Using Penalized Synthetic Controls on Truncated data: A Case Study on Effect of Marijuana Legalization on Direct Payments to Physicians by Opioid Manufacturers

Bikram Karmakar, Gourab Mukherjee and Wreetabrata Kar[1]

Abstract

Amid increasing awareness regarding opioid addiction, medical marijuana has emerged as a substitute to opioids for pain management. Concurrently, opioid manufacturers are putting significant research into making opioids safer yet effective. Interactions between these manufacturers and physicians are critical to advance existing pain management protocols. Direct payments from opioid manufacturers to physicians are established practices that often moderates such interactions. We study the effects of passage of a medical marijuana law (MML) on these direct payments to physicians. To draw causal conclusions, we develop a novel penalized synthetic control (SC) method that accommodates the zero-payment related latent structures inherent in these payments. Under a truncated flexible additive mixture model, we show that the SC method has uncontrolled maximal risk without the penalty; by contrast, the proposed penalized method provides efficient estimates. Our analysis finds a significant decrease in direct payments from opioid manufacturers to pain medicine physicians as an effect of MML passage. We provide evidence that this decrease is due to the availability of medical marijuana as a substitute. Finally, our heterogeneity analyses indicate that the decrease in direct payments are comparatively higher for female physicians and in localities with higher white, less affluent, and more working-age populations.

Keywords: Access to medication; average treatment effect; latent structure; pain management; penalized estimation

1 Introduction

Opioids are a class of drugs used to reduce pain. Opioids can be prescribed by physicians to treat moderate to severe pain but may also involve serious risks and side effects. Misuse and overuse of opioids have led to significant increase in opioid addictions and deaths. Opioid overdose-related deaths in the US rose from $21,088$ in 2010 to $68,630$ in 2020 (NIDA, 2022). As such, opioid consumption and its effects are highly debated objects in the current public discourse as well as a topic of vibrant academic research (Blanco et al., 2007, Cohn and Zubizarreta, 2022, Jacobs et al., 2022, Nam et al., 2020, Neuman et al., 2020, Prochaska et al., 2021, Zhang et al., 2020).

We consider two notable consequences in light of the opioid epidemic. First, advocacy of marijuana as a substitute for opioids gained traction (Cooper et al., 2018, Geluardi, 2016, Hollenbeck and Uetake,

[1]B. Karmakar is an Assistant Professor in the Department of Statistics, University of Florida, G. Mukherjee is an Associate Professor in the Department of Data Sciences & Operations, University of Southern California and W. Kar is an Assistant Professor of Purdue University. Corresponding author: BK, 102 Griffin-Floyd Hall, University of Florida, Gainsville, Florida, 32611, bkarmakar@ufl.edu. September 6, 2023

2021), arguing for its effectiveness with respect to opioids, both as a painkiller as well as in lowering the chances of addiction and overdose death than opioids (NIDA, 2021). Many states have legalized medical consumption of marijuana, which in part is aimed at reducing opioid-induced harm (Bachhuber et al., 2014, Powell et al., 2018, Shi, 2017). However, there is limited medical congruence regarding the efficacy of marijuana in treating acute and chronic pain, and the Food and Drug Administration (FDA) has advocated for more clinical studies before it approves marijuana for pain management (FDA, 2020). Second, opioid manufacturers are increasingly spending more into research and development to make opioids safer, e.g., by including an abuse-deterrent formulation (Evans et al., 2019, FDA, 2015). However, an increased adoption of marijuana could lead to opioid being a niche product or, in the extreme, could lead to severely diminished usage of opioids (Feinberg, 2019, Szalavitz, 2023). Thus, in response to marijuana's entry into pain management, opioid manufacturers are likely to adjust their push-marketing strategies to interact with physicians (Levy et al., 1983, Scherer, 1980). One of the most common practices to facilitate such interactions in pharmaceuticals is through direct payments to physicians from opioid manufacturers (Jones and Ornstein, 2016, Schwartz and Woloshin, 2019). These direct payments may be in the form of consulting and speaker fees, conference travel reimbursements, or meal vouchers.

In this paper, we study the effects of legalization of medical marijuana on these direct payments made by opioid manufacturers to opioid-prescribing physicians. In 2021, the direct payments to physicians made by US pharmaceutical companies amounted to $10.88 billion.[2] Some stakeholders in this ecosystem, who have justified these payments, have argued that these payments serve as a conduit to engage with physicians and foster collaboration (Donohue et al., 2007, Korenstein et al., 2010, Rosenbaum, 2015). However, these payment practices have been historically found to have caused biased endorsement of manufacturers' drugs by the payment-receiving physicians (Carey et al., 2021, DeJong et al., 2016, Jones and Ornstein, 2016) and also contributed to higher health care costs (CMS, 2013). Given the potential impact of these payments, it is of societal interest to study how a law affecting a critical domain, such as pain management, impacts these payment-to-physician strategies.

