Are two-way fixed-effect difference-in-differences estimates blowing smoke? A cautionary tale from state-level bank branching deregulation

Anthony Zdrojewski
Rice University
anthony.m.zdrojewski@rice.edu

Alexander W. Butler
Rice University
alex.butler@rice.edu

November 15, 2021

We illustrate the sensitivity of two-way fixed effects difference-in-differences estimates to innocuous changes in data structure. Using the staggered rollout of state-level bank branching deregulations, three outcome variables are brought to bear on the interventions: personal income growth (a replication), house prices (new to the literature), and per capita cigarette purchases (a falsification test). Estimates are sensitive to panel length, and the data structure creates the false impression of a causal effect of the interventions on all three outcome variables. We contend that any two-way fixed effects regression using this set of interventions is at risk of generating spurious results.

Keywords: banking, deregulation, economic growth, difference-in-differences, two-way fixed effect estimation

JEL Classification: C13, C18, C23, G21, G28, O47

Acknowledgments: We thank an anonymous referee, Lizzy Berger, Lee Ann Butler, Alan Crane, Kevin Crotty, Ron Liu, Sandra Mortal, Ioannis Spyridopoulos, Cihan Uzmanoglu, Livia Yi, and Ivo Welch (the editor) for their helpful comments. Biz Rasich provided expert copyediting. Author order is reverse alphabetical because otherwise Zdrojewski would probably never be a first-listed co-author. Send correspondence to alex.butler@rice.edu.

Are two-way fixed-effect difference-in-differences estimates blowing smoke? A cautionary tale from state-level bank branching deregulation

We illustrate the sensitivity of two-way fixed effects difference-in-differences estimates to innocuous changes in data structure. Using the staggered rollout of state-level bank branching deregulations, three outcome variables are brought to bear on the interventions: personal income growth (a replication), house prices (new to the literature), and per capita cigarette purchases (a falsification test). Estimates are sensitive to panel length, and the data structure creates the false impression of a causal effect of the interventions on all three outcome variables. We contend that any two-way fixed effects regression using this set of interventions is at risk of generating spurious results.

Studying the causal effect of finance on economic outcomes is notoriously difficult due to the endogenous nature of financial development. One fruitful approach to this challenge uses deregulation of state banking industries throughout the 1970s and 1980s as an exogenous source of variation in states' financial development. The staggered nature of the deregulations is a serendipitous feature of the setting that mitigates concerns of confounding events or trends that might bias the estimated causal effect on economic growth. Jayaratne and Strahan (1996) originate the application of this experimental design, an approach that was pathbreaking and state-of-the-art at the time and remains an enduring empirical setting for researchers (Huang (2008), Chava et al. (2013), Berger et al. (2021)). Despite the attractive features of the experimental setting, recent methodological papers suggest that empirical designs such as this one may inadvertently produce biased estimates. In this paper, we examine whether this indeed occurs.

In settings where units receive treatment at different times, like that of the staggered bank branching deregulations, researchers frequently estimate a Two-Way Fixed Effect (TWFE) regression of the outcome variable on a treatment indicator and fixed effects for unit and year. Goodman-Bacon (2021) decomposes this estimate into a weighted average of difference-in-differences estimates for each intervention, using different treatment timing groups as control groups. The decomposition shows that weights are sensitive to the intervention's position in the panel with respect to time (i.e., when the treatment occurs), such that units that are treated towards the middle of the panel receive relatively large weights. Furthermore, component estimates may be biased if the treatment effect varies over event time. For example, if a positive treatment effect attenuates over time following treatment, then previously treated control units understate the increase in the counterfactual outcome, biasing the difference-in-differences estimate upward. Similarly, if the treatment effect varies over calendar time, the single parameter from a two-way fixed effect regression may weight each calendar-time effect arbitrarily, resulting in a nonsensible summary of the time-varying effects. Recent research (e.g., Callaway and Sant'Anna (2020), Sun and Abraham (2020), among others) offers methodological solutions to these problems.

We contribute to this methods literature by applying it to the setting of staggered bank branching deregulation, an empirical setting that is well-suited for studying the sensitivity of estimates in staggered difference-in-differences tests. We first replicate and extend Jayaratne and Strahan's (1996) seminal result. In their main test, the authors employ a two-way fixed effect regression of real per capita personal income growth on fixed effects for state and year and a treatment indicator equal to one if the state has legalized bank branching by merger and acquisition by that year. They report that bank branching deregulation increases personal income growth by 0.94%. However, other researchers document an effect of deregulation on income growth (Berger et al. (2021)) and firm-level innovation (Chava et al. (2013)) that grows over time. These findings suggest a time-varying treatment effect that, coupled with the staggered implementation of deregulation across states, makes the original estimates sensitive to panel length and prone to bias (Goodman-Bacon (2021)). Whether these problems substantially affect the original Jayaratne and Strahan (1996) estimates is an empirical question that our paper aims to answer.

Based on the traditional approach of a difference-in-differences test with two-way fixed effects, we replicate the estimated treatment effect of deregulation on personal income growth of 0.94%. This estimate is statistically significant and economically large (for context, the United States' real gross domestic product increased about 3% per annum over the sample period). Then, like previous researchers, we apply the Jayaratne and Strahan (1996) empirical design with new outcome variables. We introduce two variables that highlight the possibility of biased estimates. The first is House Price Index (HPI) growth. We reason that deregulation makes banks more competitive, resulting in attractive interest rates for homebuyers and increased housing demand and prices. An alternate channel is that an increase in the supply of bank capital may facilitate quality improvements (Reher (2021)). In standard fashion, we regress state-year House Price Index growth on state and year fixed effects and an indicator for whether the state had deregulated yet. We estimate a statistically significant treatment effect of deregulation on growth in the housing price index of 1.98% per year, which is directionally intuitive but unexpectedly large; the effect equals about a third of the 6% average growth in the housing price index for the state-years in our sample. The large magnitude points to the possibility that the effect may be driven by bias.

We then estimate the same regression with our second novel outcome variable: per capita cigarette purchases, the same data analyzed in Abadie et al. (2010). We select this variable as one that banking regulations are unlikely to affect, therefore making it suitable for a falsification test. Nevertheless, bank branching deregulation appears to have a causal effect on smoking as well: the point estimate of the treatment effect is a statistically significant increase of 4.98 cigarette packs per capita (full population, not just smokers) per year.

The first two effects we present above have surprisingly large magnitudes, whereas the third effect is, also surprisingly, non-zero. A possible explanation is that our estimates suffer from the problems discussed by Goodman-Bacon (2021). We examine this possibility directly. We begin by performing the Goodman-Bacon (2021) decomposition on each of our three estimates. For each outcome variable, over 60% of the weight is attributable to comparisons of treated units to already-treated units—those which are most prone to bias (Goodman-Bacon (2021)). In fact, this problem would present itself using *any* outcome variable over this sample period and set of shocks; the weights are completely determined by the panel of interventions and are not influenced by the outcome variable. To abuse a Latin phrase, *caveat regressor.*

Recall that the influence of an intervention on the TWFE estimate depends partially on its position in the panel; observations near the middle of the time series receive relatively large weights (Goodman-Bacon (2021)). When we use a panel ranging from 1972-1992, deregulation events near the middle of the panel (i.e. Pennsylvania in 1982) are heavily weighted as treatment units. However, changing the sample period may change these interventions' relative positions, reducing their influence in the estimation, and hence changing the estimated treatment effect. We quantify this sensitivity to a seemingly innocuous research decision in our setting.

With the benefit of data not available to the original researchers, we can expand Jayaratne and Strahan (1996)'s original panel by seven years beyond their last year of data (and in the Appendix, up to ten years prior to their first year of data). Adding data from 1993 through 1999 decreases the size of the estimate by 13% from 0.82 to 0.71 (percentage points).[1] Housing price index growth and cigarette sales show an even greater sensitivity to panel length, as we document in Section IV.c.

The sensitivity of our results to innocuous design choices is troubling, but a straightforward alternative estimator mitigates these problems. Callaway and Sant'Anna (2020) estimate "group-time" treatment effects, similar in spirit to an event study, that allow the estimated effects to be dynamic, and they specify either never-treated units or not-yet-treated units as controls for each treatment event. This approach eliminates the use of previously treated units as controls and the bias that results.

We use the Callaway and Sant'Anna (2020) approach to estimate the treatment effect of bank branching deregulation on all three outcome variables. When we do, we obtain treatment effect estimates that are each statistically indistinguishable from zero. Because these results contradict those of the previous approach, we conclude that the apparent results of the two-way fixed effect estimation likely arise due to bias.

