EMPIRICAL BAYESIAN ESTIMATION OF TREATMENT EFFECTS

CHRISTOPHER P. ADAMS,
FEDERAL TRADE COMMISSION

Abstract. This paper presents an alternative approach to difference in difference estimation of treatment effects when the number of treated units are small and the number of control units are large. Information from all of the cross sectional units are used to determine a consistent estimate of the distribution of the treatment effect. The empirical Bayesian estimator appropriately accounts for sampling variation. In particular, the proposed method tends to "shrink" the estimate of the treatment effect. The approach is illustrated with an analysis of the impact of German reunification on per capita GDP growth rates. Using data from the World Bank on GDP per capita, the proposed method suggests that there no effect on growth rates from reunification, where the difference in difference method finds a decrease in the growth rate of -0.014.

The views expressed in this article are those of the author and do not necessarily reflect those of the Federal Trade Commission. I'm grateful for continued discussions about this problem with Emek Basker, Dan Greenfield, Dan Hosken, Nick Kreisle, Joris Pinske, Jeremy Sandford, Chris Taylor and Nathan Wilson. Thanks also to Chuck Manski. All errors are my own.

1. Introduction

Difference in difference is a standard method for estimating the effects of policy changes using panel data sets (Abadie et al., 2010; Simpson and Taylor, 2008; Hosken et al., 2011; Silvia and Taylor, 2013; Kreisle, 2015). Unfortunately, difference in difference has a number of drawbacks. Of particular concern is that the estimated magnitude of the treatment effect may be too large because sampling variation is not accounted for appropriately. This paper presents a method that uses information across the large number of cross sectional units to provide a consistent estimate of the distribution of the treatment effect. Using data on country level GDP per capita, the method finds no difference in growth rates for Germany before and after reunification, with the 90% confidence interval spanning -0.020 to 0.029. Using the standard difference in difference approach, the treatment effect of reunification is -0.014, with the approximate 90% confidence interval spanning -0.072 to 0.044.

Difference in difference is an analog estimate of the treatment effect, and as such may be preferred by some researchers (Manski, 1988). Under linearity assumptions, difference in difference is also an unbiased estimate of the treatment effect. Over a large enough number of "experiments," the average estimate will equal the true treatment effect. However, it is not clear that the difference in difference estimate is close to the true treatment effect in any particular experiment. We say that an estimate is consistent if for a large number of observations, the estimator is equal to the true value. If we are willing to assume that the actual number of observations is "large" then we think of the estimator is "approximately" equal to the true value (Goldberger, 1991). In panel data sets such as the one used here, the number of observations can be measured in two dimensions, number of cross sectional units and number of time periods. In the case considered here, the difference in difference estimator is consistent if *both* are large (Cameron and Trivedi, 2005). Unfortunately, we have only a small number of time periods. We cannot appeal to asymptotic results to reassure ourselves that our actual difference in difference estimate is close to the true value. In fact, the difference in difference estimate may tend to be larger in magnitude than the true value.

To see why this would be the case, consider an example like the one analyzed below. In the example we have a large number of countries and each country is characterized by an average growth rate. This average growth rate may be determined by various characteristics of the country like infrastructure, education, taxes, democratic institutions and regulations. We are interested in measuring the average growth rate in a particular country. The most obvious way to do this is to use annual national accounts data and calculate a sample average. While we may not know the average growth in any particular country, assume we know the distribution of actual averages across the countries. Assume that we also observe the sample average from the data for a large number of countries. These two distributions are not the same. The sample average is an unbiased estimate of the actual average, but the distribution of sample averages across all the countries is not equal to the distribution of actual averages across the countries (Efron and Hastie, 2016). Figure 1 is from the numerical example discussed below. The figure illustrates that the actual distribution of averages is quite different from the distribution of sample averages. In particular the distribution of actual averages has much thinner tails than the distribution of sample averages.

