restaurant i for the low minimum wage is independent of the actual level of the minimum wage D_{st}, once we condition on a restaurant specific effect and a time effect. Furthermore, we assumed a linear additive structure for the restaurant and the time effect.

The DD model simply relaxes this assumption to $E(\mathrm{Y}_{0ist}|s, t, \mathrm{D}_{st}) = E(\mathrm{Y}_{0ist}|s, t)$, and we now have

$$E(\mathrm{Y}_{0ist}|s, t) = \gamma_s + \lambda_t \tag{6}$$

[5] The DD idea is at least as old as IV. Kennan (1995) reports that a BLS report in 1915 used this methodology already to study the employment effects of the minimum wage (Obenauer and von der Nienburg, 1915).

where s denotes the state (New Jersey or Pennsylvania) and t denotes the period (February, before the minimum wage increase or November, after the increase). This says that in the absence of a minimum wage change employment is given by an additive state effect and a time effect, which is the same in both states. D_{st} only varies at the state level over time, so we only need to worry about confounders at the state and time level as well. Whether employment levels at restaurants within a state are the same or they differ is immaterial. Conditional on s, restaurant specific counterfactual employement Y_{0ist} cannot be correlated with the minimum wage.

As before, we can write

$$\mathrm{Y}_{ist} = \gamma_s + \lambda_t + \beta \mathrm{D}_{st} + \varepsilon_{ist} \tag{7}$$

where D_{st} is the dummy for the treatment, a high minimum wage, which was in place in New Jersey in November. Using (7) it is easy to see that

$$E[\mathrm{Y}_{ist}|s = PA, t = Nov] - E(\mathrm{Y}_{ist}|s = PA, t = Feb) = \lambda_{Nov} - \lambda_{Feb}$$

and

$$E(\mathrm{Y}_{ist}|s = NJ, t = Nov) - E(\mathrm{Y}_{ist}|s = NJ, t = Feb) = \lambda_{Nov} - \lambda_{Feb} + \beta.$$

The population difference-in-difference

$$\begin{gathered}[E(\mathrm{Y}_{ist}|s = PA, t = Nov) - E(\mathrm{Y}_{ist}|s = PA, t = Feb)] \\ - [E(\mathrm{Y}_{ist}|s = NJ, t = Nov) - E(\mathrm{Y}_{ist}|s = NJ, t = Feb)] = \beta\end{gathered}$$

identifies the treatment effect, hence the name of the strategy. β can easily be estimated by the sample analogue.

Table 3 in Card and Krueger (1994) displays average employment at fast food restaurants in New Jerseay and Pennsylvania before and after the change in the New Jersey minimum wage. There are four cells, as well as the state differences, the changes over time, and the difference-in-difference. Employment in PA restaurants is somewhat higher than in NJ in Februrary and falls by November. Employment in NJ, in contrast, increases slightly. This results in a positive estimate for the difference-in-difference, the opposite from what we might expect if restaurants were moving up their labor demand curve as the minimum wage increases.

The key identifying assumption here is that employment trends are the same in both states in the absence of the treatment. Hence, the employment trend in the treatment state has the same slope as in the control state,

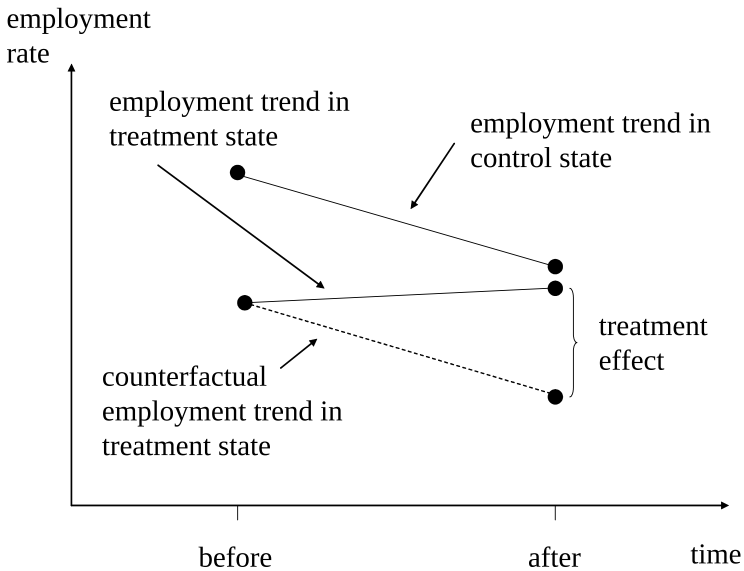


Figure 1: Identification in the differences-in-differences model

but is displaced to account for the different employment levels before the treatment, as in figure 1. A similar figure applies to the fixed effects case, where the control state is simply replaced by a control individual.

Depending on the context, there may be different forms of the linear and additive state and time effects assumption (6) which are reasonable. Card and Krueger (1994) assume that it is the levels of employment which evolve in the same way in PA and NJ. If employment levels were somewhat different ex ante, an equally reasonable assumption might be that the log of employment evolves in the same way absent minimum wage changes, or

$$E(\log \mathrm{Y}_{0ist}|s,t) = \gamma_s + \lambda_t.$$

This implies the multiplicative relationship

$$\mathrm{Y}_{0ist} = e^{\gamma_s+\lambda_t}\varepsilon_{ist}.$$

which is different from (6), and hence involves a different assumption about the counterfactual trends. If one assumption is true, the other one must be necessarily false. Since the assumption is about an unobserved counterfactual, it is not testable with the data we have examined so far.[6]

Much of the recent discussion of DD models has been about ascertaining whether the underlying assumption of equal trends in the absence of treatment is a reasonable one. One possible way to look at this issue is if there are data available on multiple periods. For a later update of their study, Card and Krueger (2000) obtained time series of administrative payroll data for restaurants in New Jersey and Pennsylvania. These data are plotted in Figure 2 in their paper. The vertical lines indicate the dates when their original surveys were conducted, and the third vertical line denotes the increase in the federal minimum wage to $4.75 in October 1996, which affected Pennsylvania but not New Jersey.

The administrative data also show a slight decline in employment from February to November 1992 in Pennsylvania, and little change in New Jersey. However, the data also reveal a large amount of ups and downs in employment in the two states in other periods. The employment trends in periods when the minimum wage was constant are often not the same in the two states. In particular, employment in New Jersey and Pennsylvania was rather similar at the end of 1991. Relative employment in Pennsylvania declined over the next three years (at least using the larger set of 14 PA counties), with much of this trend occuring at periods unrelated to the 1992 minimum wage change. Hence, eastern Pennsylvania restaurants may not be a very good control group for New Jersey restaurants, because employment trends differ somewhat in periods with no treatment.

A more encouraging example comes from the paper by Hastings (2004). She studies the effect of the competitive environment in the retail gasoline market on gasoline prices. She uses the takeover of a large number of previously independent Thrifty gas stations in southern California in September 1997 by ARCO, a large, vertically integrated gasoline retailer. Gas stations belonging to a vertically integrated retailer typically sell gasoline at a higher price than independent stations. The hypothesis is that the presence of more independent gas stations in a local market increases cometition and therefore lowers the market price of competitors as well. Hastings investigates this hypothesis by looking at the prices of other gas stations before and after the ARCO purchase of the Thrifty stations. The treatment group in her setup are gas stations which are located near a Thrifty station, while the control

[6] Other identifying assumptions are possible. Instead of applying (6) to means, it could be applied to quantiles as in Poterba, Venti and Wise (1995) and Meyer, Viscusi, and Durbin (1995). Athey and Imbens (2006) present a non-parametric DD estimator which makes weaker assumptions than the standard DD model.

group are gas stations with no Thrifty station nearby.

Figures 1a and 1b in the Hastings (2004) paper plot gasoline prices for Thrifty competitors and other stations during 1997 in two metropolitan areas. While there are large price changes from quarter to quarter, prices move in parallel throughout the period, except between June and October, the period of the ARCO purchase. Prices at Thrifty competitors increase by more during this period than at comparison stations. The graphs are highly compelling that the comparison stations provide a good control group for the Thrifty competitors.[7] Being able to produce pictures like these should be the goal of any good DD analysis.