Although many papers make use of these staggered deregulations to examine questions like those in Jayaratne and Strahan (1996), ours is not about whether financial development causes economic growth (or house price changes or cigarette consumption). Rather, we use the setting to explore the challenges to traditional difference-in-differences empirical designs. We show the event structure *itself* is generally unsuitable for a TWFE approach; the panel of interventions, under a TWFE methodology, generates a bias

[1] Some states deregulate during the years we add to the panel. To isolate the effect of the panel expansion from the effect of these additional treatments, we exclude these states in this analysis. For example, the 0.82% effect reported over the years 1972 to 1992 and 0.71% over the years 1972 to 1999 each exclude those states that deregulated post-1992. We elaborate in section IV.c.

appearing in estimated effects on the original variable, a related variable, and a completely unrelated variable.

A contemporary paper, Baker et al. (2021), analyzes staggered treatment methodologies through simulations and by re-examining three extant results: Fauver et al. (2017), which studies the effects of staggered adoption of corporate board reforms on firm value, Wang et al. (2021), which studies the effects of staggered legalization of stock repurchases on corporate behaviors, and Beck et al. (2010), which studies the effects of staggered bank branch deregulation (our focus) on income inequality. Comparatively, we emphasize the sensitivity of two-way fixed effect estimates *in practice* by replicating a seminal finding and documenting two implausibly large effects using the same setting. Moreover, our examination of variation in panel length further underscores the sensitivity of two-way fixed effect estimates. Both our paper and Baker et al. (2021) demonstrate that commonly used natural experiments are susceptible in practice to the theoretical critiques of Goodman-Bacon (2021), Callaway and Sant'Anna (2020), and others.

I. Background

Until the 1970s, regulation of banking in the United States largely occurred at the state level. In particular, state branching restrictions limited the ways in which banks could expand their branching footprints, both interstate and intrastate. There were two main types of restrictions that existed on intrastate bank branching: restrictions on mergers and acquisition (M&A) branching and restrictions on de novo branching. M&A branching restrictions prevented banks from converting purchased or subsidiary banks into branches. De novo banking restrictions prevented banks from forming new branches outright within the state. Over Jayaratne and Strahan's sample period, 1972 to 1992, deregulation of branching by M&A mostly occurred before deregulation of de novo branching for each state. In many cases, the two types of deregulation occurred in quick succession, so Jayaratne and Strahan (1996, 1998) use the years of M&A deregulation as the year of deregulation for each state. We do the same and throughout use "deregulation" to refer to deregulation of bank branching by M&A. Figure 1 illustrates graphically the staggered timing of the deregulation events across states.

II. Data

We collect state-level data from the Bureau of Economic Analysis (BEA), the St. Louis Federal Reserve, the Federal Housing Finance Agency (FHFA), and the Center for Disease Control and Prevention (CDC). Our measure of state-level economic growth is personal income growth per capita, as in Jayaratne and Strahan (1996). We take state-level personal income and population figures from the BEA and the consumer price index (CPI) deflator in chained 2015 dollars from the St. Louis Federal Reserve. We express personal income growth per capita in real terms by scaling state-level income by state population and then deflating this per capita figure by CPI. We obtain House Price Index (HPI) levels from the Federal Housing Finance Agency (FHFA) and tobacco consumption data from the CDC.

Over time, the BEA has adjusted its methodology for calculating personal income and therefore revised historical data. So, in order to convincingly replicate the original results of Jayaratne and Strahan (1996), we obtain the earliest available post-1996 vintage of personal income data from the BEA, which was published in 1999 and includes real personal income growth from the period spanning 1961 to 1999. Finally, HPI data spans 1975 to 2020, whereas tobacco consumption data spans 1970 to 2019.

III. Methods

In the canonical difference-in-differences framework, a group of units is treated at the same time, and their changes in outcomes are compared to those of an untreated control group.[2] Parallel trends is the identifying assumption, which states that the outcome variable for the treated group would have changed by the same amount as that of the untreated group, in expectation, if the treated group were left untreated. Using the potential outcome framework, this assumption can be written as $E[Y_1^0|Post] - E[Y_1^0|Pre] = E[Y_0^0|Post] - E[Y_0^0|Pre]$, where Pre and $Post$ are intervals of time on either side of the intervention and Y_g^d refers to the outcome for group g if it were to receive treatment d. The Average Treatment Effect on the Treated (ATT) can then be estimated as β from the equation:

$$Y_{it} = \alpha TREAT_i + \gamma POST_t + \beta(POST * TREAT)_{i,t} + \varepsilon_{i,t} \qquad (1)$$

where $TREAT_i$ is an indicator equal to one if unit i is in the treatment group and $POST_t$ is an indicator equal to one in the second period (the period in which treatment occurs for treated observations).

Researchers have extended this approach to settings where treatment occurs at different times for different units. In particular, they use panel data to estimate Two-Way Fixed Effect (TWFE) regressions of the form $Y_{i,t} = \alpha_t + \gamma_i + \beta D_{i,t} + \epsilon_{i,t}$, where $D_{i,t}$ is an indicator equal to one if unit i is treated in period t. While differential timing is thought to

---

[2] We note that Baker et al. (2021) and Barrios (2021) also provide a discussion and review of the new approaches discussed above, as well as other similar approaches.

alleviate concerns of omitted concurrent events biasing $\hat{\beta}$, recent econometric research has shown that $\hat{\beta}$ in most instances is not interpretable as an estimate of the treatment effect. We begin with a brief review of this research and explain how we will implement its findings in the setting of banking deregulation laws.

a. Goodman-Bacon (2021)

Goodman-Bacon (2021) constructs a setting in which there are K different treatment timing groups such that each unit in group k becomes treated at time t_k (where time is indexed by $t = 1, \dots, T$). For a particular treatment group k, there are three sensible groups one might use as a counterfactual to ascertain the treatment effect associated with group k: units that are never treated, units that are not yet treated by time t_k, and units that are already treated by time t_k. In particular, consider two treatment timing groups, k and l, with $t_k < t_l$, and the group of units that are never treated (denote them as group U). There are four distinct difference-in-differences comparisons one might make using these groups: treatment group k compared with groups l or U around t_k or treatment group l compared with groups k or U around t_l. Letting $POST(a) = [t_a, T]$, $PRE(a) = [1, t_a)$, and $MID(a, b) = [t_a, t_b)$, Goodman-Bacon (2021) defines the 2x2 estimates:

$$\hat{\beta}_{kU}^{2x2} = \left(\bar{y}_k^{POST(k)} - \bar{y}_k^{PRE(k)}\right) - \left(\bar{y}_U^{POST(k)} - \bar{y}_U^{PRE(k)}\right) \tag{2}$$

$$\hat{\beta}_{kl}^{2x2,k} = \left(\bar{y}_k^{MID(k,l)} - \bar{y}_k^{PRE(k)}\right) - \left(\bar{y}_l^{MID(k,l)} - \bar{y}_l^{PRE(k)}\right) \tag{3}$$

$$\hat{\beta}_{kl}^{2x2,l} = \left(\bar{y}_l^{POST(l)} - \bar{y}_l^{MID(k,l)}\right) - \left(\bar{y}_k^{POST(l)} - \bar{y}_k^{MID(k,l)}\right) \tag{4}$$

$\bar{y}_k^{PRE(k)}$ denotes, for instance, the average of the outcome variable y for group k over the period $PRE(k)$. $\hat{\beta}_{kU}^{2x2}$ is the 2x2 difference-in-differences estimate obtained by comparing treatment group k with the untreated group U around t_k. Similarly, $\hat{\beta}_{kl}^{2x2,k}$ is the 2x2 estimate obtained by comparing treatment group k with treatment group l around t_k. Note that the second and third equations involve the same treatment groups, k and l, but the second equation estimates the treatment effect on group k using group l as a comparison, and the third equation estimates the treatment effect on group l using group k as a comparison.

The central result of Goodman-Bacon (2021) is that the $\hat{\beta}$ from the balanced two-way fixed effect regression is a weighted average of all the 2x2 estimates above for each of the K treatment groups. The weights that are applied to each of the 2x2 estimates are an increasing function of the size of the treatment groups. Interestingly, they are also the largest for treatment units that are towards the middle of the panel with respect to time. This implies that by lengthening or shortening the panel, a researcher is implicitly assigning different units to be closer to the middle and therefore to receive greater weight as treatment units. As a result, the two-way fixed effect estimate is sensitive to the length of the panel the researcher has available (or selects), an otherwise innocuous choice.

The Goodman-Bacon (2021) decomposition also reveals that for many of the 2x2 estimates, previously treated units are used as the comparison group. In many situations, using previously treated units as a comparison group generates a biased estimate of the treatment effect. In particular, if the true treatment effect gradually increases/declines over time, then previously treated units are not suitable control units for newly treated units. For example, in a setting with a positive treatment effect that gradually grows stronger in post-treatment periods, using previously treated units as a control group in a difference-in-differences estimate will bias the estimate of the effect downward and so the two-way fixed effect estimate will be biased.