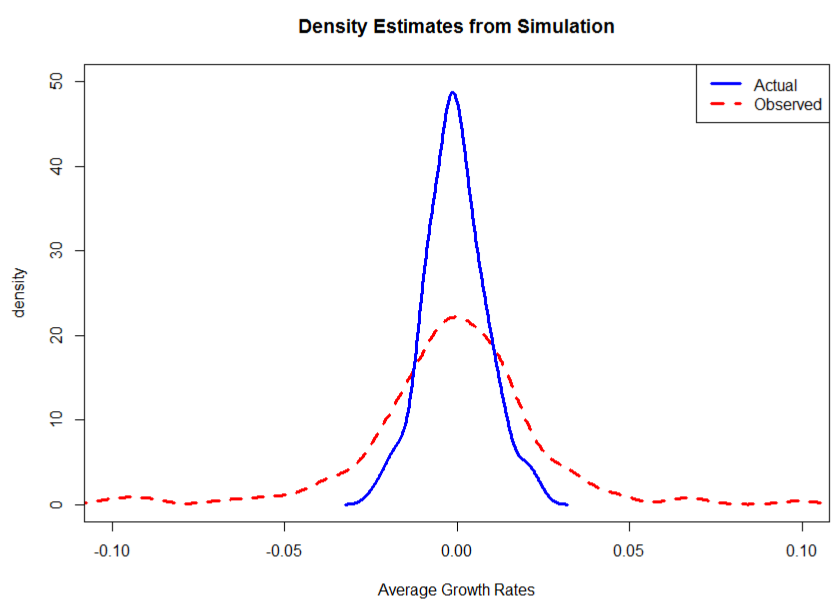

Figure 1. Plot of density functions associated with the simulation of 190 countries with growth rates from 4 time periods. The blue density function represents the true distribution of country types (country average growth rate), the red density function is the distribution of sample averages for the countries.

In this paper we are interested in how a change in a countries political structure, German reunification, may change the average growth rate in the country. Assume that reunification leads to a new draw from the distribution of average growth rates. The treatment effect is the difference in two draws from the distribution of average growth rates.[1] Given this assumption, our *a priori* expectation is that the treatment effect is zero. To estimate the treatment effect we compare the sample average before reunification to the sample average after reunification. This is equivalent to the difference between two draws of a the distribution of sample averages. Figure 2 presents absolute differences from a large number of differences between two random draws from the actual distribution (blue) and the distribution of sample averages (red). The absolute difference from the sample estimates will tend to be larger than the absolute difference of the actual values. Therefore, the measured treatment effect will tend to be larger in magnitude relative to the actual treatment effect.

The solution is to use a Bayesian estimator. If we know the true distribution of average growth rates then we can combine this prior with the observed samples via Bayes Rule. Through this process we get an estimate of the posterior distribution of the treatment effect. The treatment effect will tend to "shrink," as the Bayesian estimator appropriately account for the sampling variation. The problem with this approach is that we do not observe the true distribution of average growth rates across countries. The empirical Bayesian approach notes that in certain cases, such as the case analyzed here, it is possible to consistently estimate the true prior distribution. The method has two steps. In the first step, the method uses the observed distribution of sample average growth rates to estimate the true distribution of country average growth rates. The second step uses Bayes Rule, the sample data and the estimated prior, to estimate the posterior distribution over the treatment effect.

Robbins (1956) notes that the density over estimated averages is a mixture of density functions over sample averages conditional on the true type, weighted by the density over types. He notes that the solution to this mixture model provides both an estimate of the true prior over average growth rates and an estimate of the posterior distribution over each country's average growth rate. This is called a deconvolution problem. This paper shows that the deconvolution problem is solved for arbitrary approximations of the conditional density functions. Efron (2014) notes that the mixture model is uniquely determined when the conditional density functions are

[1] Assume for simplicity that growth rates are stable over time.