1.2.1 Regression DD

As in the case of the fixed effects model, we can use regression to estimate (7). If there are only two groups and two periods, then

$$\begin{aligned} \mathrm{Y}_{ist} &= \gamma_s + \lambda_t + \beta \mathrm{D}_{st} + \varepsilon_{it} \\ &= \alpha + \gamma 1(s = NJ) + \lambda 1(t = Nov) + \beta 1(s = NJ) \cdot 1(t = Nov) + \varepsilon_{ist} \end{aligned}$$

where $1(\cdot)$ is the indicator function. There are two main effects, a dummy for New Jersey, and a dummy for November, and an interaction term for observations from New Jersey in November (as well as a constant). Since this is a saturated model, taking conditional expectations for different states and periods, and subtracting easily yields

$$\begin{aligned} \alpha &= E(\mathrm{Y}_{ist}|s = PA, t = Feb) = \gamma_{PA} + \lambda_{Feb} \\ \gamma &= E(\mathrm{Y}_{ist}|s = NJ, t = Feb) - E(\mathrm{Y}_i|s = PA, t = Feb) = \gamma_{NJ} - \gamma_{PA} \\ \lambda &= E(\mathrm{Y}_{ist}|s = PA, t = Nov) - E(\mathrm{Y}_i|s = PA, t = Feb) = \lambda_{Nov} - \lambda_{Feb} \\ \beta &= [E(\mathrm{Y}_{ist}|s = PA, t = Nov) - E(\mathrm{Y}_i|s = PA, t = Feb)] \\ &\quad - [E(\mathrm{Y}_{ist}|s = NJ, t = Nov) - E(\mathrm{Y}_i|s = NJ, t = Feb)]\,. \end{aligned}$$

The regression formulation of the difference-in-difference model is useful for a variety of reasons. First of all, it is a convenient way of obtaining the DD estimate, standard errors, and t-statistics. Second, it is easy to incorporate additional states or periods in the analysis now. For example, instead of just comparing the impact of the change in the minimum wage in New Jersey in a particular period, we may want to look at the impact comparing many state pairs, or comparing different periods. In this case, the

[7] But see the comment by Taylor, Kreisle, and Zimmerman (2007) who try to replicate these results with different data and reach somewhat different conclusions.

formulation of the model would not change but s and/or t would take on more than two values. For example, we could use the data from Figure 2 in Card and Krueger (2000) for New Jersey and Pennsylvania up to 1997, and incorporate the increase in the federal minimum wage in 1996. D_{st} would be 1 for New Jersey after April 1992, and for Pennsylvania after October 1996.

But the minumum wages in New Jersey and Pennsylvania were actually not the same after October 1996. The minimum wage was \$5.05 in New Jersey and \$4.75 in Pennsylvania. This immediately suggests a third advantage of the regression formulation. The treatment may not be binary but multivalued since different states could have different levels of the minimum wage, or the same nominal minimum wage may have a different impact depending on the distribution of wages in the state. The regression formulation becomes

$$\text{Y}_{ist} = \gamma_s + \lambda_t + \beta \text{S}_{st} + \varepsilon_{ist} \tag{8}$$

where the variable S_{st} is either the level of the minimum wage or a measure of the "bite" of the minimum wage in state s at time t. Despite the continuous nature of the treatment, this formulation still retains the basic features of the differences-in-differences model.

An example of the model in (8) is the paper by Card (1992). He studies the effect of the federal increase in the minimum wage in April 1990 using all the US states. The federal minimum wage was \$3.35 before the increase, and was raised to \$3.80. Some states already had state minimum wages of \$3.80 or higher at the time of the federal increase. Moreover, the same increase will have more of an effect in a low wage state, where many workers are subject to the minimum, than in a high wage state. Card's measure of the impact of the increase of the minimum wage is the fraction of workers who are paid less than \$3.80 just before the increase of the minimum wage, something he calls the "fraction of affected workers."

There are still only two time periods in the Card (1992) setup, before and after the minimum wage increase. Difference eq. (8) over time to obtain

$$\begin{aligned} \Delta \text{Y}_{st} &= \lambda_t - \lambda_{t-1} + \beta \text{S}_{st} + \Delta \varepsilon_{st} \\ &= \lambda + \beta \text{S}_{st} + \Delta \varepsilon_{st}. \end{aligned}$$

OLS estimation of the differenced equation is the same as OLS of the levels equation (8) with dummy variables when there are only two periods (but this isn't true for more periods). The difference in the time effect simply becomes a constant term $\lambda = \lambda_t - \lambda_{t-1}$, so the differenced equation is a standard bivariate relationship for the outcome and the treatment variable. Furthermore, the data here are not individual firms as in Card and Krueger

(1994) but rather averages of employment for a group like teenagers in state s at time t. Hence, we have indexed all the variables simply by s and t.

Table 3 in Card (1992) shows that wages increased more in states where the minimum wage had more bite (column 1). Employment, on the other hand, seems largely unaffected by the bite of the minimum wage as can be seen in column (3).

Given the multiple contrasts from using 51 state level changes (50 states plus the District of Columbia) instead of two we have gained many degrees of freedom. So we can use up a few of them by adding controls for other state specific factors. The idea is now that the counterfactual employment in the absence of a change in the minimum wage is

$$E[\mathrm{Y}_{0ist}|s, t, \mathrm{X}_{st}] = \gamma_s + \lambda_t + \mathrm{X}_{st}\delta.$$

Regression offers a straightforward way to do this, which is the fourth advantage of the regression formulation of DD. Card uses the wages and employment outcomes of teenagers, a group typically strongly affected by the minimum wage but also by other time varying factors like the business cycle. Card therefore controls for adult employment trends in the state in X_{st}. Different states do not need to have the same counterfactual employment trends in the absence of the minimum wage anymore. These trends only have to be the same conditional on adult employment. A key requirement for the variables X_{st} is that they are not "bad controls" as discussed before. I.e. adult employment cannot be affected directly by the minimum wage itself. This suggests that we may want to be careful in constructing such variables (for example, by using only older or more educated male workers, whose wage distribution lies well above the minimum wage).

Card runs regressions of the form

$$\Delta\mathrm{Y}_{st} = \lambda + \beta\mathrm{S}_{st} + \Delta\mathrm{X}_{st}\delta + \Delta\varepsilon_{st}.$$

The results including changes in the adult employment to population rate are shown in columns (2) and (5) in his Table 3. There is little change in the coefficient on the fraction of affected teens, which is a comforting result.

In many applications (as in the New Jersey-Pennsylvania study by Card and Krueger, 1994) the outcome is measured at the individual level rather than at the state and time level. In fact, the employment rates in Card (1992) come from the micro data of the Current Population Survey. Instead of aggregating everything to the state level it is also possible to run the regression at the individual level, as we did before in (7) and (8). Returning to the model in levels this yields

$$\mathrm{Y}_{ist} = \gamma_s + \lambda_t + \beta \mathrm{S}_{st} + \mathrm{X}_{ist}\delta + \varepsilon_{ist}. \tag{9}$$

Variables in X_{ist} could now be either individual level variables like race or family income or time varying variables at the state level. Only the variation in the controls at the state and time level matters for identification (because S_{st} is orthogonal to any within state and time variation). Including individual level variables may therefore not primarily help to control for confouning trends, but may help reduce the variance of ε_{ist} and hence the standard errors of the estimate of β.