The problems pointed out by Goodman-Bacon (2021) present a threat to the credibility of research designs that have employed this technique, even if they have an otherwise strong causal argument. We conduct a series of analyses to determine to what extent these problems affect tests using this staggered deregulation setting.

b. Callaway and Sant'Anna (2020)

In light of the critique by Goodman-Bacon (2021) and others of two-way fixed effect estimation with staggered intervention, a number of methodological improvements have been suggested to identify the treatment effect in this setting. Intuitively, these approaches correct the biases pointed out in Goodman-Bacon (2021) by being more explicit about which units serve as a control group for a treated unit. In particular, they do not include as controls units that have already been treated and instead make comparisons of treated units to units that are never treated or units that have not yet been treated.

Callaway and Sant'Anna (2020) introduce a setting with τ periods, indexed by $t = 1, \ldots, \tau$. The variable G is defined as the time period in which a unit is first treated so that G defines the treatment timing group to which the unit belongs. For units that are never treated, $G = \infty$. G_g is an indicator variable that equals one if the unit is first treated at time g. Dynamic potential outcomes are denoted by $Y_t(g)$, which is the outcome attained at time t if the unit belongs to treatment group g (i.e., is first treated at time g). Note that in this potential outcome framework, g is the dimension that hypothetically varies. The authors use $Y_t(0)$ to denote the untreated potential outcome at time t for a unit that is never treated. In other words, it is the outcome at time t if the unit were to never receive treatment. They generalize the concept of the Average Treatment effect on the Treated

(ATT) by allowing it to vary over time. Specifically, they define the following, which serves as the building block for the rest of their analysis:

$$ATT(g,t) = E\left[Y_t(g) - Y_t(0) \middle| G_g = 1\right] \tag{5}$$

This is the ATT in period *t* for units treated at *g*. The authors then propose a method of estimating the above, which is unbiased under mild assumptions. An advantage of their approach is that it only requires the assumption to hold conditional on included control variables. This additional degree of flexibility makes the approach valid for a much broader range of research applications. Nevertheless, the approach simplifies dramatically in the case where additional controls are not needed for identification, as in our setting. For expositional clarity, we therefore only explain the approach in a setting without additional covariates. For the full approach including a set of controls, we refer readers to Callaway and Sant'Anna (2020).

The most salient assumption that the researcher must make is analogous to the parallel trends assumption in a standard difference-in-differences setting. In particular, they assume that one of the following holds for each treatment group g and each time $t \geq g$ and $s \geq t$:

$$E\left[Y_t(0) - Y_{t-1}(0) \middle| G_g = 1\right] = E[Y_t(0) - Y_{t-1}(0) | g = \infty] \tag{6}$$

$$E\left[Y_t(0) - Y_{t-1}(0) \middle| G_g = 1\right] = E\left[Y_t(0) - Y_{t-1}(0) \middle| D_s = 0, G_g = 0\right] \tag{7}$$

D_t is an indicator variable equal to one if the unit has received treatment by time t. In words, (6) states that, in expectation, if the units receiving treatment at time g had instead been never treated, they would have changed the same amount between $t-1$ and t as did the never treated group. Similarly, (7) states that, in expectation, if the units receiving treatment at time g had instead been never treated, they would have changed the same amount between $t-1$ and t as did the group of not-yet-treated units (had they also been never treated). Naturally, the authors refer to these assumptions as parallel trends based on "never-treated" (*nev*) and "not-yet-treated" (*ny*) groups, respectively. Under these assumptions, simple algebra shows that that ATT(g,t) can be identified by one of the following:

$$ATT^{nev}(g,t) = E\left[Y_t - Y_{g-1} \middle| G_g = 1\right] - E\left[Y_t - Y_{g-1} \middle| g = \infty\right] \tag{8}$$

$$ATT^{ny}\,(g,t) = E\left[Y_t - Y_{g-1} \middle| G_g = 1\right] - E\left[Y_t - Y_{g-1} \middle| D_t = 0, G_g = 0\right] \tag{9}$$

From inspection, it is clear that the above can be simply estimated by the appropriate sample analogs:

$$\widehat{ATT}^{nev}(g,t) = \left(\bar{Y}_t^g - \bar{Y}_{g-1}^g\right) - \left(\bar{Y}_t^{nev} - \bar{Y}_{g-1}^{nev}\right) \tag{10}$$

$$\widehat{ATT}^{ny}(g,t) = \left(\bar{Y}_t^g - \bar{Y}_{g-1}^g\right) - \left(\bar{Y}_t^{ny} - \bar{Y}_{g-1}^{ny}\right) \tag{11}$$

Callaway and Sant'Anna (2020) also suggest some helpful ways of aggregating the $\widehat{ATT}(g,t)$s into interpretable parameters. In this paper, we employ two in particular, but different settings may make other aggregations more attractive. The first method of aggregation we employ results in parameters interpretable in a similar fashion to standard event studies. In particular, each of the group-time estimated effects is averaged into event-time buckets. In other words, we average the estimated treatment effects based on how many periods a unit has been exposed to treatment. Letting $e = t - g$ denote the number of periods elapsed since treatment, Callaway and Sant'Anna define:

$$\theta_{es}(e) = \sum_g 1\,\{g + e \leq \tau\} P(G = g | G + e \leq \tau) ATT(g, g + e) \tag{12}$$

The above is the average estimated treatment effect e periods after treatment for all those units for which the eth period after treatment is observed. Throughout, we make event-study-style plots by plotting the values of this parameter on the vertical axis and event time on the horizontal.

In order to arrive at a single summary parameter to estimate the treatment effect, we use the approach mentioned by Callaway and Sant'Anna (2020) of simply averaging the group-time treatment effects together. Formally, they define:

$$\theta_W^O = \frac{1}{\kappa}\sum_g \sum_{t=2}^{\tau} 1\{t \geq g\} ATT(g,t) P(G = g | G \leq \tau), \tag{13}$$

where $\kappa = \sum_g \sum_{t=2}^{\tau} 1\{t \geq g\} P(G = g | G \leq \tau)$. θ_W^O is a weighted average of the estimated $ATT(g,t)$s, where the weights increase proportionately to the number of units in the timing group g. The authors note that θ_W^O also puts more weight on groups that participate in treatment for longer, which may or may not be desirable depending on the researcher's goal. In our case, we are searching for a single per-year estimate of the treatment effect to compare with the estimate of Jayaratne and Strahan, so we therefore estimate the treatment effect, θ_W^O, of bank deregulation on all three of our outcome variables.

Because difference-in-difference approaches rely on the validity of a parallel trends assumption, it is important to reconsider equations 6 and 7 in the context of bank branching deregulation and economic growth. In this setting, equation 6 means that in expectation, if states that deregulated in a given year had instead never deregulated, they would have experienced the same sequence of changes in economic growth in the ensuing years as those experienced by states that *never* deregulate. Likewise, equation 7 means that in expectation, if states that deregulated in a given year had instead never deregulated, they would have experienced the same sequence of changes in economic growth in the ensuing years as those experienced by states that had *not yet* deregulated.

IV. Results

Table 1 exhibits our replication of the Jayaratne and Strahan (1996) personal income growth regressions side-by-side with the original. The basic model is the standard TWFE regression, while the regional model replaces year fixed effects with region-year fixed effects, with four geographic regions as defined in Jayaratne and Strahan (1996). Our sample inclusion criteria follow Jayaratne and Strahan (1996). We use data from 1972 to 1992 and exclude state-years in which the deregulation occurred. For example, Texas deregulated in 1988, so we drop the observation of Texas in 1988 from our sample. We exclude Delaware in both specifications and Alaska and Hawaii in the regional specification. Our replicated estimates of 0.94% in the basic model and 0.51% in the regional model are identical to the original estimates to the second decimal place. Thus, we are confident that our data is practically the same as the original authors'.

a. Two-Way Fixed Effect Regressions

In the remainder of our analyses, we depart from the data approach of Jayaratne and Strahan (1996) by including state-years in which deregulations occurred. We do this to construct a balanced panel, which is a requirement of the Goodman-Bacon (2021) decomposition. Columns (1) and (2) of Table 2 show the estimates of the same regressions as Table 1 but now includes these observations. This inclusion reduces the estimated treatment effect of deregulation on personal income growth from 0.94% to 0.82% for the basic model and from 0.51% to 0.38% for the regional model.

We next consider what other economic variables would be plausibly influenced by bank branching deregulation. One such variable is growth in state-level House Price Index (HPI) values. We offer several economic justifications. Deregulation seems to have increased personal income levels, suggesting that people are wealthier overall and thus more able to invest in housing. Bank deregulation may also increase competitiveness in the local banking market, leading banks to offer compelling interest rates to marginal homebuyers, increasing the demand and price for housing. Further, there is some empirical work suggesting a relation between housing costs and access to finance, including Reher (2021), which documents that increased financing for rental improvements causes an increase in average rent.