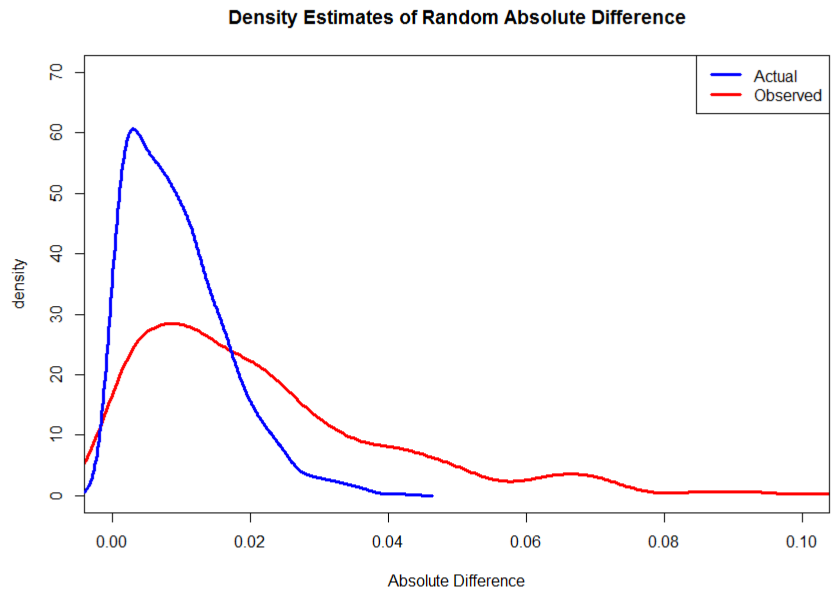

Figure 2. Plot of density functions associated with the simulation of 190 countries with prices from 4 time periods. The blue density function represents the absolute difference in random draws from the true distribution of country types (average growth rates), the red density function is the density of the absolute difference in two draws from the distribution of sample averages for the countries.

known (assuming a rank condition holds).[2] While Efron and Narasimhan (2016) notes that the conditional density function is a multinomial function when the set of prices is partitioned into a finite number of sets. That is, for any arbitrary approximation, the conditional density functions are known. Therefore we can solve the deconvolution problem and estimate the true prior and true posterior distributions. This paper uses a method for solving the deconvolution problem adapted from the iterative approach of Benaglia et al. (2009).

[2]Rao (1992) points out that there is an identification concern for binomial functions when the number of observations is small or doesn't vary across experiments.

This estimator is not unbiased and it is not an analog estimator, but it has a number of virtues.[3] The paper shows that the empirical Bayesian estimator is consistent when the number of cross sectional units gets large. To the extent that we are willing to think of this as an approximation for the actual data, the estimator is close to the true treatment effect. In addition, the estimator belongs to a class of estimators that are generically "closer" to the true treatment effect than the class of estimators that the difference in difference estimator belongs to.[4]

This paper adds to a growing literature expressing concerns with estimation and interpretation of difference in difference estimates (Bertrand et al., 2004; Abadie et al., 2010; Xu, 2015; Greenfield et al., 2015; Hosken et al., 2017). Bertrand et al. (2004) suggests that some researchers are not appropriately accounting for serial correlation in the data.[5] Abadie et al. (2010) raises two concerns. First, the estimator makes strong assumptions about how the control units are weighted to construct the counter-factual estimates. See Doudchenko and Imbens (2016) for a survey of various approaches to solving this problem. Second, the approximation error of the counter-factual estimates is not appropriately accounted for in standard reports of uncertainty. While these issues are important, this paper focuses on the problem that sampling variation tends to inflate the treatment effect estimates.

The paper proceeds as follows. Section 2 presents a model of the difference in difference estimator and the empirical Bayesian estimator. Section 3 presents an example with simulated data. Section 4 presents analysis of the impact of reunification on the growth rate of GDP per capita in Germany. Section 5 concludes.

2. Model

Consider a standard difference in difference model for panel data. Some potential outcome $Y_{it}(D_{it})$ is (potentially) observed for country $i \in \{1, ..., N\}$ in time $t \in$

[3]There is some debate about the relative virtues of empirical Bayesian estimators in the "value added" literature, which looks at evaluating teachers using standardized test data (McCaffrey et al., 2003).

[4]James and Stein show that an analog estimator, such as the difference in difference estimator, can be dominated in terms of risk. They show that an analog estimator is dominated by a Bayesian estimator when the statistician has access to 3 or more independent by related samples (Efron and Morris, 1973). Here we have access to 197 independent but related samples.

[5]Here, the paper aggregates annual data up to 5 year periods because this aggregation accounts for the serial correlation in the data.

$\{1, ..., T\}$, where D_{it} is an indicator for whether or not country i receives the treatment in period t. Here Y_{it} will be the average growth rate for country i in a 5-year period t. The treatment effect is received in country $i = 1$ in period $t > T_0$. Here the treatment is the reunification of Germany which occurred in 1990.

$$Y_{it}(D_{it}) = \alpha_i + \beta_t + \gamma_i D_{it} + \epsilon_{it} \tag{1}$$

where

$$D_{it} = \begin{cases} 1 \text{ if } i = 1 \text{ and } t > T_0 \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

The parameter α_i is the fixed effect measuring the mean of the growth rate for country i, β_t is the time-dummy which accounts for variation over time that impacts all of the countries, γ_i is the individual treatment effect for country i and ϵ_{it} is the idiosyncratic error that is assumed to be distributed iid across countries and time.