In a model with multiple treatment groups (states) and multiple periods, it becomes more difficult to provide a simple visual inspection for the evolution of state specific trends in the periods when there is no treatment, as in Card and Krueger (2000) and Hastings (2004). Of course, identical counterfactual trends in treatment and control states, conditional on any X_{ist}, is still the identifying assumption. One way to test this assumption is to allow for leads and lags of the treatment. In order to see how this works, return to the model with a binary treatment D_{st}. Let k_s be the time at which the treatment is being switched on in state s. Our regression model is now

$$\mathrm{Y}_{ist} = \gamma_s + \lambda_t + \sum_{j=-m}^{q} \beta_j \mathrm{D}_{st}(t = k_s + j) + \mathrm{X}_{ist}\delta + \varepsilon_{ist}.$$

The "treatment variables" $\mathrm{D}_{st}(t = k_s + j)$ take on the value 1 in state s and period t if we are exactly j periods after the start of the treatment in state s. Instead of a single treatment effect, we now also include m "leads" and q "lags" of the treatment effect for a total of $m + q + 1$ treatment dummies. β_j is the coefficient on the jth lead or lag. A test of the DD assumption is $\beta_j = 0\ \forall j < 0$, i.e. the coefficients on all leads of the treatment should be zero.[8] Moreover, the $\beta_j, j \geq 0$ may not be identical. For example, the effect of the treatment could accumulate over time, so that β_j increases in j.

An example of this approach is the paper by Autor (2003). He investigates the effect of employment protection on outsourcing by firms. To this end, he relates the employment of temporary help workers in a state to indicators whether the state courts had adopted more stringent exceptions to the employment at will doctrine. Figure 3 in his paper plots the β_j coefficients. These coefficients are zero in the two years before the courts adopted

[8] This is essentially a Ganger causality test: we expect the treatment to affect the outcome only after the treatment actually took place but not before.

the new rule, increasing in the first few years after the adoption, and then flat. This indicates that the DD strategy seems successful in this context.

An alternative way to probe the robustness of the DD identification is to include state specific parametric time trends among the regressors in X_{ist}. Of course, this is only possible with multiple periods again. This is done, for example, in the paper by Besley and Burgess (2004). They examine the effect of labor regulation on the performance of firms in Indian states. Different states change the regulatory regime at different times, giving rise to a DD design. Their Table IV shows the key results. Column (1) shows that labor regulation leads to lower output per capita. In columns (2) and (3) they include state specific-time varying regressors like development expenditures per capita. This is a similar strategy to using the adult employment rate in Card (1992) above. This affects the estimates little. However, when they include linear state specific trends in column (4) the coefficient on labor regulation drops to zero. This suggests that the introduction of additional labor regulation correlates with other trends in state level output, and it is not possible to disentangle the causal effect of the regulation from these underlying trends. Effectively, after including a parametric trend, the identification hinges on there being a (relatively) sharp change in the outcome at the date of the treatment.

Controlling for state specific trends only works well when there is a sufficient sample period available before the treatment starts. This is particularly true when there is a dynamic response to the treatment (for example, it takes some time for treatment to take effect or the initial effects are large but they peter out later). In this case it is not particularly informative looking at trends after the treatment. The paper by Wolfers (2003) vividly illustrates this. He discusses the impact of unilateral divorce laws on divorce in the US. Before the 1970s, a divorce was only possible if both spouses agreed. In the 1970s, states introduced unilateral divorce laws, which allow a divorce if one spouse wants the divorce. An influential paper by Friedberg (1998) estimated the effect of the introduction of unilateral divorce on divorce rates, and found a sizeable and lasting effect.

Wolfers (2003) reanalyzes the data, and points out that much of the result hinges on Friedberg's treatment of state specific trends. Figure 5 in his paper illustrates the problem for one state: California. Friedberg's sample starts only one year before California introduced unilateral divorce. Her estimate of the California specific trend therefore relies almost completely on the post-law trend in the state. Wolfers, using a sample going back to the late 1950s, demonstrates that California's pre-exisiting trend was very different from that after the treatment. Extrapolating this pre-existing trend results in a very different estimate of the divorce effect.

1.2.2 Picking Controls

In the discussion so far we have labeled the two dimensions s and t in the DD setup "states" and "time." While there are many applications were the treatment or policy is time varying at the regional level, the identification strategy is not limited to these dimensions. s and t can be any two dimensions, so that treatment only takes place for particular combinations of s and t. While contrasts simply across s or t may not plausibly identify the treatment effect, this may be more likely for the DD estimator. Instead of states, s may denote different demographic groups, some of which are affected by a policy and others are not. For example, unemployment benefits may be changed differentially for various age groups. Anti-discrimination or job protection legislation may not apply to firms below a particular size cutoff but could be extended to additional firms. Welfare benefits may only be payable to low income families with a single parent, but not other demographic groups. The excluded groups may or may not be appropriate comparison groups. It is often the main challenge for the researcher to identify a particularly appropriate comparison group, which satifies the necessary identifying assumption, that the treated groups would behave similarly as t is varied as the untreated groups (and remember that t may not necessarily be time).

One potential pitfall in defining treatment and control groups in a DD setup is that s or t may be directly affected by the treatment. For example, if s is a state as before, we may be concerned that the policy induces some inter-state migration. Hence, the population resident in the treatment state before and after the policy becomes effective may not be identical. Say state $s = 1$ lowers wefare benefits, and this leads some poor families to move to another state $s = 0$, which forms the control group. We are interested in estimating how the lower benefits affect the fraction of the population on welfare. Also suppose that in the absence of the policy change welfare receipt would not have changed over time. In this case, the change of welfare receipt over time in the control state does not provide a valid counterfactual anymore: in the absence of the policy, welfare receipt in state $s = 1$ would have been unchanged. Instead, welfare receipt in state $s = 0$ increases because of the induced welfare migration. Hence, we are overestimating the effect of the policy.

Sometimes this problem can be overcome if we know where an individual starts out. Say we know the state of residence in the period before treatment, or we know the individual's state of birth. Either of these is immutable, i.e. cannot be affected by the treatment itself or by individuals' reaction to the treatment. If we assign individuals to the treatment or control group on the basis of this immutable characteristic, we can circumvent the problem outlined above. This introduces a new problem, however, that the new dimension, say state of birth, is not really the correct delineation for the treatment. Some individuals "born" (or previously residing) in the treatment state move, and we would now assign them to the treatment group, even so they are not affected by the policy after their move. However, this divergence of treatment group assignment and actual treatment can easily be addressed by using IV methods as discussed before. Treatment status assigned on the basis of state of birth is like an "intention to treat."[9]

Rather than using single differences, the treatment assignment rule may sometimes suggest a triple or higher order differences setup for the estimation. An example is the extension of Medicaid coverage in the U.S. studied by Yelowitz (1992). Medicaid, health insurance for the poor, was traditionally tied to eligibility for AFDC, the cash welfare program. In the late 1980s, various states introduced extensions of Medicaid coverage for families with earnings high enough so they would not qualify for AFDC anymore. These extensions happened at different times for different states. This gives rise to a classical DD design exploiting changes across states over time. However, different states introduced these extensions for children in different age groups. Hence, the age of the youngest child is a third dimension along which the treatment varies. Yelowitz analyzes employment effects of these extensions using the model

$$\mathrm{Y}_{iast} = \gamma_{st} + \lambda_{at} + \theta_{as} + \beta \mathrm{D}_{ast} + \mathrm{X}_{iast}\delta + \varepsilon_{iast}.$$

There are now three dimensions, state (s), time (t), and age of the youngest child (a). This allows the researcher to control non-parametrically for state specific shocks γ_{st}, i.e. each time period receives a separate dummy variable in each state. In order to only exploit the triple differences, it is also necessary to include interactions of age and time effects λ_{at}, and age and state effects θ_{as}. Sometimes it may not be possible to identify the effect of the treatment with such a rich set of controls, and some of the second level interactions may have to be excluded. However, when the full set of controls is feasible, triple differences may allow for a more credible analysis.

[9] Sometimes this problem can be much thornier. If the treatment group is large enough then treatment may affect prices and individuals in the control group react to these price changes. Such general equilibrium effects are much more difficult to address.