To investigate whether deregulation causes an increase in HPI growth, we again estimate basic and regional models in Columns (3) and (4) of Table 2. We obtain estimated effects of deregulation of 1.98% and 0.94%, respectively. In the basic model, our estimated effect is statistically significant at the 10% level and is economically large. The standard interpretation of this estimate is that the removal of bank branching restrictions by M&A caused HPI growth to increase 1.98% per year. Although statistically significant and directionally intuitive, the magnitude of this estimate is surprisingly large. The estimate is on a per-year basis, meaning that bank branching deregulation causes housing prices to increase by almost two percentage points more each year. This effect compounds such that, after five years, housing prices would increase more than 10% simply due to the removal of bank branching restrictions.

Next, we repeat this test with a third outcome a variable that is unlikely to be affected by bank deregulation to serve as a falsification test. In Columns (5) and (6) of Table 2, we replace the outcome variable of interest with per capita cigarette pack purchases. We estimate that bank branching deregulation increased per capita cigarette purchases each year by 4.98 packs in the basic model and 4.39 in the regional model. Both estimates are statistically significant at the 1% level. This effect is large, especially because it is not a per-smoker effect but rather a per-person effect. Further, there is no clear economic mechanism connecting bank regulation with smoking behavior, consistent with this result being a false positive.

b. Goodman-Bacon (2021) decomposition

We next consider to what extent the Goodman-Bacon (2021) critique can shed light on the results we obtain in our two-way fixed effect regressions. For each of the three estimates, we perform the Goodman-Bacon (2021) decomposition.[3] We then group the 2x2 estimates into one of four Treatment vs. Control group types: Earlier vs. Later Treated compares treated units to units that are not yet treated, Later vs. Always Treated compares treated units to units that are always treated in the sample, Later vs. Earlier Treated compares treated units to units that are previously treated in the sample, and Treated vs. Untreated compares treated units with units that are never treated in the sample. Under the Goodman-Bacon (2021) critique, it is the Later vs. Always Treated and Later vs. Earlier Treated comparisons that may be biased.

Table 3, Panel A presents this analysis for our estimated effect on personal income growth. We show the total weight, the average estimate, and the product (total weight)*(average estimate) for each of the above comparison groups. Inappropriate comparison groups, Later vs. Always Treated and Later vs. Earlier Treated, receive a total of 61% of the weight. While this is an item of concern, we note that the average estimates for these groups curiously seem to be bringing the estimate down. Specifically, the

[3] We perform the Goodman-Bacon (2021) decomposition using the R package "bacondecomp".

weighted average estimate across the unbiased comparison groups is 1.17%, which is higher than the overall estimate of 0.82%. At first glance this suggests the two-way fixed effect approach of Jayaratne and Strahan (1996) may have understated the true effect of deregulation on personal income growth.

In Table 3 Panels B and C, we perform the same analysis for the housing price index growth and cigarette pack purchases regressions, respectively. Panel B shows that 67% of the weight of the estimated effect on housing price index growth comes from biased Later vs. Always Treated or Later vs. Earlier Treated comparisons. Again, a cursory glance suggests that the two-way fixed effect estimate may be biased downward by these comparisons; the weighted average estimate across only the unbiased groups is 2.56%, greater than the 1.98% two-way fixed effect estimate. Panel C reveals that 61% of the estimated effect of deregulation on cigarette purchases comes from biased Later vs. Always Treated or Later vs. Earlier Treated comparisons. For this variable, however, the biased comparisons are responsible for the estimate. The far right column shows that the almost-five-pack estimated effect stems mostly from the Later vs. Always Treated comparison group.

We present graphical depictions of the Goodman-Bacon decompositions of our estimates in Figure 2. The vertical axis is the estimated effect of each component estimate, and the horizontal axis is its weight. Each point represents a single 2x2 comparison. For example, one of the Earlier vs. Later Treated points represents a 2x2 estimate using units treated in 1988 as controls for units treated in 1987, one of the Later vs. Always Treated points represents a 2x2 estimate using units that are always treated in the sample as controls for units treated in 1988, one of the Later vs. Earlier Treated points represents a 2x2 estimate using units treated in 1987 as controls for units treated in 1988, and one of the Treated vs. Untreated points represents a 2x2 estimate using never treated units as controls for units treated in 1987. All three plots show a handful of green triangles pushed to the right; these are Later vs. Always Treated estimates with outsized weights, as we would expect to see based on the previously discussed tables. Each plots also shows a wide dispersion of estimated effects coming from these comparisons ranging from -3.5% to 5% for personal income growth, -13% to 26% for housing price index growth, and -22 packs to 28 packs for cigarette purchases. In short, the figure illustrates the wide dispersion of component estimates and their weights, which jointly make the two-way fixed effect estimate prone to false positive results.

c. Estimates' sensitivity to panel length

As we evaluate a parallel trends assumption in a staggered setting like this one, we are averaging over many pre- and post-intervention comparisons, a crucial point made clear in Goodman-Bacon (2021). Moreover, as we add more time periods to the data, the pre- and post-intervention averages can change, in turn changing our estimates of a treatment effect.[4] We now demonstrate the sensitivity of our two-way fixed effect estimate of the effect of deregulation on personal income growth to panel length. Our data cover 1962 to 1999, allowing us to extend our original sample ten years prior to and up to seven years after the original sample. A number of states experience deregulation during these additional years. Therefore, to isolate the effect of panel expansion from the effect of additional treatments, we exclude these states when making our expansions. For example,

[4] We thank the referee for emphasizing this point.

Arkansas deregulated in 1994 and so is excluded from our panel when we compare the estimate over the years 1972 to 1992 to the estimate over the years 1972 to 1999.

Table 4 presents our two-way fixed effect regressions for each variable with the panel expanded up to ten years into the future. The different sample periods used for the variables are due to data constraints; the personal income data ends in 1999 and the housing price index begins in 1976. We exclude states that deregulated after the end of the original sample (1992), to facilitate comparisons. We provide other expansions in the Appendix, including demonstrations of the component estimates and weights changing across subsamples (see Table A4 for personal income growth regressions, Table A5 for housing price index, and Table A6 for cigarette purchases; regional model expansions for personal income growth can be found in Table A7).

Columns (1) and (2) of Table 4 show an estimated treatment effect on personal income growth of 0.82% over the original sample period and of 0.71% over the period expanded seven years into the future. In other words, the size of the effect was reduced by 13.4% simply by including an additional seven years. Columns (3) and (4) show an estimated treatment effect on housing price index growth of 2.10% over the original sample period and of 2.31% over the period expanded ten years into the future. This is a reduction of the size of the effect by 10% due to the inclusion of ten additional years. Columns (5) and (6) show an estimated treatment effect on per capita cigarette purchases of 6.71 packs over the original sample period and of 11.75 packs over the period expanded ten years into the future. The size of the effect nearly doubles simply due to the inclusion of ten additional years.

Table 4 demonstrates the sensitivity of our two-way fixed effect estimates to the length of the panel, an undesirable feature of this approach. Further, it reveals that the weights assigned to a component estimate may be economically arbitrary. This is a major reason why one should not estimate a treatment effect by simply taking the weighted average of the Goodman-Bacon decomposed estimates for unbiased comparison groups; doing so still arbitrarily weights comparisons. Another reason is the lack of a standard error to use in statistical inference. For these two reasons, we proceed in implementing the Callaway and Sant'Anna (2020) estimates.

d. Callaway and Sant'Anna (2020) estimates

We now present our estimated effect of deregulation on each of our three outcome variables, using the approach of Callaway and Sant'Anna (2020).[5] Throughout, we use the not-yet-treated units as controls, as there are more of them available as data points than the never treated units, which Callaway and Sant'Anna (2020) also allow as controls. For each of the three outcome variables, we estimate the single summary parameter estimate, θ_W^O, as introduced in section III.b. Recall that this parameter is constructed as a probability-weighted average of group-time treatment effects, $ATT(g,t)$, which are estimated as described in section III.b. Additionally, we present an event-study-style depiction of the dynamic specification of each treatment effect. Figure 3 presents this information.

In the top panel of Figure 3, we estimate a per-year effect (ATT) of deregulation on personal income growth of 0.44% that is not statistically significant. The panel also shows the dynamic specification of the treatment effect, with year 0 corresponding to the year of

[5] The authors provide a companion R package called "did" that we use to implement their approach.

deregulation. There is no treatment effect that is statistically greater than zero, even up to ten years out. This finding contradicts the result of our original two-way fixed effect regression. To be confident that the approach of Callaway and Sant'Anna (2020) is a methodological improvement, we test whether it is able to diagnose our other two-way fixed effect results as false positives.

The second and third panels of Figure 3 report per-year estimated effects (ATT's) of deregulation of 1.65% on housing price index growth and of 0.59 packs on cigarette purchases. Neither effect is statistically significant, and the point-estimated effect on cigarette purchases is significantly reduced towards zero, as we would expect if the two-way fixed effect result is spurious. The second and third panels of Figure 3 also depict dynamic plots of these effects with confidence bands. Even allowing for a dynamic specification, neither variable shows a statistically significant response to the bank branching deregulation. We offer this result as evidence that the Callaway and Sant'Anna (2020) approach is indeed the correct technique for this setting, as it appropriately fails to find an effect in our falsification tests.