In the context of this research, we study the impact that passage of marijuana legalization laws (MML) in different US states had on the opioid ecosystem by analyzing the changes in these direct payments to

[2]Based on OpenPayment data from CMS: https://openpaymentsdata.cms.gov/summary

opioid prescribers over time. To derive causal conclusions, we follow the popular synthetic control (SC) method (Abadie et al., 2010, Abadie and Gardeazabal, 2003). The widely used SC criterion of Abadie et al. (2010) cannot be directly applied in our context due to an idiosyncratic nature of the physicians' payment data, which we describe in detail later. To provide consistent inference we develop a novel penalized SC method akin to Abadie et al. (2015) and Ben-Michael et al. (2021b). In addition to estimating the overall effect of MML on payments to opioid-prescribing physicians, we also explore how this effect varies across physician specialties, experience levels, gender, and the communities they serve.

1.1 Causal Study of Marijuana Legalization Effects on Direct Payments to Physicians by Opioid Manufacturers

We study whether the passage of a law legalizing medical marijuana consumption (MML) affects direct payments from opioid manufacturers to physicians. These payments are part of the traditional push marketing strategies employed by pharmaceutical companies (Levy et al., 1983), strategically aimed at physicians based on patient demographics and prescription preferences (Angell, 2018, Schwartz and Woloshin, 2019). In MML states, where physicians can recommend medical marijuana for pain relief (Black, 2022), marijuana emerges as an opioid substitute. Subsequently, opioid manufacturers may adjust these payments in response to this new competition, potentially impacting the pain management ecosystem significantly.

Our research explores how direct payments from opioid manufacturers to physicians change in the states where a law legalizing medical marijuana was passed. We use a synthetic control method to match a physician from a state with MML, on payments they received before MML, to physicians in states without MML. Synthetic control methods are suitable for panel data because they provide causally interpretable estimates of post-treatment effect over time under appropriate assumptions; see details in Abadie et al. (2015) and Ben-Michael et al. (2021b). These methods are further appropriate for us since an MML passage is a staggered treatment; consequently, there are distinguished pre and post-treatment periods. However, as we discuss below, some care in using SC methods is warranted for our study.

Briefly, the synthetic control (SC) method is used with panel data where it (Abadie et al., 2010, Abadie and Gardeazabal, 2003) fits the pre-treatment observations of a target treated unit using a convex combi-

nation of the pre-treatment observations of the control units, which is called the synthetic control (SC) unit for this treated unit. The post-treatment outcome of the SC unit estimates the unobserved counterfactual post-treatment outcome of the target unit. Most applications of the SC method in the literature have used aggregated units, e.g., states and countries, as their study units. Aggregated observations average over latent patterns in the finer units' data and retain the common factors that might have different loadings in different units. Intuitively, in aggregated data, weights in the SC methods attempt to equate the factor loading of the target unit to the weighted average of the loadings of the control units.

Our direct payments data are available at the physician-level and have latent patterns. Specifically, these payments received by a physician are typically discontinuous, with significant periods of time when no payments are made to the physician. Nonetheless, controlling for these no-payment periods in our estimation is critical. For example, consider a physician in the treated group who received payment periodically, every six months, while many control physicians received payments every five months. For such data (Abadie et al., 2015) note that in direct use of the SC method, "interpolation biases may be severe [when] the donor pool contains units with characteristics that are very different from those of the unit representing the case of interest." The usual SC method that looks at the overall fit of the payments received by the target physician (in our example), will likely give positive weights to control physicians who received payments every five months. This will lead to biased estimates of post-treatment counterfactual outcomes. As a possible remedy to this bias, Abadie et al. (2015) recommend "restricting the donor pool to units that are similar to the [target unit]." For our study, we customize the synthetic control method (Abadie et al., 2010, Abadie and Gardeazabal, 2003) to account for varied zero-payments patterns in physician payments while matching as these patterns are latent.

Using detailed physician-level data, we study the heterogeneity in the effect based on physicians' specialties and their genders. We also study the heterogeneity in the said effects based on income, age, and racial composition of the respective patient communities these physicians serve by combining demographic and socioeconomic zip code data to each physician's area of practice. Our contributions are outlined below.

1. We develop a novel penalized synthetic control method to accommodate the zero-payment related idiosyncrasies of our physician payments data set. Most physicians' payment histories contain

instances of no payments, which do not allow direct application of the widely used synthetic control (SC) method of Abadie et al. (2010). Motivated by penalized SC (PSC) approaches suggested in Abadie et al. (2015), Ben-Michael et al. (2021a), we develop a novel penalty that can prevent interpolation biases and can capture the varied patterns of non-payments in the pre-treatment period. The proposed penalty involves two parameters λ and γ (defined in Sec. 3.2). While γ is associated with a pooling-penalty akin to Ben-Michael et al. (2021b), λ involves a new penalty that is designed to adjust for different patterns of non-payments.