1.3 Fixed Effects versus Lagged Dependent Variables

In the previous sections we have adressed indentification in the case where we believe a variant of the following condition to hold: $E(\text{Y}_{0it}|\alpha_i,\text{X}_{it},t,\text{D}_{it}) = E(\text{Y}_{0it}|\alpha_i,\text{X}_{it},t)$. This is tantamount to saying that selection into treatment is on the basic of a vector of fixed but unobservable characteristics of unit i, as in the case when union status depends on individual ability and motivation, or certain states are more likely to raise their minimum wage but this does not depend on the time-varying employment conditions in the state. An alternative scenario is that selection is based on the past realization of the outcome variable under study. For examle, we might be interested in the earnings of participants in a training program, and trainees are selected on the basis of their pre-training earnings. This implies $E(\text{Y}_{0it}|\text{Y}_{it-h},\text{X}_{it},t,\text{D}_{it}) = E(\text{Y}_{0it}|\text{Y}_{it-h},\text{X}_{it},t)$ where selection was determined h periods ago. With panel data, running a regression of the form

$$\text{Y}_{it} = \alpha + \theta\text{Y}_{it-h} + \lambda_t + \beta\text{D}_{it} + \text{X}_{it}\delta + \varepsilon_{it} \tag{10}$$

identifies the treatment effect.[10]

In many circumstances knowledge of the selection process will give some clear indication whether either $E(\text{Y}_{0it}|\alpha_i,\text{X}_{it},t,\text{D}_{it}) = E(\text{Y}_{0it}|\alpha_i,\text{X}_{it},t)$ or $E(\text{Y}_{0it}|\text{Y}_{it-h},\text{X}_{it},t,\text{D}_{it}) = E(\text{Y}_{0it}|\text{Y}_{it-h},\text{X}_{it},t)$ might be a more reasonable assumption. But sometimes we simply don't know. Does it matter whether we estimate regression (3) or (10)? The answer is yes since neither the model (3) nests (10) nor vice versa. It is sometimes alleged that setting $\theta = 1$ in (10) results in (3). Notice that setting $\theta = 1$ yields

$$\Delta\text{Y}_{it} = \alpha + \lambda_t + \beta\text{D}_{it} + \text{X}_{it}\delta + \varepsilon_{it}$$

which is not the same as (4). It is necessary to difference the regressors in order to obtain the fixed effects estimats while $\theta = 1$ in (10) only amounts to differencing the dependent variable. In the appendix, we show more formally what happens when we estimate (3) when (10) is the correct regression, and vice versa.

One possible solution to this is to be agnostic, assume only $E(\text{Y}_{0it}|a_i,\text{Y}_{it-h},\text{X}_{it},\text{D}_{it}) = E(\text{Y}_{0it}|\alpha_i,\text{Y}_{it-h},\text{X}_{it})$, and estimate

$$\text{Y}_{it} = \alpha_i + \theta\text{Y}_{it-h} + \lambda_t + \beta\text{D}_{it} + \text{X}_{it}\delta + \varepsilon_{it} \tag{11}$$

which allows for both a fixed effect and a lagged dependent variable. One problem with this model is that the conditions for fixed effects estimators

[10] There are non-parametric methods for this case as well, see Abadie, Diamond, and Hainmueller (2007) for an example.

with lagged dependent variables to be consistent are much more demanding than those without. For example, with a once lagged dependent variable, i.e. setting $h = 1$, the differenced equation is

$$\Delta \mathrm{Y}_{it} = \theta \Delta \mathrm{Y}_{it-1} + \Delta\lambda_t + \beta \Delta \mathrm{D}_{it} + \Delta \mathrm{X}_{it}\delta + \Delta\varepsilon'_{it}. \quad (12)$$

While this still removes the fixed effect it introduces a new problem. Even if the residual ε_{it} in the regression (11) is serially uncorrelated, ε_{it-1} is correlated with Y_{it-1} by construction, hence the residual $\Delta\varepsilon'_{it}$ from running (12) cannot be the same as $\Delta\varepsilon_{it}$. This means that the coefficients from the regression (12) will not be consistent for the population regression coefficients in (11). This problem was first pointed out by Nickell (1981). There are various potential solutions to this problem, typically involving instrumenting $\Delta \mathrm{Y}_{it-1}$ with further lags of the dependent or independent variables, which are uncorrelated with $\Delta\varepsilon_{it}$ under some assumptions about the error process.[11]

2 References

Alberto Abadie (2005) Semiparametric Difference-in-differences, Review of Economic Studies, 72, 1-19

Alberto Abadie, Alexis Diamond, Jens Hainmueller (2007) Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program, NBER Working Paper No. 12831, January 2007.

Athey, Susan and Guido Imbens (2006) Identification and Inference in Nonlinear Difference-in-Differences Models, Econometrica, 74, March 2006, 431-497.

Meyer, Bruce, Kip Viscusi, and David Durbin (1995) Workers' Compensation and Injury Duration: Evidence from a Natuarl Experiment, American Economic Review 85, June 1995, 322-340.

Poterba, James, Steven Venti, and David Wise (1995) Do 401(k) contributions crowd out other personal savings? Journal of Public Economics, 58, 1-32.

Richard Blundell, Monica Costa Dias, Costas Meghir and John Van Reenen (2004) Evaluating the employment impact of a mandatory job search assistance program, Journal of the European Economic Association, June 2004

[11] See Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991), Blundell and Bond (1998) for examples on how to approach this problem.

David Card, “Using Regional Variation to Measure the Effect of the Federal Minimum Wage”, Industrial and Labor Relations Review, vol. 46, October 1992, pp. 22-37

David Card and Alan Krueger, “Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania”, American Economic Review 84, no. 4, September 1994, pp 772-784.

David Card and Alan Krueger, “Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania: Reply”, American Economic Review 2000, 1397-1420

Justine Hastings, “Vertical Relationships and Competition in Retail Gasoline Markets: Empirical Evidence from Contract Changes in Southern California”, American Economic Review 94, March 2004, 317-28.

John Kennan, The Elusive Effects of Minimum Wages, Journal of Economic Literature, Vol. 33, No. 4. (Dec., 1995), pp. 1950-1965.

Marie Obenauer and Berta von der Nienburg (1915) Effect of Minimum Wage Determinations in Oregon, Bulletin of the US Bureau of Labor Statistics, Number 176, Washington DC: US GPO, 1915.

Wolfers, Justin, “Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results,” NBER Working Paper 10014, October 2003.

L Friedberg, Did Unilateral Divorce Raise Divorce Rates? Evidence from Panel Data, The American Economic Review, 1998, 608-627

David Autor, “Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing”, Journal of Labor Economics, vol. 21, January 2003, pp. 1-42.

Orley Ashenfelter and David Card, “Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs”, The Review of Economics and Statistics, vol. 67, no. 4 (Nov. 1985), pp. 648-660.

Chamberlain, Gary, “Panel Data,” in Zvi Griliches and Michael Intriligator (eds.) Handbook of Econometrics, vol II, Amsterdam: North Holland, 1984, chapter 22.

Douglas Holtz-Eakin; Whitney Newey; Harvey S. Rosen, Estimating Vector Autoregressions with Panel Data, Econometrica > Vol. 56, No. 6 (Nov., 1988), pp. 1371-1395

Manuel Arellano; Stephen Bond, Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, The Review of Economic Studies > Vol. 58, No. 2 (Apr., 1991), pp. 277-297

Richard Blundell and Stephen Bond, Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics vol 87, Aug 1998, Pages 115-143

D. Card, "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," Econometrica, 64, July 1996, 957-979.

Richard Freeman, "Longitudinal Analyses of the Effect of Trade Unions", Journal of Labor Economics, vol. 3, 1984, pp. 1-26

Z. Griliches and J. Hausman, "Errors in Variables in Panel Data," Journal of Econometrics, 1986, 93-118.

Ashenfelter, Orley and Alan Krueger, "Estimates of the Economic Return to Schooling from a New Sample of Twins," American Economic Review, vol. 84, no.5, December 1994, pp. 1157-73.