Figure 3 also provides an opportunity to highlight an advantage of the Callaway Sant'Anna (2020) approach. In each of the figures, the pre-treatment years are in red, and are statistically indistinguishable from zero. This indicates that there are no observable departing trends in the outcome variables between deregulating states and control states prior to deregulation. Although the parallel trends assumption is inherently untestable, a visual inspection of the figures suggests that there are no obvious pre-treatment trends by using the event-study approach of Callaway and Sant'Anna (2020).

V. Conclusions

This paper applies new econometric techniques for staggered adoption difference-in-differences designs to re-examine the use of staggered bank branching deregulations in the U.S. as a natural experiment. We demonstrate that the setting is prone to delivering false positive results by extending the specification with two novel outcome variables, housing price index growth and per capita cigarette purchases. Under the two-way fixed effect specification with staggered bank branching deregulations, both novel variables have implausible effect magnitudes. We show that these estimates suffer from the problems pointed out by the Goodman-Bacon (2021) critique, including biased component estimates, overweighting of particular observations, and sensitivity to the panel length. Finally, we implement the bias-corrected approach of Callaway and Sant'Anna (2020), which reveals statistically insignificant estimated effects on each of the three outcome variables. Despite our three results using the two-way fixed effect specification, bank branching deregulation did not cause a statistically significant increase in personal income growth, housing price index growth, or cigarette consumption.

The paper contributes to the burgeoning literature on difference-in-differences with staggered treatment by applying econometric theory to the data in a familiar, intuitive, and memorable setting. It demonstrates the practical hazards of naively implementing a two-way fixed effect regression in a setting with staggered adoption. Finally, it provides an intuitive exposition of an appropriate way to estimate treatment effects in this setting. Altogether, it is a useful reference for future empirical research using staggered adoption difference-in-differences methods.

References

Abadie, A., A. Diamond, and J. Hainmueller, 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of American Statistical Association.* 105, 493-505.

Baker, A., D. Larcker, and C. Wang, 2021. "How Much Should We Trust Staggered Difference-In-Differences Estimates?", *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3794018*

Barrios, J.M., 2021. "Staggeringly Problematic: A Primer on Staggered DiD for Accounting Researchers.", *Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3794859*

Beck, T., R. Levine, and A. Levkov, 2010. "Big bad banks? The winners and losers from bank deregulation in the United States." *Journal of Finance*. 65(5), 1637-1667.

Berger, E., A. Butler, E. Hu, and M. Zekhnini, 2021. "Financial Integration and Credit Democratization: Linking Banking Deregulation to Economic Growth." *Journal of Financial Intermediation.* 45, 100857.

Callaway, B. and P. H. Sant'Anna, 2020. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics.*

Chava, S., A. Oettl, A. Subramanian, and K. Subramanian, 2013. "Banking Deregulation and Innovation." *Journal of Financial Economics*. 109, 759-774.

Fauver, L., M. Hung, X. Li, and A.G. Taboada, 2017. "Board reforms and firm value: Worldwide evidence." *Journal of Financial Economics.* 125, 120-142.

Goodman-Bacon, A., 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics.* Forthcoming.

Huang, R., 2008. "Evaluating the Real Effect of Bank Branching Deregulation: Comparing Contiguous Counties across US State Borders." *Journal of Financial Economics*. 87, 678-705.

Jayaratne, J. and P. Strahan, 1996. "The Finance-Growth Nexus: Evidence from Bank Branch Deregulation." *Quarterly Journal of Economics*. 111, 639–670.

Jayaratne, J. and P. Strahan, 1998. "Entry Restrictions, Industry Evolution, and Dynamic Efficiency: Evidence from Commercial Banking." *Journal of Law and Economics*. 41, 239–274.

Reher, M., 2021. "Finance and the Supply of Housing Quality." *Journal of Financial Economics*, *Forthcoming.*

Sun, L. and S. Abraham, 2020. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” *Journal of Econometrics*.

Wang, Z., Q. E. Yin, and L. Yu, 2021. “Real effects of share repurchases legalization on corporate behaviors.” *Journal of Financial Economics.* 140, 197-219.

Table 1: Replication of JS (1996) Personal Income Growth Regression

	Basic Model		Regional Model	
	JS-Original	Replication	JS-Original	Replication
	(1)	(2)	(3)	(4)
Deregulation Coefficient	0.94*	0.94*	0.51*	0.51*
Observations	1015	1015	974	974
R^2		51%		67%
Adjusted R^2	49%	48%	62%	62%

Note: * denotes statistical significance at the 5% level

Description: This table presents original results from Table II of Jayaratne and Strahan (1996) side-by-side with our replications. The basic model is given by $Y_{i,t}/Y_{i,t-1} = \alpha_t + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where $Y_{i,t}$ is state i's real per capita personal income in year t, and $D_{i,t}$ is an indicator equal to one for state-years with no M&A banking restrictions. The regional model is $Y_{i,t}/Y_{i,t-1} = \alpha_{t,j} + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where j indexes the four regions specified in Jayaratne and Strahan (1996): Northeast, South, Midwest, and West. Coefficients are reported in percentage points. The data spans years 1972 to 1992. State-years in which the deregulation occurred are excluded, as are all observations from Delaware. The regional regressions also exclude Alaska and Hawaii. We follow Jayaratne and Strahan in using White heteroskedasticity robust standard errors.

Interpretation: We are able to replicate the original result of Jayaratne and Strahan (1996) that deregulation of bank branching causes real personal income growth to increase by 0.94% per year under the basic model and by 0.51% per year under the model with region-year fixed effects.

Table 2: Balanced Panel Regressions

	Personal Income Growth (%)		House Price Index Growth (%)		Per Capita Cigarette Purchases (Packs)	
Model:	Basic	Regional	Basic	Regional	Basic	Regional
	(1)	(2)	(3)	(4)	(5)	(6)
Deregulation Coefficient	0.82***	0.38	1.98*	0.94	4.98***	4.39***
	(2.89)	(1.52)	(1.96)	(1.09)	(4.09)	(3.42)
Observations	1050	1008	850	816	1050	1008
R^2	0.51	0.67	0.22	0.42	0.89	0.89
Adjusted R^2	0.48	0.62	0.15	0.33	0.88	0.87

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Description: This table presents the estimated effect of deregulation on our three outcome variables using two-way fixed effects regressions. The top row identifies the dependent variable of the regression, with its units in parentheses. The second row specifies whether a basic or regional model is employed. The basic model is given by $Y_{i,t} = \alpha_t + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where $Y_{i,t}$ is state i's outcome variable in year t, and $D_{i,t}$ is an indicator equal to one for state-years with no M&A banking restrictions. The regional model is $Y_{i,t} = \alpha_{t,j} + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where j indexes the four regions specified in Jayaratne and Strahan (1996): Northeast, South, Midwest, and West. Delaware is excluded in all regressions, and the regional regressions also exclude Alaska and Hawaii. Personal income and cigarette purchases data span years 1972 to 1992, and House Price Index data spans years 1976 to 1992. Standard errors are clustered at the year and state level, and T-statistics appear in parentheses below the relevant coefficient.

Interpretation: Standard two-way fixed effect regressions suggest that bank branching deregulation causes real per capita personal income growth to increase by 0.82 percentage points, house price index growth to increase by 1.98 percentage points, and per capita cigarette purchases to increase by 4.98 packs. Regressions with regional time-varying fixed effects suggest that bank branching deregulation causes real per capita personal income growth to increase by .38 percentage points, house price index growth to increase by 0.94 percentage points, and per capita cigarette purchases to increase by 4.39 packs.

Table 3: Decomposition of Effect of Deregulation on Outcome Variables

Panel A: Personal Income Growth (%)			
Comparison Type	Weight	Average Estimate	Weight*Average Estimate
Earlier vs Later Treated	0.29	1.31	0.38
Later vs Always Treated	0.43	0.45	0.19
Later vs Earlier Treated	0.18	0.9	0.16
Treated vs Untreated	0.11	0.8	0.09
Total:	1.00		0.82

Panel B: House Price Index Growth (%)			
Comparison Type	Weight	Average Estimate	Weight*Average Estimate
Earlier vs Later Treated	0.23	3.26	0.74
Later vs Always Treated	0.48	1.34	0.64
Later vs Earlier Treated	0.19	2.61	0.50
Treated vs Untreated	0.10	0.95	0.10
Total:	1.00		1.98

Panel C: Per Capita Cigarette Purchases (Packs)			
Comparison Type	Weight	Average Estimate	Weight*Average Estimate
Earlier vs Later Treated	0.29	5.01	1.45
Later vs Always Treated	0.43	12.01	5.14
Later vs Earlier Treated	0.18	-3.37	-0.59
Treated vs Untreated	0.11	-9.44	-1.01
Total:	1.00		4.98

Description: This table decomposes the TWFE-estimated effects of branching deregulation on our three outcome variables. For a unit that is treated in a given period, a 2x2 estimate of the treatment effect is obtained by comparing its change in the outcome variable with that of a control group. In particular, such a unit can be compared to a unit that is treated later, always treated, treated earlier, or never treated in the sample. This table presents the weight and average estimated treatment effect across treated units for each comparison group. The weights are derived in Goodman-Bacon (2021). Panels A, B, and C show the decomposition for personal income growth, house price index growth, and cigarette purchases, respectively.