2. We explain the role of the penalty and the working principle behind the developed PSC method in a truncated flexible additive mixture model that consists of a latent factor model and a mixture process. The model is more complex than the models for which operating characteristics of SC methods have been studied in the existing literature (Abadie et al., 2010, Ben-Michael et al., 2021a,b). The truncation is for non-negative payments and the mixture accommodates varying patterns of zero-payments among the physicians. In Section 3.3, we rigorously explain how the proposed penalty produces efficient SC estimates by accurately learning the factor model coefficients as well as mixture group memberships (see Theorem 1). Further, we illustrate the necessity of the penalty by showing that unpenalized SC method will have uncontrolled maximal risk in the concerned additive mixture models (see Lemma 3). These results may be of independent interest in understanding the role of SC methods in mixture models.

3. We analyze the impact on pain-medicine physicians' direct payments using our PSC method. Quarterly, 5%-15% of physicians had no payments. In the pre-treatment period, physicians in MML states (treated) and non-MML states (control) had no payments on an average of 0.99 and 1.04 quarters respectively (see Section 2). Our penalized SC method effectively matches physicians with synthetic counterparts during this period. Assuming the validity of the proposed synthetic control method, we find a statistically significant payment decrease due to MML passage.

4. We stress-test the effect of MML passage by examining a potential substitution mechanism. First, we identify a consistent effect in Florida, despite Florida passing MML two quarters after other treated states in our main analysis. Second, for Anesthesiologists, who are less prone to shifting

from opioids to marijuana, we observe an initial negative post-MML effect that subsequently levels to a non-significant impact. Third, we find a negative correlation between increased marijuana patient registration and opioid-prescribing physician payments.

5. Finally, we investigate the variability in the MML effect on payments across different subgroups. This heterogeneity analysis uses the estimated individualized treatment effect of the pain-medicine physicians. The effect varies between areas with comparatively higher white and black populations and seems more substantial in areas with more lower-income and working-age populations. Additionally, the effects of MML passage are lower in magnitude on female physicians.

1.2 Organization of the paper

Section 2 describes our data. We develop our PSC method aimed at varied zero-payment patterns and study its theoretical properties in Section 3. Section 4 presents simulation experiments comparing our PSC method with existing methods. Section 5.1 presents the primary analyses for pain-medicine physicians. Section 5.2 provides the mechanism analysis. Section 6 probes the heterogeneity in effects across physicians' gender, experience, and demographics of their patient communities. We conclude with additional discussion in Section 7. The supplement includes proofs and additional results.

2 Data Description

To meet our research goals, we needed access to the details on the direct payments from opioid manufacturers to prescribing physicians. These payments, although endogenously decided by each opioid manufacturer, are now legally mandated to be reported under the "Sunshine Act" (Richardson et al., 2014). This law was a federal response to address concerns over possible conflicts of interest, potential treatment bias, and healthcare costs (Carey et al., 2021, DeJong et al., 2016, Engelberg et al., 2014, Jones and Ornstein, 2016). Data became publicly available in September 2014, including payment amounts between physicians and manufacturers, drugs promoted through these payments, and payment dates.

We aggregated the payment information for each physician in our treated and control states for each of the 16 quarters from 2014 to 2017. No states passed an MML in 2015, and in 2016, six states passed an MML: Pennsylvania (PA), Ohio (OH), North Dakota (ND), Louisiana (LA), Florida (FL) and Arkansas (AR). We excluded the two small states, ND and AR, which had less than three eligible physicians for our

primary analysis. Three out of these four states, PA, OH, and LA passed an MML in the second quarter of 2016, while FL passed the law in the last quarter of 2016. This gap of over a quarter in Florida's MML passage could sway attitudes of physicians and patients on marijuana adoption, based on outcomes seen in PA, OH, and LA. FL policymakers and voters might also have been influenced by these states' MML passage, posing potential confounding concerns. Therefore, to prevent this confounding bias, we analyze FL separately from PA, OH, and LA. We use 10 control states that did not pass an MML till 2016 Q2.[3]

For each payment made to a physician, the data indicates the drug category promoted during the interaction, with up to five drugs listed per entry. To isolate opioid-related payments, we first flagged payments mentioning "pain," and subsequently retained those payments that mentioned opioids among the promoted drugs. We focused only on opioid manufacturer-drug combinations with payments in our pre-MML (pre-treatment) period (2014–2016). This led to 15 opioid brands from 5 manufacturers.[4] Our analysis precisely examines how MML impacts payments to physicians from these 15 opioid brands.[5]

A single payment can involve both opioids and non-opioids. As there was no logical way to allocate a fraction of the payment solely to opioids, to be conservative, we deemed any transaction as opioid-related if one or more opioids were mentioned. Figure 1 depicts the distribution of opioid proportion promoted in each payment and its corresponding average amount. The graph highlights two prominent payment types: payments promoting a single opioid, and payments involving two drugs, one or both of which could be opioids. Notably, payments featuring a single opioid tend to be associated with higher amounts.