Note that prior to the treatment the outcome for the treated unit is, $Y_{1t}(0) = \alpha_1 + \beta_t + \epsilon_{1t}$. Post treatment the outcome is $Y_{1t'}(1) = \alpha_1 + \beta_{t'} + \gamma_1 + \epsilon_{1t}$. If $E(\epsilon_{it}) = 0$, then the analog estimator of the treatment effect can be derived from the following equations.

$$\hat{\alpha}_1 = \frac{1}{T_0} \sum_{t=1}^{T_0} (Y_{1t} - \hat{\beta}_t) \tag{3}$$

and

$$\hat{\gamma}_1 = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} (Y_{1t} - \hat{\beta}_t - \hat{\alpha}_1) \tag{4}$$

and

$$\hat{\beta}_t = \frac{1}{N - 1} \sum_{i=2}^{N} (Y_{it} - \hat{\alpha}_i) \tag{5}$$

If $E(\alpha_i) = 0$ then $\hat{\beta}_t$ is an unbiased and consistent estimator of β_t. It is straight forward to see that $\hat{\gamma}_1$ is unbiased but not a consistent estimator as the number of cross sectional units (N) goes to infinity.

Here, we will estimate a slightly more general model, allowing for the "error" term to have a distribution that varies with both the country and the treatment.

$$Y_{it}(D_{it}) = \alpha_i + \beta_t + \gamma_i D_{it} + \epsilon_{iDt} \tag{6}$$

Let $\tilde{Y}_{it}(D_{it}) = Y_{it}(D_{it}) - \beta_t$, where $\tilde{Y}_{it}(D_{it}) \sim \theta_{iD}$ and θ_{iD} denotes i's type for potential outcome D. While we cannot observe all the potential outcomes and thus all types, if we assume that all types are drawn from the same distribution we can observe a large random sample of types. That is, the treatment changes the country's type, but all types are drawn from the same distribution irrespective of treatment.[6]

Given this assumption, let $g(\theta_{iD})$ denote the prior density over types. Let $f(\hat{\theta}_{iD})$ denote the density over the observed sample $\{\tilde{Y}_{it}(D_{it})\}_{t \in \{1,...,T\}}$. Lastly, let $h(\hat{\theta}_{iD}|\theta_{iD})$ denote the conditional density of the observed distribution given the true type (Efron and Narasimhan, 2016).

$$f(\hat{\theta}_{iD}) = \int_{\theta} h(\hat{\theta}_{iD}|\theta) g(\theta) \tag{7}$$

Robbins (1956) notes that if the left-hand side of Equation (7) is observed and if the equation can be solved, we can determine the prior distribution over the true types. Unfortunately, without further information or restrictions Equation (7) cannot be solved uniquely (Adams, 2016). However, Efron (2014) points out that if $h(.|.)$ is known and a rank condition holds, then the equation can be solved uniquely. If the set of outcomes is discrete, then $h(.|.)$ is known. It is a multinomial function (Efron and Hastie, 2016).

Let $\tilde{Y}_{it}(D_{it}) \in [\underline{Y}, \bar{Y}] \subset \Re$. Partitioning this set into K subsets $\sum_{k=1}^{K} [Y_k, Y_{k+1}] = [\underline{Y}, \bar{Y}]$. Given this, for any θ_{iD}, the function $h(.|\theta_{iD})$ is a multinomial function,

$$h(\hat{\theta}_{iD}|\theta_{iD}) = \frac{T_{iD}!}{\Pi_{k=1}^{K} (\hat{p}_{kiD} T_{iD})!} \Pi_{k=1}^{K} p_{kiD}^{\hat{p}_{kiD} T_{iD}} \tag{8}$$

where $p_{kiD} = \Pr(\tilde{Y}_{it}(D) \in [Y_k, Y_{k+1}])$, $\hat{p}_{kiD} = \frac{1}{T_{iD}} \sum_{t=1}^{T_{iD}} 1|\tilde{Y}_{it}(D) \in [Y_k, Y_{k+1}]|$, $\theta_{iD} = \{p_{1iD}, p_{2iD}, ..., p_{KiD}\}$ and $\hat{\theta}_{iD} = \{\hat{p}_{1iD}, \hat{p}_{2iD}, ..., \hat{p}_{KiD}\}$ and T_{iD} is the number of observations (time periods). If the set $[0, 1]$ is partitioned into L sets, then the problem can be approximated with L^K types and written in matrix form (Efron and Narasimhan, 2016).

$$\mathbf{f} = \mathbf{H}\mathbf{g} \tag{9}$$

[6]This restriction could be relaxed if both treatment groups were large and there is no selection into the observed treatment.

where $\mathbf{f}$ is $L^K \times 1$ vector representing the probability distribution over the observed sample distributions, $\mathbf{H}$ is $L^K \times L^K$ square matrix of mapping from the true distribution to the observed sample frequencies, and $\mathbf{g}$ is a $L^K \times 1$ vector representing the prior distribution over types.