Interpretation: More than 60% of the weight of each TWFE estimate is attributable to inappropriate control group comparisons, namely, Later vs Always Treated comparisons and Later vs Earlier Treated comparisons. The deregulation coefficients obtained in Table 2 (Columns 1,3, and 5) are indeed the weighted average of these 2x2 estimates, exhibited in the bottom-right of each panel.

Table 4: Expanded Panel Regressions

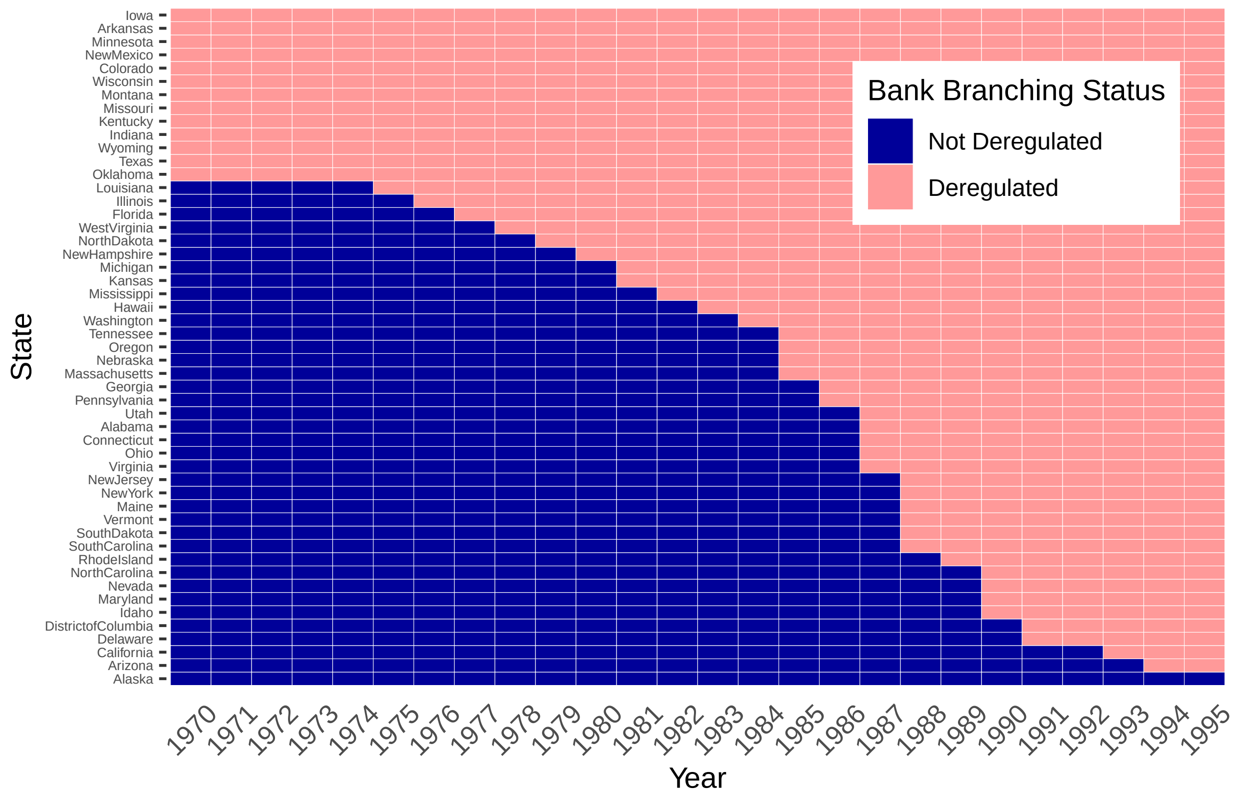
	Personal Income Growth (%)		House Price Index Growth (%)		Per Capita Cigarette Purchases (Packs)	
Sample:	Base 1972-1992	Expansion 1972-1999	Base 1976-1992	Expansion 1976-2002	Base 1972-1992	Expansion 1972-2002
	(1)	(2)	(3)	(4)	(5)	(6)
Deregulation Coefficient	0.82**	0.71***	2.10*	2.31***	6.71***	11.75***
	(2.81)	(3.14)	(1.98)	(3.23)	(5.38)	(10.31)
Observations	987	1316	799	1269	987	1457
R^2	0.49	0.49	0.21	0.21	0.89	0.89
Adjusted R^2	0.46	0.46	0.14	0.16	0.88	0.88

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Description: This table presents the same regressions as Table 2 but now post-expands the panel to include up to 10 additional years. The regression is $Y_{i,t}/Y_{i,t-1} = \alpha_t + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where $Y_{i,t}$ is state i's housing price index level in year t, and $D_{i,t}$ is an indicator equal to one for state-years with no M&A banking restrictions. The column header indicates whether the regression is using a base or extended sample and the years it includes. Delaware and three states that deregulated after 1992 are excluded. Standard errors are clustered at the year and state level, and T-statistics appear in parentheses below the relevant coefficient, which is expressed in percentage points. The different sample periods used for the variables are due to data constraints; the personal income data ends in 1999 and the housing price index begins in 1976.

Interpretation: In the setting of staggered state-level bank branching deregulation, TWFE estimators may be sensitive to changes in panel length in terms of the coefficients' magnitudes and statistical significance.

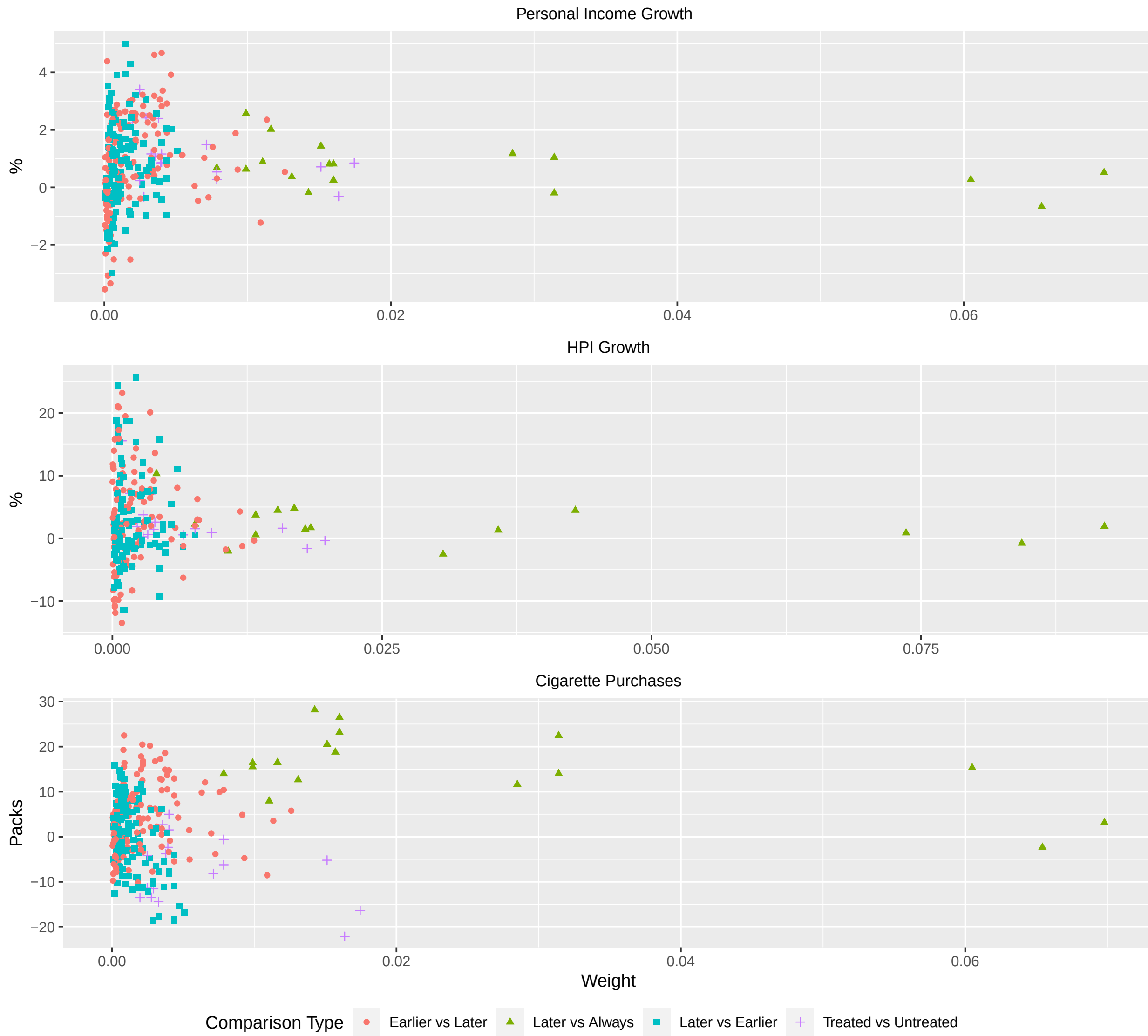
Figure 1 : Deregulation Timing Across States



Description: The figure shows the timing of bank branch deregulation for each state and the District of Columbia. Pink/rose corresponds to years in which deregulation occurs or has already occurred. Blue corresponds to years prior to deregulation.