Figure 2 outlines physician payments in the states under our study. While not identical, the payments exhibit similar patterns between treated and control states pre-MML. These payments, however, differ based on physician specialties. In 2015, the year before our treatment, 'Anesthesiologists' received roughly 30% of payments by dollar value, with 'Pain Medicine' physicians following at around 19%. Anesthesiologists and pain medicine physicians likely prescribe opioids for different purposes. Medi-

[3]These are: VA, NC, IN, GA, TX, WI, NE, SC, UT, AL. Twenty states had not passed an MML till 2016 Q2. Our control states were selected from them first, based on their geographical proximity to the treated states. Then, because of the population and economy sizes relative to our treated states, we included TX as a control state and removed smaller states TN and WV. The scarcity in just 20 donor states also leads to poor balance in synthetic control for state-level aggregated data. The quarter-year time scale also minimizes treatment anticipation bias in our analysis.

[4]Different dosages of the same drug are considered as a single opioid brand.

[5]We anticipate minimal interference from new opioid brand introductions in our post-MML analysis period, as only one brand from a new manufacturer was introduced in 2017, our post-MML period. (For more details, refer to Supplement S3).

Table 1: Percent of physicians by the number of quarters with zero payments

Number of quarters	0	1	2	3	4	5	6	Total
Control states	50	26	8	7	6	3	0	100
Treated states	53	23	10	9	2	2	2	100

cal pain management involves various specialists for treating chronic pain conditions, including family physicians, internal medicine physicians, and psychiatrists (Chou et al., 2009, INTEGRIS, 2020). Anesthesiologists, on the other hand, specialize not only in treating chronic pain but also in acute pain, and may use advanced intravenous techniques, particularly in peri-operative settings (INTEGRIS, 2020). Despite marijuana's growing use for chronic pain, its effectiveness in acute pain is limited (Corliss, 2022). With FDA yet to endorse marijuana for pain (FDA, 2020), anesthesiologists cannot use marijuana in intravenous acute pain treatment. Consequently, if medical marijuana were to work as a substitute for opioid in pain management, we are likely to see a more pronounced effect of MML passage on direct payments to pain medicine physicians than anesthesiologists. Therefore, we primarily study pain medicine physicians to explore the causal effect of MML passage on direct payments, and subsequently include the effect on anesthesiologists as part of the mechanism analysis behind the causal effect.

Combining anesthesiologists and pain medicine physicians, our analysis had 138 and 356 physicians from the four treated and ten control states, respectively, after removing 11 and 28 physicians from the treatment states and control states for extreme and irregular values. Figure 2 shows that in each quarter, 5%–15% of the physicians had no payments. Physicians in the treated and control states had zero payments on an average of 0.99 and 1.04 quarters, respectively, between 2014 Q1 and 2016 Q2. An incidence of zero payment during a period between an opioid manufacturer and a physician is informative about the latent behaviors of both parties. Clearly, the latent behaviors vary across physicians. Thus, our method, described in the next section, includes an additional penalty to closely match these zero-payment related latent patterns for a physician and its synthetic counterpart.[6]

We supplement the payments data with the corresponding prescription data for each physician from

[6]We could have avoided this technical challenge by aggregating the payments data at the state level. However, with only 20 control possible control states for 4 treated states and 10 pre-treatment time periods to match for, we found that the synthetic control method does not give good matches.

the Medicare Part D Prescriber Public Use File.[7,8] To calculate the number of opioid-related prescriptions, we separated the opioid and non-opioid drugs prescribed by the pain-medicine physicians. Figure 3 shows the yearly average opioids related payments and number of opioids related prescriptions. The figure shows a decrease in payments, while the average number of prescriptions increase marginally, although not significantly, from 2015 to 2017. Later, we look at a difference-in-differences comparison for opioid vs non-opioid prescription patterns for the treated and control states across the years.

We use additional data for further analysis of the heterogeneity in the effects on direct payments due to an MML passage. Supplementary analyses involve zip-code level data on demographics and income characteristics from the US Census Bureau's American Community Survey. We also use data on physicians' experience, gender, and practice size. Further, we analyze longitudinal data on medical marijuana patients in Florida post-MML. Section 6 gives additional information on these datasets.

3 Methodology

3.1 Set-up and notations

Let b be arbitrary unit that received treatment. The set $\mathcal{C}$ of all control units is indexed by $c = 1, \ldots, C$. We observe payments y_{ct}, $c = 1, \ldots, C$ and $t = 1, \ldots, T$ received by units in $\mathcal{C}$. For simplicity, assume that the treatment was applied between time $T-1$ and T. We observe payments $y_{bt}, t = 1, \ldots, T-1$ received by unit b in the pre-treatment period and the payment $\check{y}_{bT}$ received by unit b post-treatment. Noting that unit b would have received y_{bT} if they were not in the treatment set, the treatment effect is given by $\text{TE}_b = \check{y}_{bT} - y_{bT}$.