Efron (2014) notes that if $\mathbf{H}$ is full-rank and known then

$$\mathbf{g} = \mathbf{H}^{-1}\mathbf{f} \tag{10}$$

and the prior distribution can be determined from observed distribution of types. Rao (1992) notes that the binomial model is only identified up to a function of the number of possibly observed frequencies. Here the number of possibly observed frequencies includes all the frequencies for samples of size 1 to 9.[7] Note that as sample size increases, the matrix $\mathbf{H}$ will converge to the identity matrix which is full-rank.

Theorem 1. *Given an approximation denoted by L and K (Equation 9), then if $\mathbf{H}$ is full-rank,* $\lim_{N \to \infty} \hat{g}(\theta|\hat{\theta}) = g(\theta|\hat{\theta})$

Proof. Robbins (1956) shows that $\hat{f}(\hat{\theta})$ is consistently estimated and so by Bayes Rule and Efron (2014) the result holds. Note that identification may only up to set of moments (Rao, 1992). □

While it is not possible to consistently estimate the country's type using a standard analog estimator, we can use the empirical Bayesian approach to consistently estimate the posterior distribution over each country's type.

3. Example with Simulated Data

Consider a simulated set of 190 countries, where the average growth rate in each country i is drawn from a normal distribution with mean μ_i and standard deviation σ_i. The distribution of country growth rate means is drawn from a normal distribution with mean zero and standard deviation equal to the standard deviation of the sample means in the data presented below divided by a factor (5 in the analysis below). Note that while the data in this example is generated with a normal distribution, the estimation procedure makes no assumption on the characteristics of the prior distribution. The procedure estimates the prior distribution non-parametrically. Each simulated country's standard deviation is drawn at random from the actual

[7]This issue is somewhat mitigated in the multinomial case by choosing a larger number of partitions.

distribution of sample standard deviations. Given this true distribution of country-types, simulated growth rates are drawn for 4 time periods. Note that the country-type distributions are fixed over time in the simulations.

The problem is approximated by assuming that all growth rates lie between the smallest and largest observed growth rates and partitioned that into 20 subsets ($K = 20$). A set of 100,000 distributions are drawn at random to be the approximate set of possible distribution types. A multinomial function determines the probability of each observed sample conditional on each of the 100,000 possible distributions. The model is solved using an iterative approach based on Benaglia et al. (2009).

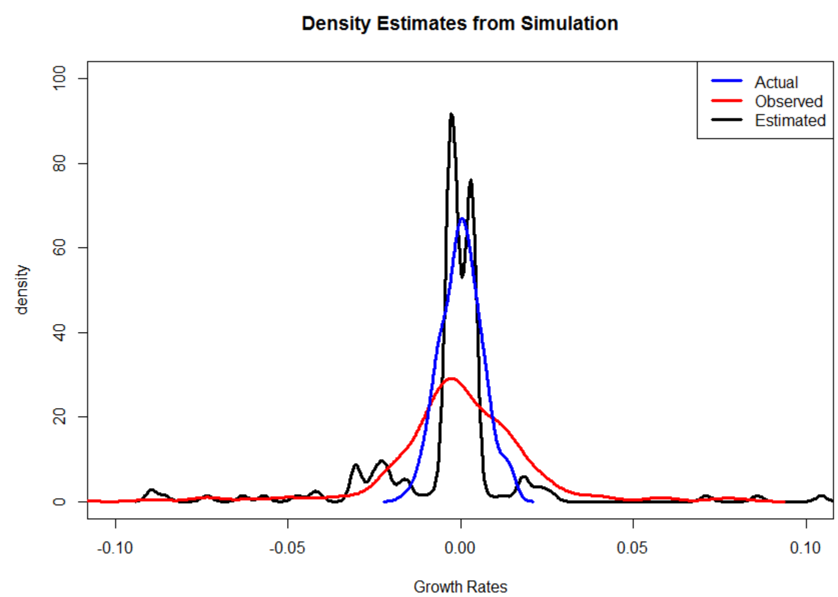

FIGURE 3. Plot of three density functions associated with the simulation of 190 countries with growth rates from 4 time periods. The blue density function represents the true distribution of country types (country average growth rates), the red density function is the distribution of sample averages for the countries and the black density function represents the distribution of posterior of country average growth rates from the empirical Bayesian estimator.