Ashenfelter, Orley and Cecilia Rouse "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins," Quarterly Journal of Economics 113, February 1998, 253-284.

Bound, John and Gary Solon "Double Trouble: On the Value of Twins-based Estimation of the Return to Schooling" Economics of Education Review 18, 1999, 169-182.

Nickell, Stephen "Biases in Dynamic Models with Fixed Effects," Econometrica 49, 1981, 1417-1426.

3 Appendix

Suppose the population regression function is given by (10), so that selection into treatment depends on lagged outcomes. We estimate the fixed effects model (3) instead. Ignoring other covariates, and using the first difference estimator we get

$$\begin{aligned} \widehat{\beta} &= \frac{cov(\Delta \mathrm{Y}_{it}, \Delta \mathrm{D}_{it})}{var(\Delta \mathrm{D}_{it})} \\ \text{plim}\widehat{\beta} &= \beta + \frac{\theta cov(\Delta \mathrm{Y}_{it-h}, \Delta \mathrm{D}_{it})}{var(\Delta \mathrm{D}_{it})}. \end{aligned}$$

The second term will not be equal to zero because selection (and hence D_{it}) depends on Y_{it-h}. For example, if D_{it} is training, Y_{it} is earnings, and individuals with low earnings Y_{it-h} receive training, then $cov(\Delta \mathrm{Y}_{it-h}, \Delta \mathrm{D}_{it})$ is negative if the earnings process exhibits mean reversion (so that low earnings in period $t-h$ are likely to be associated with earnings declines just before that).

Now consider the case where the population regression function is given by the fixed effects regression (3) but we erroneously estimate the lagged outcome model (10). To see this, consider the bivariate regression of Y_{it} on

Y_{it-h} and D_{it} ignoring λ_t and X_{it}. Then we have

$$\widehat{\beta} = \frac{var(\mathrm{Y}_{it-h})cov(y_{it}, \mathrm{D}_{it}) - cov(\mathrm{Y}_{it-h}, \mathrm{D}_{it})cov(\mathrm{Y}_{it}, \mathrm{Y}_{it-h})}{var(\mathrm{D}_{it})var(\mathrm{Y}_{it-h}) - cov(\mathrm{Y}_{it-h}, \mathrm{D}_{it})^2}.$$

If $\mathrm{D}_{it-h} = 0$ and $E(\mathrm{D}_{it}) = p$, then we have $var(\mathrm{D}_{it}) = p(1-p)$ and hence

$$\begin{aligned}
\text{plim } \widehat{\beta} &= \frac{(\sigma_\alpha^2 + \sigma_\varepsilon^2)\left[cov(\alpha_i, \mathrm{D}_{it}) + \beta p(1-p)\right] - cov(\alpha_i, \mathrm{D}_{it})\sigma_\alpha^2}{p(1-p)(\sigma_\alpha^2 + \sigma_\varepsilon^2) - cov(\alpha_i, \mathrm{D}_{it})^2} \\
&= \frac{\beta p(1-p)(\sigma_\alpha^2 + \sigma_\varepsilon^2) + cov(\alpha_i, \mathrm{D}_{it})\sigma_\alpha^2}{p(1-p)(\sigma_\alpha^2 + \sigma_\varepsilon^2) - cov(\alpha_i, \mathrm{D}_{it})^2}.
\end{aligned}$$

If $cov(\alpha_i, \mathrm{D}_{it}) > 0$ then plim $\widehat{\beta} > \beta$ and we will be overestimating the treatment effect.

Card and Krueger 1994: Table 3

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

	Stores by state			Stores in New Jersey[a]			Differences within NJ[b]	
Variable	PA (i)	NJ (ii)	Difference, NJ − PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	−2.69 (1.37)	−2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	−2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores[c]	−2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	−2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	−2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	−2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

[a]Stores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ($N = 101$), is between \$4.26 and \$4.99 per hour ($N = 140$), or is \$5.00 per hour or higher ($N = 73$).

[b]Difference in employment between low-wage (\$4.25 per hour) and high-wage (≥ \$5.00 per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

[c]Subset of stores with available employment data in wave 1 and wave 2.

[d]In this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the subset of stores with available employment data in wave 1 and wave 2.

Card and Krueger 2000: Figure 2

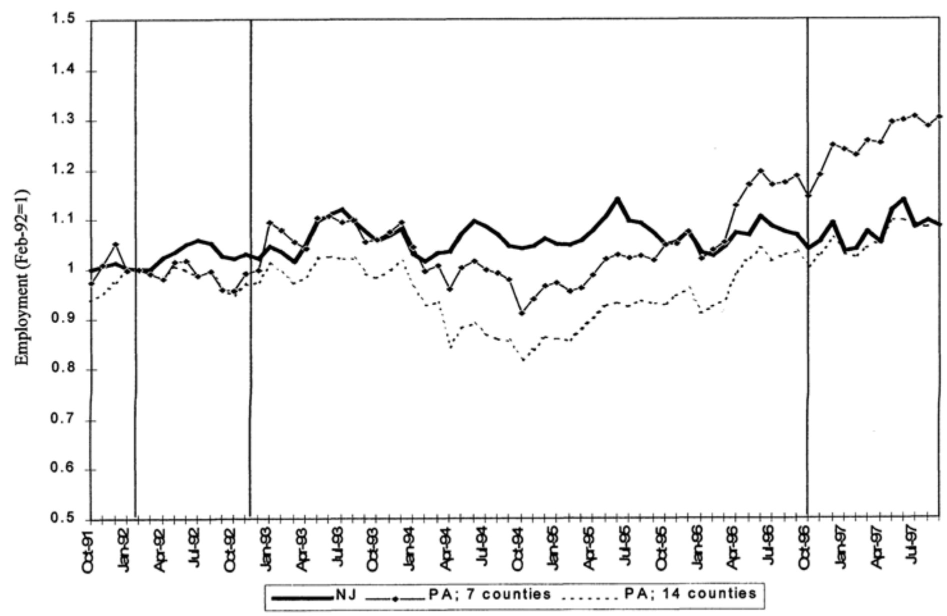
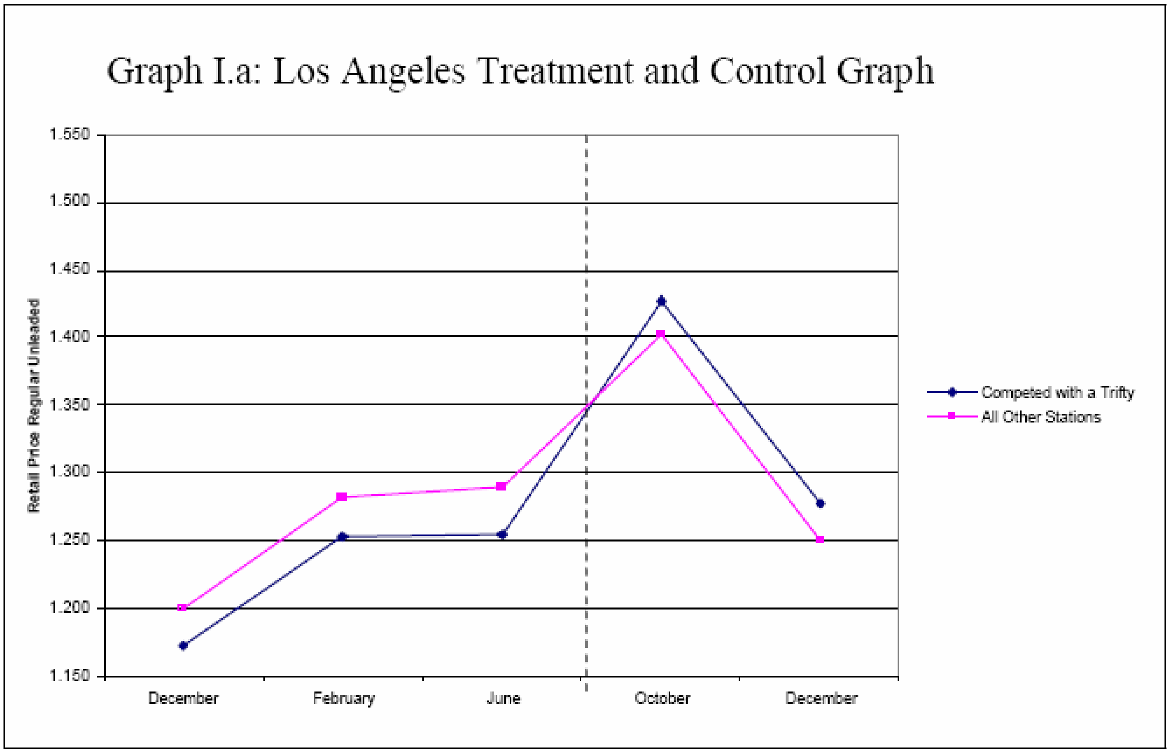