Now, if $\mathcal{B}$ be a set of treated units indexed by $b = C+1, \ldots, C+B$, the average treatment effect on the treated (ATT) over the set $\mathcal{B}$ is given by $\text{ATT}_{\mathcal{B}} = B^{-1} \sum_{b=C+1}^{C+B} \text{TE}_b$. Our goal is to estimate $\text{ATT}_{\mathcal{B}}$ as well as the subgroup average treatment effect $\text{ATT}_{\mathcal{A}} = |\mathcal{A}|^{-1} \sum_{b \in \mathcal{A}} \text{TE}_b$ over various interesting subsets of $\mathcal{A} \subseteq \mathcal{B}$, where $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$. For that purpose we next develop a synthetic control method to estimate the unknown y_{bT} for each $b \in \mathcal{B}$. The estimates $\hat{y}_{bT}$ are then used to estimate $\text{ATT}_{\mathcal{A}}$

[7]The Part D Prescriber PUF is from CMS's Chronic Conditions Data Warehouse, which contains Prescription Drug Event records submitted by Medicare Advantage and stand-alone Prescription Drug Plans (https://www.cms.gov/Research-Statistics-Data-andSystems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/ PartD2013.html).

[8]Unlike the detailed payments dataset, the prescription dataset only gives yearly aggregated information per physician.

by $\widehat{\text{ATT}}_{\mathcal{A}} = |\mathcal{A}|^{-1} \sum_{b \in \mathcal{A}} (\check{y}_{bT} - \hat{y}_{bT})$.

3.2 Proposed Penalized Synthetic Control Method

For unit $b \in \mathcal{B}$, we estimate y_{bT} by using the synthetic control (SC) method (Abadie, 2021, Abadie et al., 2010, Abadie and Gardeazabal, 2003) that prescribes estimating y_{bT} by linearly aggregating the payments received by the controls $\hat{y}_{bT} = \sum_{c=1}^{C} w_{bc} y_{cT}$ where the weights $w_{bc} \geq 0$ and $\sum_{c=1}^{C} w_{bc} = 1$ for all $b \in \mathcal{B}$. Let $\boldsymbol{w}_b$ be the C dimensional vector $(w_{b1}, \ldots, w_{bC})$ and W denote the $B \times C$ matrix whose row b is $\boldsymbol{w}_b'$. Define

$$f(W; \lambda, \nu) = \frac{1}{B} \sum_{b \in \mathcal{B}} \left[\sum_{t=1}^{T-1} \left(y_{bt} - \sum_{c=1}^{C} w_{bc}\, y_{ct} \right)^2 + \sum_{c=1}^{C} w_{bc} \exp\left\{ \lambda \left(\sum_{t=1}^{T-1} \left(y_{bt} + y_{ct} \right) I\{\, y_{bt} \cdot y_{ct} = 0 \} \right) \right\} \right] + \nu \sum_{t=1}^{T-1} \left\{ \frac{1}{B} \sum_{b \in \mathcal{B}} y_{bt} - \sum_{c=1}^{C} \left(\frac{1}{B} \sum_{b \in \mathcal{B}} w_{bc} \right) y_{ct} \right\}^2, \quad (1)$$

where $I\{\}$ denotes the indicator function. For any fixed $\lambda, \nu \geq 0$ consider the following minimization:

$$\arg\min_{W} f(W; \lambda, \nu) \text{ such that } \boldsymbol{w}_b \geq 0 \text{ and } ||\boldsymbol{w}_b||_1 = 1 \text{ for all } 1 \leq b \leq B. \quad (2)$$

The objective criterion produces a penalized synthetic control (PSC) estimator. Penalized synthetic controls are increasingly being used (Abadie, 2021, Abadie et al., 2015, Ben-Michael et al., 2021a) to incorporate relevant structural constraints particularly while dealing with disaggregate level data. See Section 1 of Ben-Michael et al. (2021a) for a comprehensive review on usages of penalized synthetic controls. Here, we have two penalty parameters λ and ν which imparts two different types of regularization on the estimators. We next elaborate on the motivation behind (1) and the role of the penalization parameters.

We are interested in not only estimating the average treatment effect on the treated $\text{ATT}_{\mathcal{B}}$ over different concerned subsets of physicians $\mathcal{B}$ but also in studying the heterogeneity among the individual treatment effects TE_b. For the first goal, it is best to use pooled SC based criterion that minimizes the average pre-treatment imbalance across members in $\mathcal{B}$. However, for the second goal it is optimal to use separate SC criterion which estimates weights by separately minimizing the pre-treatment imbalance for each treated unit $b \in \mathcal{B}$. The estimators from the pooled SC and the separate SC based criteria often significantly disagree and subsequently producing highly sub-optimal inference in either one of the two

goals. Partially pooled SC (Ben-Michael et al., 2021b) provides a framework for construction of SC estimator whose risk can be simultaneously well-controlled in both the aforementioned inferential goals. We consider a partially pooled SC framework. The ν hyper-parameter in (1) balances the sum of squared imbalances (Im) from the individual SC and the pooled SC criteria. As such, note that the objective criterion minimized here is the sum of three components. Denote the three terms in (1) respectively by

(a) $\text{Im}_{\texttt{sep}}$, which is the sum of squared pre-treatment imbalances for each separate treated units,

(b) $\text{Pen}_{\texttt{sep}}(\lambda)$, which is an additive penalty that is separable across treated units, and

(c) $\text{Im}_{\texttt{pool}}$, which is the sum of squared pre-treatment imbalances for the average payment in $\mathcal{B}$.