Figure 3 presents the density estimates from the simulation where the true distribution of average growth rates is distributed normal with mean of zero and standard deviation equal to the standard deviation of the sample averages from the data below divided by a factor of 5. The figure illustrates the issue that the observed distribution of sample averages has thicker tails than the actual distribution. The implication is that a country with an observed average that sits in the tails, that average is more likely to be due to sampling variance than be equal to the true country mean growth rate. The empirical Bayesian approach shrinks the estimates of the country mean growth rates down toward their true values.

The implication for treatment effect estimation is that if the observed treatment effect is the difference between two draws from the red density function, then the estimate will be tend to be too large in magnitude. One or both of the country averages will be in the tails of the distribution. For these estimates it is more likely that the observed average is due to sampling variance than a true estimate of the country's type.

4. Reunification of Germany

On November 9, 1989, a spokesman mistakenly announced that citizens of the German Democratic Republic would be allowed unrestricted travel to West Germany. This led to a series of events that ended with the reunification of Germany. We are interested in whether this major change in the political institutions and population had an impact on economic growth.

The base data set is annual GDP per capita in constant 2010 US dollars, from the World Bank's Data Bank. The annual growth rate is calculated as the difference in log GDP per capita between years. In order to account for serial correlation in the data, the data is aggregated up into 5 year periods from 1961 to 2016.[8] The pre-treatment is the 6 half-decades before 1990. There are 217 countries in this data set, although not all countries have data for all periods. The final data set includes growth rate estimates for 197 countries. The treated country is Germany. Abadie et al. (2010) are careful to estimate the impact on West Germany, rather than Germany. Here we use the definition of Germany used by the World Bank.

The estimated treatment effect from the difference in difference model is $\hat{\gamma} = -0.014$ with a standard error of 0.030. Using the Abadie et al. (2010) data, $\hat{\gamma} = -0.010$ with a standard error of 0.012. Although neither estimate is statistically different from zero at standard levels, there is still concern that the point estimate is too large

[8] Data that is aggregated up to shorter time periods shows signs of serial correlation.

in magnitude. The concern is that the estimate is based on the difference between average growth rates that lie in the tail of the distribution. Analysis presented above suggests that an estimate in the tail is more likely to be due to sampling variance than be a true measure of country's growth rate.

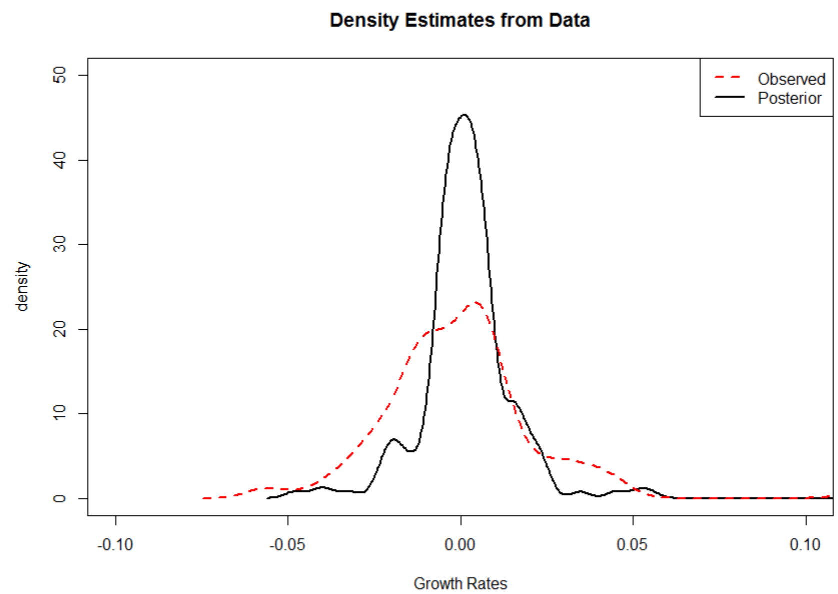

FIGURE 4. Density of country growth rates net of time dummies (and normalized to have mean zero). The dashed red density is the observed distribution of country sample average growth rates and the solid black density function is the estimated density of the mean of the country posterior distribution.