Interpretation: The figure demonstrates the variation in treatment timing across states.

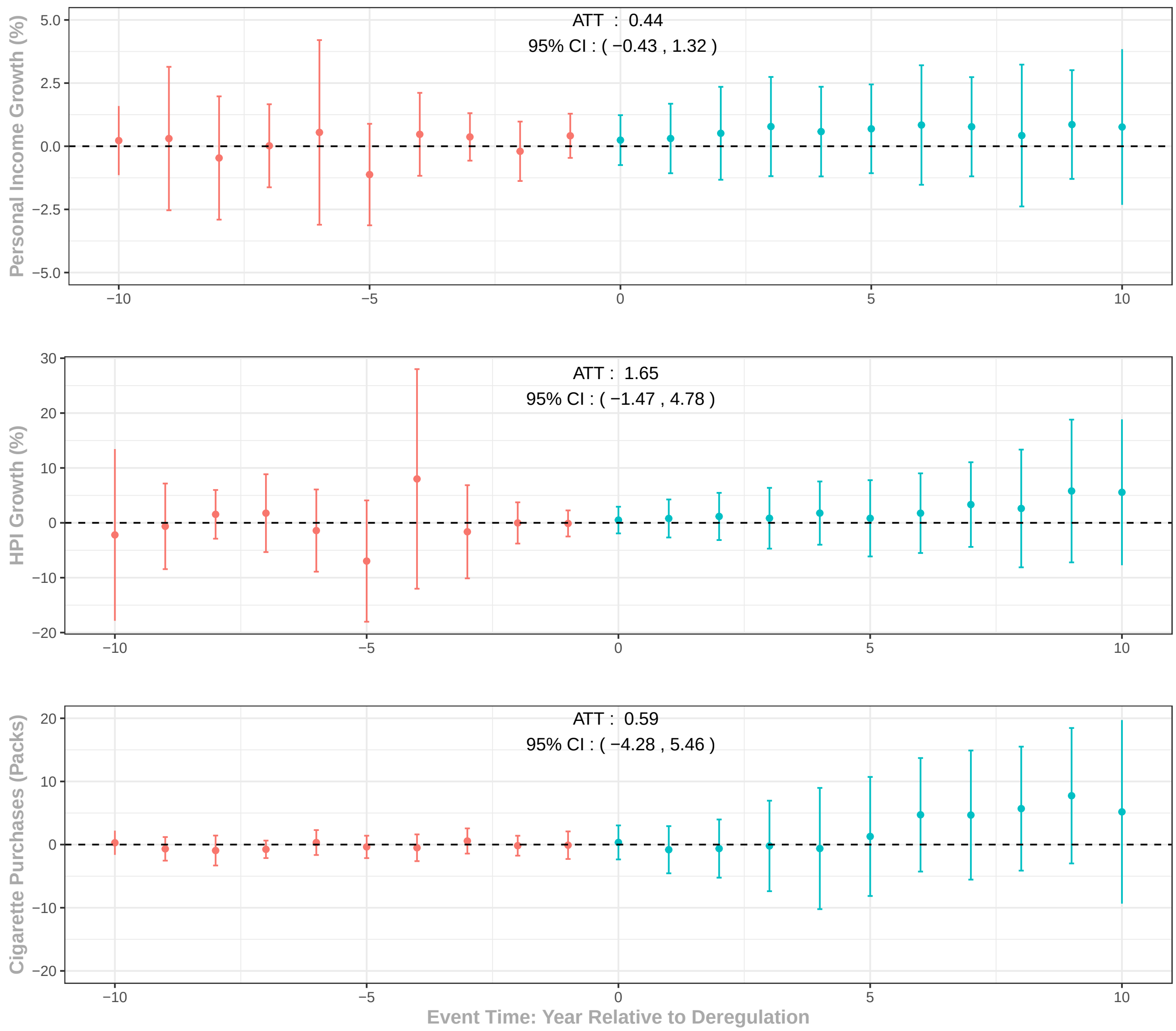
Figure 2: Decomposed Effect of Bank Deregulation



Description: The figure plots each 2x2 difference-in-differences component estimate from the decomposition theorem against its weight in the context of banking deregulation. The green triangles are terms in which one timing group acts as the treatment group and the always-deregulated states act as the control group. The purple crosses are terms in which one timing group acts as the treatment group and the never-deregulated states act as the control group. The red circles (blue squares) are the timing-only terms, which use earlier (later) treated units as a treatment group and later (earlier) treated units as a control group.

Interpretation: The component estimates of the TWFE-estimated effect on each outcome variable demonstrate a wide range of estimated effects and associated weights, and the TWFE estimate is heavily influenced by a small number of component estimates.

Figure 3: Dynamic Effects of Banking Deregulation



Description: This figure presents a depiction of the dynamic treatment effect of deregulation on our three outcome variables in event time, where time 0 corresponds to the year of deregulation. The treatment effects and standard errors are calculated and aggregated into a dynamic event-time plot following the recommended approach of Callaway and Sant'Anna (2020). The bands around each point are 95% confidence intervals using their suggested bootstrap methodology. The overall ATT (θ_W^O in equation 13) is overlaid with its confidence interval in the top center of each graph.

Interpretation: Using the event-study-style, bias-corrected approach of Callaway and Sant'Anna (2020), we find no statistically significant effect of bank branching deregulation on any of the outcome variables, even up to ten years following treatment.

Appendix

Table A1 : TWFE Regressions, Excluding "Always Treated" States

	Personal Income Growth	HPI Growth	Cigarette Packs Sold
	(1)	(2)	(3)
Deregulation Coefficient	1.09***	2.50**	-0.27
	(3.59)	(2.42)	(-0.22)
Observations	798	646	798
R^2	0.56	0.26	0.91
Adjusted R^2	0.53	0.2	0.9

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Description: This table is similar to Table 2, except it removes all "Always Treated' ' states from the sample. We include only the basic model here, given by $Y_{i,t} = \alpha_t + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where $Y_{i,t}$ is state i's real per capita personal income growth, housing price index growth, or per capita cigarette pack purchases in year t, and $D_{i,t}$ is an indicator equal to one for state-years with no M&A banking restrictions. Personal income and cigarette purchases data span years 1972 to 1992, and House Price Index data spans years 1976 to 1992. Standard errors are clustered at the year and state level. Growth rates are reported in percentage points, and T-statistics appear in parentheses below the relevant coefficient.

Interpretation: The removal of "Always Treated" states from the two-way fixed effect regressions alters the magnitudes of the estimated effects on personal income growth and housing price index growth slightly, but they each remain statistically and economically significant and positive. However, removing the "Always Treated" states caused the estimated effect on cigarette pack purchases to change sign and become insignificant.

Table A2 : CSA Estimator of Deregulation Effect on Personal Income Growth by Region

				95% Confidence Interval	
Region	Number of Control States	ATT	Standard Error	Lower Bound	Upper Bound
Northeast	8	-0.090	0.798	-1.653	1.474
West	7	1.895	0.574	0.770	3.020
Midwest	11	0.797	0.841	-0.851	2.445
South	11	0.528	0.641	-0.728	1.784

Description: This table presents the estimated treatment effect of bank branching deregulation on personal income growth, using the approach of Callaway and Sant'Anna (2020) with not-yet-treated units as controls. Here, we conduct the analysis separately over sub-samples for each of the four regions identified in Jayaratne and Strahan (1996). The effect has been aggregated into a single overall parameter for each region, across treatment groups and event times, following the suggested methodology of Callaway and Sant'Anna (2020). Similarly, the standard errors and confidence intervals are constructed according to their bootstrap methodology. It also shows the maximum number of possible control states within each region, ie, how many had not deregulated by 1972.

Interpetation: Due to the small number of states within each cluster, statistical inference from this table is probably not reliable. Still, the bias-corrected approach of Callaway and Sant'Anna (2020) shows that in three of the four regions, the estimated treatment effect is still indistinguishable from zero, and in the Northeast it's point estimate is even negative.