Thus, we have: $f(W;\lambda,\nu) = \text{Im}_{\texttt{sep}} + \text{Pen}_{\texttt{sep}}(\lambda) + \nu\,\text{Im}_{\texttt{pool}}$. When $\nu = 0$, $f(W;\lambda,\nu)$ decouples into B separate unit-level minimization problems. Also, as $y_{it} \geq 0$ for all i and t in our data application, $\text{Pen}_{\texttt{sep}}(\lambda)$ is an increasing function of λ. At $\lambda = 0$, $\text{Pen}_{\texttt{sep}}(0) = 1$. When both $\lambda = 0$ and $\nu = 0$, $f(W;\lambda,\nu)$ is the canonical SC criterion prescribed in Abadie et al. (2010). When $\lambda = 0$ and $\nu > 0$, it is the partially pooled SC criterion where ν balances the separate unit level and pooled sum of squared imbalances between the treated unit and their synthetic controls in the pre-treatment period.

We develop and use the penalty $\text{Pen}_{\texttt{sep}}(\lambda)$ in (1) to prevent interpolation biases particularly when the control set is large and have highly heterogeneous members. Such uses of penalties in SC methods were suggested in Abadie et al. (2015) and later further developed in Ben-Michael et al. (2021b). However, $\text{Pen}_{\texttt{sep}}(\lambda)$ differs in fundamental aspects from penalties that have been prescribed in the existing literature on PSC. This is because we have developed $\text{Pen}_{\texttt{sep}}(\lambda)$ so that the resulting estimators are adaptive to the following important structural characteristics of the physicians' payment data set that we analyze here. This adaption in the proposed PSC method is crucial (explained later in Section 3.3) in controlling the error rates of the synthetic control based estimators of TE in this application.

While the observed payment y_{it} is non-negative, we witness (see Table 1 and Figure 2) non-significant proportion of zero-payments, i.e., $y_{it} = 0$. As the event of a zero-payment is intrinsically much different from the event of a positive payment, considering a uniform metric such as L_2 distance used in $\text{Im}_{\texttt{sep}}$ across all time points can lead to erroneous estimation. To mitigate the severe interpolation bias that can happen due to using sum of squared differences between treated and its estimates, we append the penalty $\text{Pen}_{\texttt{sep}}(\lambda)$ to the minimization criterion. A natural choice of penalty is the weighted L_1 distance

between the treat unit b and each of the control units: $\text{Pen}^{(\ell_1)}_{\text{sep}}(\lambda) = \lambda\{\sum_{c=1}^{C} w_{bc}\,(\sum_{t=1}^{T-1}|y_{bt} - y_{ct}|)\}$. The proposed penalty $\text{Pen}_{\text{sep}}(\lambda)$ differs from it by emphasizing the difference between the treated and control units in the occurrence of zero-payments. Unlike this L_1 penalty, the proposed penalty is not linear but exponential and it only considers the gaps between the treated and control units when one of them is zero and the other positive: $\text{Pen}_{\text{sep}}(\lambda) = \sum_{b=1}^{B}\sum_{c=1}^{C} w_{bc} \exp\left(\sum_{t=1}^{T-1}\{\lambda\, y_{ct} I(y_{bt} = 0) + \lambda\, y_{bt} I(y_{ct} = 0)\}\right)$. Table S1 in the supplement compares these two choices of penalties through a simulation study and shows that the proposed penalty is more suitable for our context.

Heuristically, the penalty helps in the construction of SC estimates by restricting estimates for treated unit b to only corresponding control units that have similar patterns of zero-payments; subsequently, the ν-weighted sum of separate and pooled imbalances are minimized producing SC estimates for any treated unit $b \in \mathcal{B}$ that (a) have controlled imbalances for positive y_{bt} in $t = 1, \ldots, T-1$, and (b) are based on control units $\mathcal{C}_b \subset \mathcal{C}$ such that $\sup_{c\in\mathcal{C}} y_{ct} \approx 0$ whenever $y_{bt} = 0$ for any $t = 1, \ldots, T$. We show below in Section 3.3 that not only the former but the second condition is also needed in our application to produce good estimates of y_{bT} for $b \in \mathcal{B}$. Thus, the role of the penalty is very important in (1). Next, we formally explain the role of the penalty function and then provide the implementation details for constructing the proposed PSC estimates in Section 3.4.

3.3 Risk properties and the role of the penalties

An additive mixture model. To study the risk properties of the proposed PSC estimators we consider a flexible additive mixture model. Readers interested in the implementation of the PSC method and our empirical study may skip ahead to Section 3.4.