Figure 4 presents the actual observed distribution of country types (country sample average growth rates net of time dummies) and the empirical Bayesian estimate of the prior distribution of country types. As in the simulated data, the figure shows that the distribution of sample averages tends to have thicker tails than the true distribution. In general, this implies the estimated treatment effect is expected to be smaller in magnitude than treatment effect estimated with the standard difference in difference.

Figure 5 presents the posterior density estimator of the treatment effect. The solid-black line is the posterior density using the World Bank definition of Germany

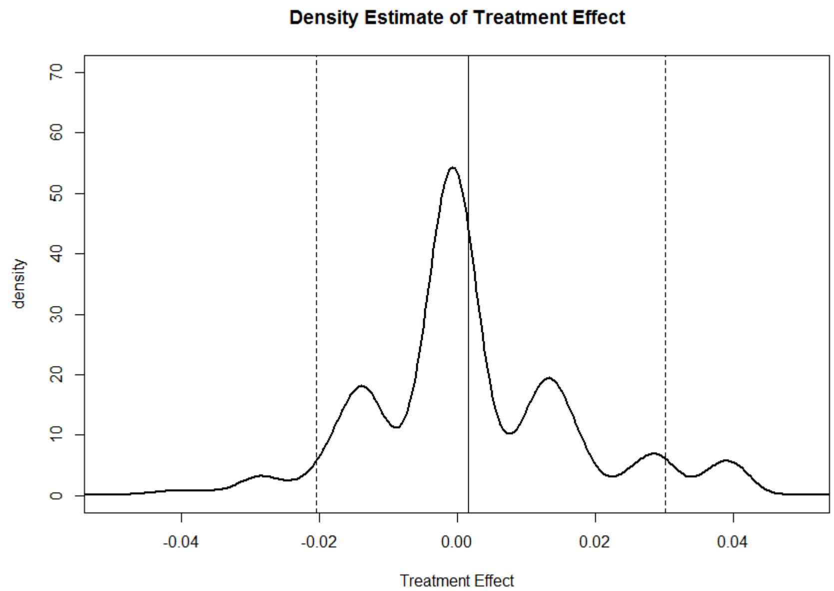

FIGURE 5. Posterior density of the treatment effect. The solid vertical line represents the mean of the posterior of the treatment effect and the dashed vertical lines represent the 5th and 9th percentiles of the posterior distribution. These are 0.001, -0.020 and 0.029 respectively.,

pre and post 1990. The solid vertical line is the mean of the posterior distribution which is very close to 0. The 90% confidence interval from the posterior distribution of the treatment effect is much tighter around zero than the classical confidence interval, it goes from -0.020 to 0.029. The 90% confidence interval for the classical estimate runs from -0.072 to 0.044.

4.1. **Serial Correlation.** Bertrand et al. (2004) raises the concern that there is often a substantial amount of serial correlation in the data. This is of particular concern here because the proposed estimator relies heavily on the assumption that the observations are drawn independently conditional on the country's type.

Figure 6 shows the correlation across adjacent time periods within the same country. The figure and regression results (not shown) suggest that the serial correlation is not present in the data.

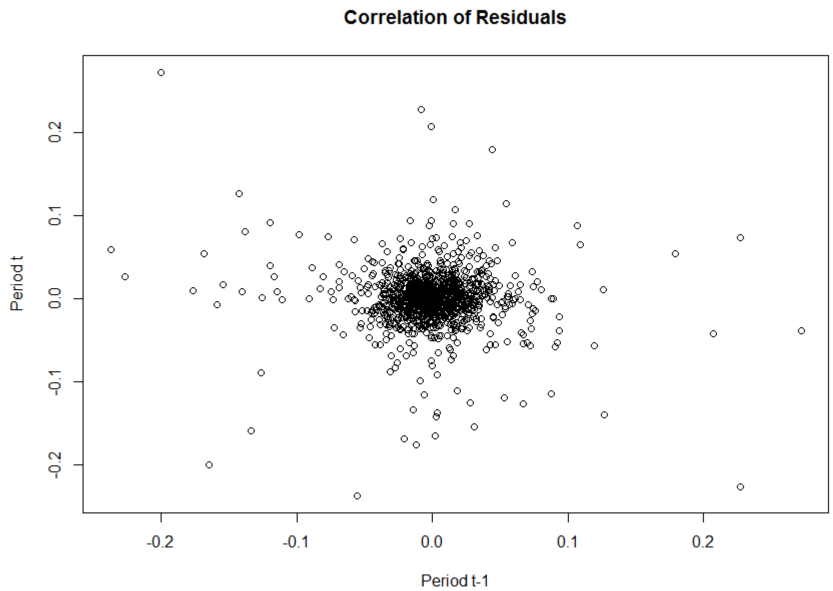

FIGURE 6. Plot of country level residuals across time periods.