Table A3 : CSA Estimator of Effect on Personal Income Growth, Controlling for Lagged Growth

		95% Confidence Interval	
ATT	Standard Error	Lower Bound	Upper Bound
-0.309	0.416	-1.126	0.507

Description: This table presents the estimated treatment effect of bank branching deregulation on personal income growth, using the approach of Callaway and Sant'Anna (2020) with not-yet-treated units as comparison units. This table differs from the ATT estimated in the top panel of Figure 3 by including as a control a one year lag of the outcome variable, personal income growth. We employ the double robust approach of Callaway Sant'Anna (2020) to implement the control. The effect has been aggregated into a single overall parameter across treatment groups and event times, following the suggested methodology of Callaway and Sant'Anna (2020). Similarly, the standard errors and confidence intervals are constructed according to their bootstrap methodology.

Interpetation: Controlling for the previous year's real personal income growth does not significantly alter our finding. The estimated treatment effect of deregulation is still statistically indistinguishable from zero, and is in fact negative.

Table A4 : Personal Income Growth Two-Way Fixed Effect Estimates : Panel Expansions

Panel A: Personal Income Regressions with Extended Samples

	Pre-Expansion			Post-Expansion		
	1962-1992	1967-1992	1972-1992	1972-1992	1972-1997	1972-1999
	(1)	(2)	(3)	(4)	(5)	(6)
Deregulation Coefficient	1.00***	1.01***	1.09***	0.82**	0.75***	0.71***
	(3.47)	(3.71)	(3.59)	(2.81)	(3.15)	(3.14)
Observations	1178	988	798	987	1222	1316
R^2	0.5	0.54	0.56	0.49	0.48	0.49
Adjusted R^2	0.47	0.51	0.53	0.46	0.45	0.46

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel B: Personal Income Growth Weights with Extended Samples

	Pre-Expansion			Post-Expansion		
	1962-1992	1967-1992	1972-1992	1972-1992	1972-1997	1972-1999
Earlier vs Later Treated	0.63	0.58	0.5	0.32	0.21	0.18
Later vs Always Treated				0.48	0.54	0.56
Later vs Earlier Treated	0.17	0.22	0.31	0.2	0.25	0.26
Treated vs Untreated	0.2	0.2	0.19			

Panel C: Personal Income Growth Estimates with Extended Samples

	Pre-Expansion			Post-Expansion		
	1962-1992	1967-1992	1972-1992	1972-1992	1972-1997	1972-1999
Earlier vs Later Treated	1.13	1.22	1.31	1.31	1.31	1.31
Later vs Always Treated				0.45	0.45	0.42
Later vs Earlier Treated	0.49	0.59	0.45	0.9	0.95	0.91
Treated vs Untreated	0.9	0.9	0.9			

Description: This table shows a broad set of both pre- and post- expansions. The regression is $Y_{i,t}/Y_{i,t-1} = \alpha_t + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where $Y_{i,t}$ is state i's real personal income per capita in year t, and $D_{i,t}$ is an indicator equal to one for state-years with no M&A banking restrictions. The column header indicates the direction of the expansion and the years it includes. Delaware and the three states that deregulated after 1992 are excluded. Standard errors are clustered at the year and state level with T-statistics in parentheses. Panel A shows regressions and Panels B and C show corresponding decomposed weights and estimates.

Interpretation: Sensitivity of estimates to changes in panel length stem from the sensitivity of both component weights and estimates.

Table A5 : House Price Index Growth Two-Way Fixed Effect Estimates : Panel Expansions

Panel A: HPI Estimates with Extended Samples

	Basic Model			Regional Model		
	1976-1992	1976-1997	1976-2002	1976-1992	1976-1997	1976-2002
	(1)	(2)	(3)	(4)	(5)	(6)
Deregulation Coefficient	2.10*	2.90***	2.31***	0.94	1.68**	1.41**
	(1.98)	(3.46)	(3.23)	(1.02)	(2.31)	(2.27)
Observations	799	1034	1269	765	990	1215
R^2	0.21	0.21	0.21	0.41	0.42	0.42
Adjusted R^2	0.14	0.15	0.16	0.31	0.33	0.34

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel B: HPI Estimates and Weights with Extended Samples

	Weights			Average Estimates		
	1976-1992	1976-1997	1976-2002	1976-1992	1976-1997	1976-2002
Earlier vs Later Treated	0.25	0.15	0.11	3.26	3.26	3.26
Later vs Always Treated	0.53	0.59	0.61	1.34	2.59	1.86
Later vs Earlier Treated	0.22	0.26	0.28	2.61	3.37	2.9

Description: This table shows a set of post-expansions of the HPI regressions: $Y_{i,t}/Y_{i,t-1} = \alpha_t + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where $Y_{i,t}$ is state i's housing price index in year t, and $D_{i,t}$ is an indicator equal to one for state-years with no M&A banking restrictions. The column header indicates the years the sample includes. Delaware and the three states that deregulated after 1992 are excluded. Standard errors are clustered at the year and state level with T-statistics in parentheses. Panel A shows basic and regional specifications (for reference) and and Panels B and C show the decomposed weights and estimates of the basic model.

Interpretation: Sensitivity of estimates to changes in panel length stem from the sensitivity of both component weights and estimates. While we don't focus on this in the paper, we note that the regional specifications are also sensitive to changes in panel length.

Table A6 : Cigarette Purchase Two-Way Fixed Effect Estimates : Panel Expansions

Panel A: Cigarette Purchase Estimates with Extended Samples

	Basic Model			Regional Model		
	1972-1992	1972-1997	1972-2002	1972-1992	1972-1997	1972-2002
	(1)	(2)	(3)	(4)	(5)	(6)
Deregulation Coefficient	6.71***	9.67***	11.75***	5.82***	8.14***	9.74***
	(5.38)	(8.15)	(10.31)	(4.33)	(6.34)	(7.91)
Observations	987	1222	1457	945	1170	1395
R^2	0.89	0.88	0.89	0.89	0.88	0.89
Adjusted R^2	0.88	0.87	0.88	0.88	0.87	0.87

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Panel B: Cigarette Purchases Estimates and Weights with Extended Samples

	Weights			Average Estimates		
	1972-1992	1972-1997	1972-2002	1972-1992	1972-1997	1972-2002
Earlier vs Later Treated	0.32	0.21	0.15	5.01	5.01	5.01
Later vs Always Treated	0.48	0.54	0.57	12.01	15.7	18.03
Later vs Earlier Treated	0.2	0.25	0.28	-3.37	0.48	2.46

Description: This table shows a set of post-expansions of the cigarette purchase regressions: $Y_{i,t} = \alpha_t + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where $Y_{i,t}$ is state i's per capita cigarette purchases in year t, and $D_{i,t}$ is an indicator equal to one for state-years with no M&A banking restrictions. The column header indicates the years the sample includes. Delaware and the three states that deregulated after 1992 are excluded. Standard errors are clustered at the year and state level with T-statistics in parentheses. Panel A shows basic and regional specifications (for reference) and and Panels B and C show the decomposed weights and estimates of the basic model.

Interpretation: Sensitivity of estimates to changes in panel length stem from the sensitivity of both component weights and estimates. While we don't focus on this in the paper, we note that the regional specifications are also sensitive to changes in panel length.

Table A7 : Personal Income Growth Regressions: Regional Model Expansions

	Pre-Expansion			Post-Expansion		
	1962-1992	1967-1992	1972-1992	1972-1992	1972-1997	1972-1999
	(1)	(2)	(3)	(4)	(5)	(6)
Deregulation Coefficient	0.64**	0.70**	0.72**	0.37	0.38*	0.35*
	(2.17)	(2.57)	(2.41)	(1.39)	(1.72)	(1.71)
Observations	1147	962	777	945	1170	1260
R^2	0.6	0.65	0.66	0.65	0.64	0.64
Adjusted R^2	0.54	0.59	0.6	0.6	0.59	0.59

Note: *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Description: This table shows a broad set of both pre- and post- expansions of the regional specification for personal income growth. The regression is $Y_{i,t}/Y_{i,t-1} = \alpha_{t,j} + \gamma_i + \beta D_{t,i} + \epsilon_{t,i}$, where $Y_{i,t}$ is state i's real personal income per capita in year t, $D_{i,t}$ is an indicator equal to one for state-years with no M&A banking restrictions, and j indexes the four regions specified in Jayaratne and Strahan (1996): Northeast, South, Midwest, and West.. The column header indicates the direction of the expansion and the years it includes. Delaware and the three states that deregulated after 1992 are excluded. Standard errors are clustered at the year and state level with T-statistics in parentheses.

Interpretation: While not a focus of our paper, we include this table to demonstrate that the regional specifications are also sensitive to changes in panel length, though perhaps less so in the case of personal income growth.