Without loss of generality, consider y_{it} as truncated observations from unobserved pay-offs z_{it} that varies over $\mathbb{R}$, i.e., $y_{it} = \max(z_{it}, 0)$. Consider an additive model for the pay-offs:

$$z_{it} = f_{it} + \delta_{it} + \epsilon_{it}, \text{ for } i = 1, \ldots, C, C+1, \ldots, B+C \text{ and } t = 1, \ldots, T, \tag{3}$$

where, f_{it} is a low-dimensional factor model and ϵ_{it} are noise with $E(\epsilon_{it}) = 0$, $E(\epsilon_{it}^2) = \sigma^2$ and $E(\epsilon_{i_1t_1} \cdot \epsilon_{i_2t_2}) = 0$ whenever $i_1 \neq i_2$ or $t_1 \neq t_2$. Let $f_{it} = \sum_{k=1}^{K} \phi_{ki}\,\mu_{kt}$ be a K dimensional latent factor model as in Abadie et al. (2010), with the coefficient $\boldsymbol{\phi}_i = (\phi_{ik} : 1 \leq k \leq K)$ varies across units but is invariant across time, whereas the factor $\boldsymbol{\mu}_t = (\mu_{kt} : 1 \leq k \leq K)$ is invariant across units but varies across time.

For each i, let $\Delta_i = (\delta_{i1}, \ldots, \delta_{iT})'$ be a dampening sequence, i.e, $\Delta_i \leq \mathbf{0}$. If $\Delta_i = \mathbf{0}$ for all i in (3), and $T - 1 \gg K$, then for any treated unit $b \in \mathcal{B}$ the parameters $\boldsymbol{\phi}_b$ and $\{\mu_t : 1 \leq t \leq T\}$ can be well approximated leading to good SC based estimates of y_{bT} (see appendix B of Abadie et al., 2010 and the proof of Thm. 1 in Ben-Michael et al., 2021a).

When δ_{it} is highly negative for some t, it would dominate the other terms in (3) producing negative pay-off z_{it} and so, the observation y_{it} would be a zero-payment. Represent the support of the negative spikes of this dampening process by the vector $\boldsymbol{q}_i = I\{\Delta_i < \mathbf{0}\}$. There can be 2^T different types of $\boldsymbol{q}_i$s. However, for our application the $\boldsymbol{q}_i$s are not random sequences but are based on specific temporal patterns. As such, we can impose further constraints on the model and assume that there are only L different types of dampening sequences where L is an unknown but fixed number. The presence of such regularity structures among the zero-payment patterns is important for consistent estimation. Under this constraint, Δ_i is generated from a mixture model, i.e., $\Delta_i = \bar{\Delta}_{h(i)}$ where $h : \{1, \ldots, B + C\} \to \{1, \ldots, L\}$ is an unknown function that maps the units to groups containing on similar dampening sequences.

We observe $y_{it} = \max(z_{it}, 0)$ for $i \in \mathcal{B} \cup \mathcal{C}$ and $t = 1, \ldots, T - 1$ and $y_{iT} = \max(z_{it}, 0)$ for $i \in \mathcal{C}$ and the goal is to estimate $y_{iT} = \max(z_{bT}, 0)$ for $b \in \mathcal{B}$. The factors $\boldsymbol{\mu}_t$ in the factor model are global (invariant across units) whereas $\bar{\Delta}_h = (\bar{\delta}_{h(c),t} : 1 \leq t \leq T)'$ varies between groups with different dampening patterns. Compared to the latent factor model analyzed in Abadie et al. (2010), it is more challenging to construct efficient SC estimates in the presence of these complex, additive structures in Δs. See supplement for an illustrative example. Equipping (1) with the penalty $\text{Pen}_{\text{sep}}(\lambda)$ helps us in correctly learning the coefficients $\boldsymbol{\phi}_b$ for any treated unit b. Next, we explain the risk properties of the proposed PSC method in an asymptotic regime where $T \to \infty$ and λ is large. In practice, the tuning parameter λ is chosen by cross validation and the penalized criterion is seen to produce good estimates across varied non-asymptotic regimes which are presented later in Section 4.

Risk analysis of the proposed estimator. To facilitate a formal but intuitive understanding on the role of the aforementioned penalty for producing consistent SC based estimates, henceforth in this subsection, we assume that the noise component ϵ_{it} in (3) are generated from Gaussian distribution. All the results in this subsection can be easily extended to additive mixture models with sub-gaussian noise. To provide rigorous mathematical proofs of the risk properties, we consider $T \to \infty$ and impose the following

assumptions on (3):

A1. The factor model has significant signal strength. Let $\gamma = 2(\log C + \log T)^{1/2}\sigma$. Assume $f_* := \inf\{f_{it} : 1 \le i \le B + C,\ 1 \le t \le T\} \ge \gamma$.