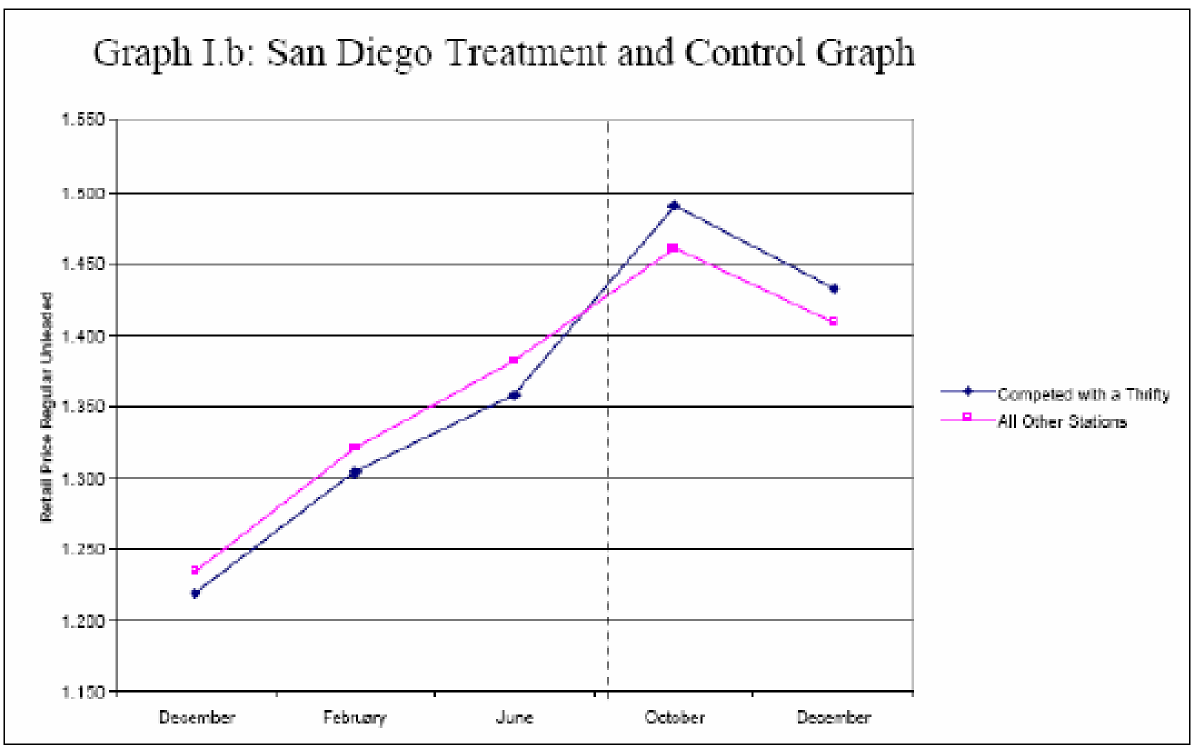
FIGURE 2. EMPLOYMENT IN NEW JERSEY AND PENNSYLVANIA FAST-FOOD RESTAURANTS, OCTOBER 1991 TO SEPTEMBER 1997

Note: Vertical lines indicate dates of original Card-Krueger survey and the October 1996 federal minimum-wage increase.
Source: Authors' calculations based on BLS ES-202 data.

Hastings 2004: Figure 1a



Hastings 2004: Figure 1b



Card 1992: Figure 4

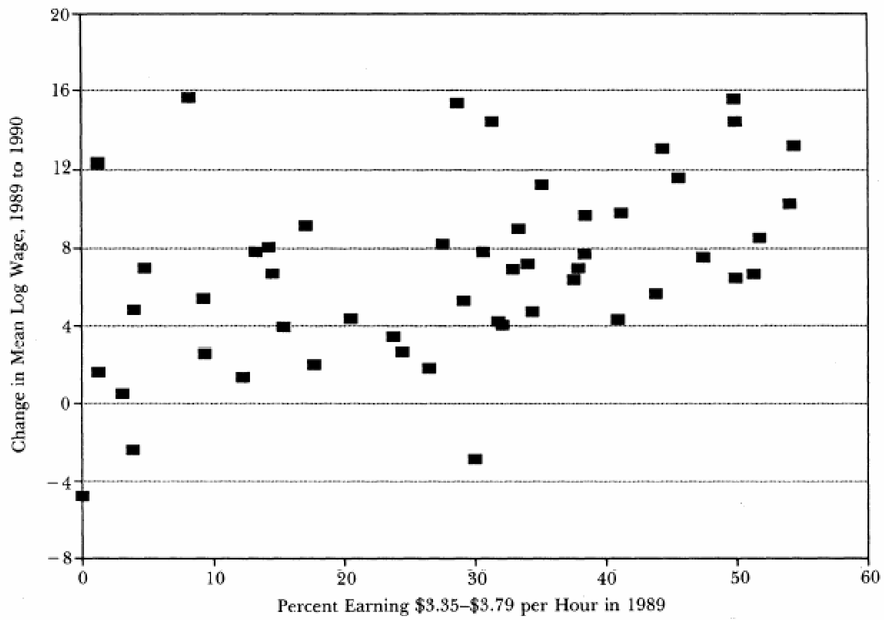


Figure 4. Change in Mean Log Wage of Teenage Workers Versus Percent Earning $3.35–$3.79 per Hour in 1989.

Card 1992: Figure 5

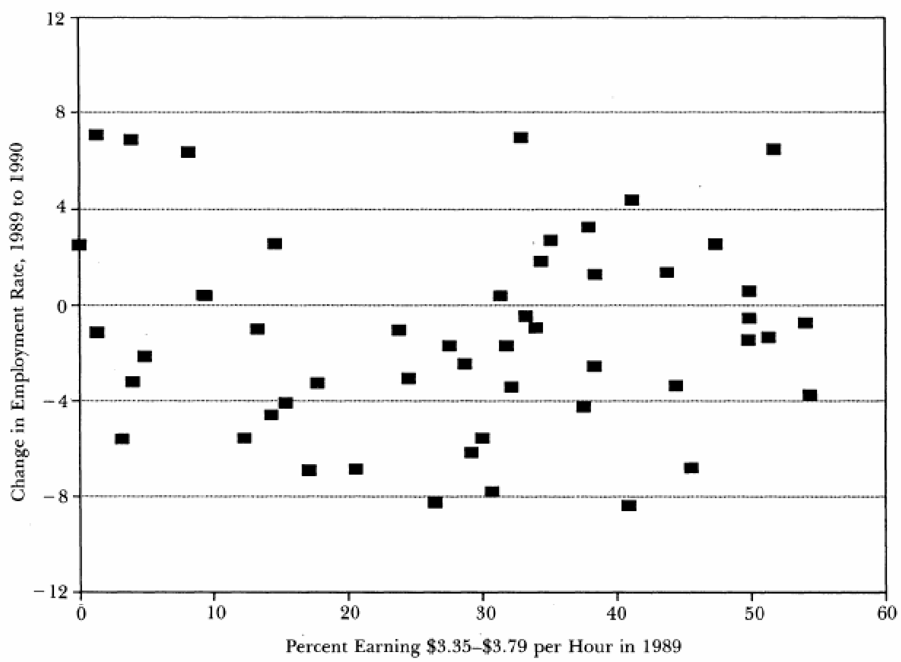


Figure 5. Change in Teenage Employment Rates Versus Percent Earning $3.35–$3.79 per Hour in 1989.