5. Conclusion

This paper presents an alternative approach to estimating difference in difference treatment effects with panel data. The empirical Bayesian estimator appropriately accounts for systematic bias due to sampling variation. When the time periods are small, the distribution of fixed effects will have thicker tails than the true distribution of fixed effects. As the treatment effect is measured as the difference between fixed effects estimates, its magnitude will tend to be too large. The paper illustrates the problem with simulation data. It compares the approach to the standard approach with data on per capita GDP growth rates before and after the reunification of Germany in 1990. The estimated treatment effect is much smaller in magnitude than the difference in difference estimate.

References

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.

Christopher P. Adams. Finite mixture models with one exclusion restriction. *The Econometrics Journal*, 19:150–165, 2016.

Tatiana Benaglia, Didier Chauveau, and David R Hunter. An em-like algorithm for semi- and non-parametric estimation in multivariate mixtures. *Journal of Computational and Graphical Statistics*, 18(2):505–526, 2009.

M. Bertrand, E. Duflo, and S. Mullainathan. How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, 119, 2004.

A. Colin Cameron and Pravin K. Trivedi. *Microconometrics: Methods and Applications*. Cambridge University Press, 2005.

Nikolay Doudchenko and Guido W. Imbens. Balancing, Regression, Difference-in-Difference and Synthetic Control Methods: A synthesis. NBER Working Paper 22791, October 2016.

Bradley Efron. Two modeling strategies for empirical bayes estimation. *Statistical Science*, 29:285–301, 2014.

Bradley Efron and Trevor Hastie. *Computer Age Statistical Inference: Algorithms, evidence, and data science*. Cambridge University Press, Boston MA, 2016.

Bradley Efron and Carl Morris. Stein's estimation rule and its competitors–an empirical bayes approach. *Journal of the American Statistical Association*, 68(341): 117–130, 1973.

Bradley Efron and Balasubramanian Narasimhan. A G-Modeling Program for Deconvolution and Empirical Bayes Estimation. Standford, 2016.

Arthur Goldberger. *A Course in Econometrics*. Harvard University Press, 1991.

Daniel Greenfield, Nicholas Kreisle, and Mark Williams. Simulating a homogeneous product merger: A case study on model fit and performance. FTC Working Paper No. 327, October 2015.

Daniel Hosken, Louis Silvia, and Christopher Taylor. Does concentration matter? measurement of petroleum merger price effects. *American Economic Review: Papers and proceedings*, 101(3):45–50, 2011.

Daniel S. Hosken, Luke M. Olson, and Loren K. Smith. Do retail mergers affect competition? evidence from grocery retailing. *Journal of Economics and Management Strategy*, August 2017. URL `http://dx.doi.org/10.1111/jems.12218`.

Nicholas Kreisle. Merger policy at the margin: Western refinings acquisition of giant industries. *Review of Industrial Organization*, 47:71–89, 2015.

Charlies Manski. *Analog Estimation Methods in Econometrics*. Chapman and Hall, 1988.

Daniel F. McCaffrey, J. R. Lockwood, Daniel M. Koretz, and Laura S. Hamilton. Evaluating value-added models for teacher accountability. RAND, 2003.

B. L. S. Prakasa Rao. *Identifiability in Stochastic Models: Characterization of probability distributions*. Academic Press, 1992.

Herbert Robbins. The empirical bayes approach to statistical decision problems. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 157–163, 1956. University of California Press.

Louis Silvia and Christopher Taylor. Petroleum mergers and competition in the northeast united states. *International Journal of the Economics of Business*, 20(1): 97–124, 2013.

John Simpson and Christopher Taylor. Do gasoline mergers affect consumer prices? the marathon ashland petroleum and ultramar diamon shamrock transaction. *The Journal of Law & Economics*, 51(1):135–152, 2008.

Yiqing Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 2015. Forthcoming.

E-mail address: cadams@ftc.gov