Card 1992: Table 3

Table 3. Estimated Regression Equations for State-Average Changes in Wages and Employment Rates of Teenagers, 1989–1990.
(Estimated Standard Errors in Parentheses)

	Equations for Change in Mean Log Wage:			*Equations for Change in Teen Employment-Population Ratio:*					
Explanatory Variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*
1. Fraction of Affected Teens	0.15 (0.03)	0.14 (0.04)	0.15 (0.04)	0.02 (0.03)	−0.01 (0.03)	0.01 (0.04)	—	—	—
2. Change in Overall Emp./Pop. Ratio	—	0.46 (0.60)	—	—	1.24 (0.60)	—	—	1.27 (0.66)	—
3. Change in Overall Unemployment Rate	—	—	−0.24 (0.92)	—	—	−0.16 (0.95)	—	—	−0.13 (0.98)
4. Change in Mean Log Teenage Wage[a]	—	—	—	—	—	—	0.12 (0.22)	−0.06 (0.24)	0.10 (0.30)
5. R-squared	0.30	0.31	0.30	0.01	0.09	0.01	0.01	0.09	0.01

Notes: Estimated on a sample of 51 state observations. Regressions are weighted by average CPS extract sizes for teenage workers in each state. All regressions include an unrestricted constant. The mean and standard deviation of the dependent variable in columns 1–3 are 0.0571 and 0.0417; the mean and standard deviation of the dependent variable in columns 4–9 are −0.0225 and 0.0361.

[a] In columns 7–9, the change in mean log is instrumented by the fraction of teenage workers earning $3.35–3.79 in 1989.

Autor 2003: Figure 3

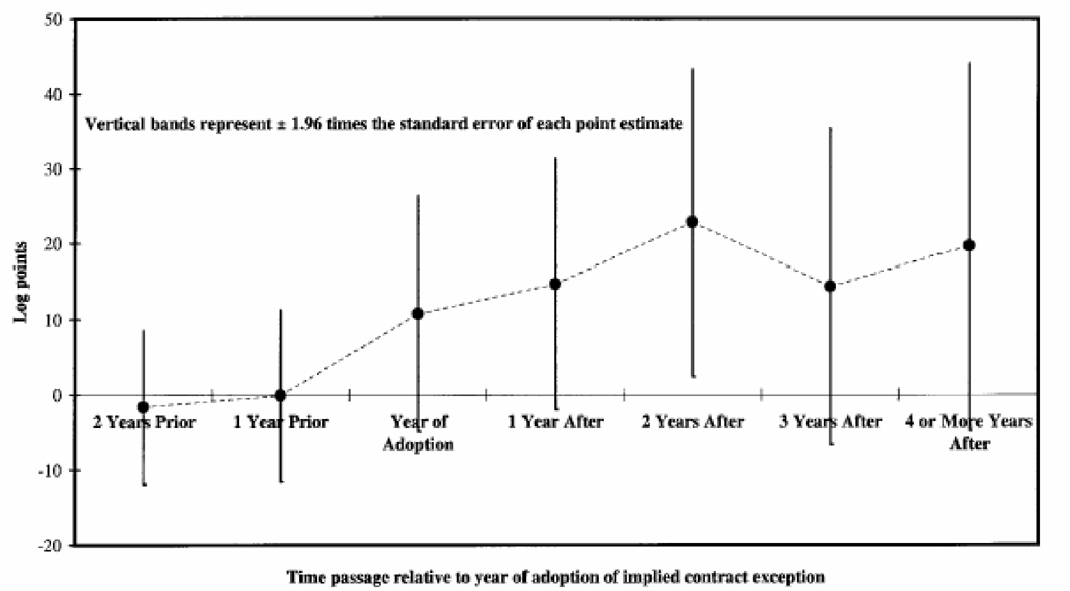


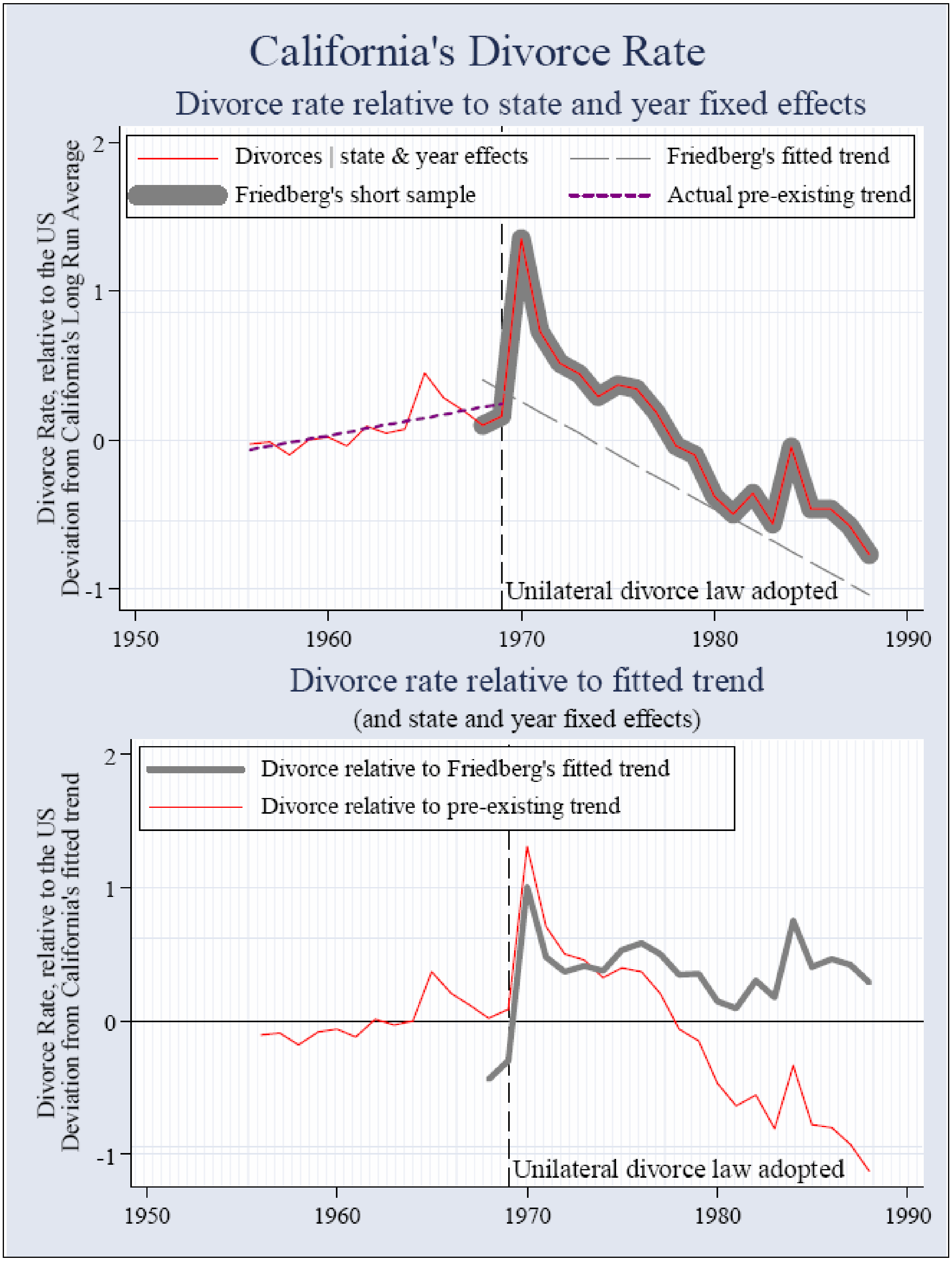
FIG. 3.—Estimated impact of implied contract exception on log state temporary help supply industry employment for years before, during, and after adoption, 1979–95.

Besley & Burgess 2004: Table 4

TABLE IV
LABOR REGULATION AND MANUFACTURING PERFORMANCE IN INDIA: 1958–1992

	(1)	(2)	(3)	(4)	(5)	(6)
	Log registered manufacturing output per capita	Log registered manufacturing output per capita	Log registered manufacturing output per capita	Log registered manufacturing output per capita	Log registered manufacturing output per capita	Log unregistered manufacturing output per capita
Method	OLS	OLS	OLS	OLS [state time trends]	OLS [no West Bengal]	OLS [no West Bengal]
Labor regulation	−0.186***	−0.185***	−0.104***	0.0002	−0.105***	0.077**
[t − 1]	(2.90)	(3.65)	(2.67)	(0.01)	(2.59)	(2.25)
Log development		0.240*	0.184	0.241**	0.208	0.492***
expenditure per capita		(1.88)	(1.55)	(2.28)	(1.69)*	(3.39)
Log installed electricity		0.089	0.082	0.023	0.053	−0.070
capacity per capita		(1.47)	(1.51)	(0.69)	(1.21)	(1.11)
Log state population		0.720	0.310	−1.419	0.629	−3.724***
		(0.75)	(0.26)	(0.61)	(0.53)	(3.18)
Congress majority			−0.0009	0.020**	−0.002	0.017
			(0.09)	(2.08)	(0.27)	(0.95)

Wolfers 2003: Figure 5



Davis & Weinstein 2002: Figure 2

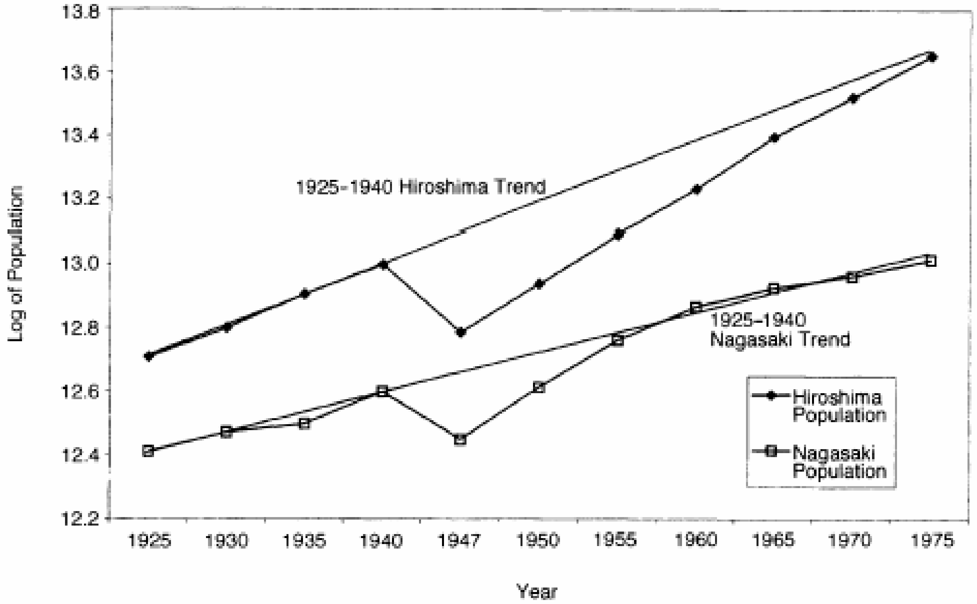


FIGURE 2. POPULATION GROWTH

Abadie & Garbeazabal 2003: Figure 1

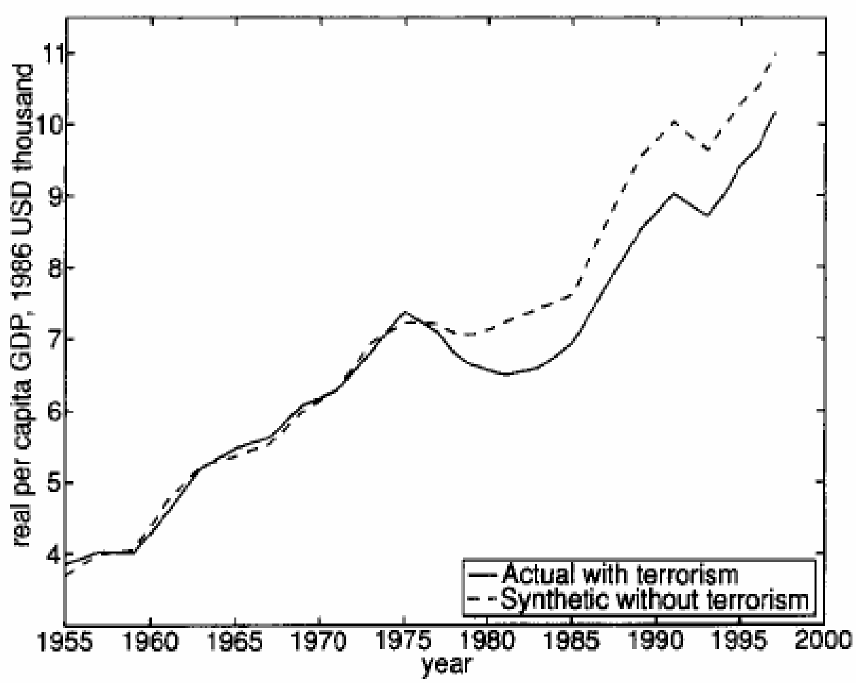


FIGURE 1. PER CAPITA GDP FOR THE BASQUE COUNTRY

Freeman 1984: Table 6

Table 6
Log Wages, Changes in Log Wages Associated with Changing Union Status, and Estimated Union Effects

Group and Survey	Log Wage Before	Log Wage After	Δ	Group	Estimated Union Effects
A. May CPS, 1974–75:					
NN	1.24	1.34	.10	*NU−NN*	.09
NU	1.28	1.47	.19	*UU−UN*	.08
UU	1.58	1.67	.09	(*NU−UN*)/2	.09
UN	1.46	1.47	.01	Cross-section	.19
B. National Longitudinal Survey of Young Men, 1970–78:					
NN	.97	1.84	.87	*NU−NN*	.12
NU	.94	1.93	.99	*UU−UN*	.09
UU	1.34	2.05	.71	(*NU−UN*)/2	.19
UN	1.22	1.84	.62	Cross-section	.28
C. Michigan PSID, 1970–79:					
NN	.95	1.61	.67	*NU−NN*	.08
NU	1.06	1.81	.75	*UU−UN*	.26
UU	1.29	2.02	.73	(*NU−UN*)/2	.14
UN	1.16	1.63	.47	Cross-section	.23
D. QES, 1973–77:					
NN	1.38	1.85	.48	*NU−NN*	.19
NU	1.24	1.91	.67	*UU−UN*	.11
UU	1.55	2.00	.45	(*NU−UN*)/2	.16
UN	1.35	1.70	.34	Cross-section	.14

SOURCE.—Calculated from the surveys. Cross-section estimates based on multivariate regression model with standard set of controls for demographic and human capital variables.

Freeman 1984: Table 1

Table 1
Example of Measurement Error Effect
A. Cross-Section Data Set

Observed	True	Number
U	*U*	23
U	*N*	2
N	*U*	2
N	*N*	73

B. Longitudinal Data Set

	Observed	Consisting of True	With Observed Means of 1	With Observed Means of 2
UU	13	13 *UU*	1.30	1.30
UN	12	9 *UN*, 1 *UU*, 2 *NN*	1.25	1.03
NU	12	9 *NU*, 1 *UU*, 2 *NN*	1.03	1.25
NN	63	61 *NN*, 1 *UN*, 1 *NU*	1.004	1.004