A2. For any two dampening sequences $\bar{\Delta}_g, \bar{\Delta}_h$, if $\bar{\boldsymbol{q}}_g = I\{\bar{\Delta}_g < \mathbf{0}\}$ and $\bar{\boldsymbol{q}}_h = I\{\bar{\Delta}_h < \mathbf{0}\}$ are such that $\sum_{t=1}^{T-1} |\bar{q}_{g,t} - \bar{q}_{h,t}| = 0$ then $q_{q,T} = q_{h,T}$. This is a benign assumption that ensures that two distinct dampening sequences must differ at least once in the pre-treatment era. It is essential for identifiable estimate of y_{bT} in (3) based on observing y_{it} for $t = 1, \ldots, T-1$ and $i = 1, \ldots, B+C$.

A3. There is at least one instance where the dampening sequence has large enough negative signal to dominate the factor model. For any h, assume $\inf_{1 \le t < T} \bar{\delta}_{h,t} \le -f^* - \gamma$ where $f^* := \sup\{f_{it} : 1 \le i \le B+C,\ 1 \le t \le T\}$. Additionally, post intervention $\delta_{h,T}$ are either very large or well-controlled: $\delta_{h,T} \in (-\infty, -f^* - \gamma] \cup [-\sigma(\log T)^{1/2}, 0]$ for all h.

A4. Two distinct dampening sequences must differ significantly in at least one time point in the pre-treatment era, i.e., $\sup_{1 \le t < T} |\bar{\Delta}_{g,t} - \bar{\Delta}_{h,t}|\, I\{\bar{\Delta}_{g,t}\bar{\Delta}_{h,t} = 0\} \ge f^* + \gamma$. Note that, by assumption A3, if two distinct dampening sequences have disjoint supports, i.e., $\sum_{i=1}^{T-1} \bar{q}_{g,t}\, \bar{q}_{h,t} = 0$, then this condition is trivially satisfied.

A5. Each dampening sequence has a non-trivial fraction of zeros: $\inf_h \underline{\lim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} (1 - \bar{q}_{h,t}) > 0$. This implies each treated unit can have a non-trivial proportion of non-zero payments in (3). This is a benign assumption mainly set to prevent degeneracy in the proof. Estimates of any treated unit not satisfying the condition can be just set to 0. We further assume that for any treated unit b, the sum of squared imbalances based on controls with same dampening sequences is well-controlled:

$$\min_{\boldsymbol{w} \ge 0,\, ||\boldsymbol{w}||_1 = 1} \sum_{t=1}^{T} \Big(y_{bt} - \sum_{c \in \mathcal{C}_b} w_c\, y_{ct}\Big)^2 I\{\delta_{bt} = 0\} \le O(T \log T) \text{ as } T \to \infty, \tag{4}$$

where, $\mathcal{C}_b = \{1 \le c \le C : \Delta_c = \Delta_b\}$. This again is a very flexible assumption as the asymptotic behavior of the square error of imbalances from any reasonable model in the pre-treatment period is typically linear in T and we allow a poly-log margin over it. It holds as long as we have a sensible control set such that no group in (3) that has too few or no control units.

A6. Our final assumption is not on the model but on criterion (1). We restrict the weight corresponding to each control unit to be either 0 or at least $1/(CT)$. Let $\mathcal{W}$ be the set of all such weight vectors

which satisfy $\sum_{c=1}^{C} w_c = 1$ and $w_c = \{0\} \cup [(CT)^{-1}, 1]$ for all c. We assume (4) also holds for the reduced weight space $\mathcal{W}$.

Under these assumptions, we concentrate on estimating Y_{bT} where Y_{bT} is generated from (3) and $b \notin \mathcal{C}$. We consider estimators of the form $\hat{y}_{bT}(\boldsymbol{w}) = \sum_{c=1}^{C} w_c\, y_{cT}$ where weight $\boldsymbol{w} \in \mathcal{W}$. We concentrate on criterion (1) with $\nu = 0$. The effect of the pooling penalty parameter has been extensively studied in Ben-Michael et al. (2021a) and similar impact will be seen here. With $\nu = 0$, the optimization of (1) decouples in optimization for each treated unit separately. For constructing $\hat{Y}_{bT}$, consider only controls in the following subset of the control set $\mathcal{C}$:

$$\hat{\mathcal{C}}_b = \{c \in \mathcal{C} : y_{ct} \leq \psi^{-1} \text{ if } y_{bt} = 0 \text{ and } y_{ct} > 0 \text{ if } y_{bt} \geq \psi^{-1} \text{ for all } t = 1, \ldots, T-1\}, \tag{5}$$

for some $\psi > 0$. Note that, unlike $\mathcal{C}_b$ which depends on the model parameters, $\hat{\mathcal{C}}_b$ depends only on the observations. Next, consider the sequence of penalty parameter $\{\lambda_T : T \geq 1\}$ with $\lambda_T^2 = 2\psi(\log(CT) + \log\log T)$. The additional off-shoot term in the penalty akin to hard thresholding penalty in Donoho and Johnstone (1994). Lemma 1 shows that with very high probability for any treated unit b the PSC estimate based on λ_T is solely based on control units in $\hat{\mathcal{C}}_b$. As such, the probability is at least $1 - T^{-2}$ as $T \to \infty$.

Lemma 1. *Under assumptions A1–A6, for any treated unit b, the optimal weight vector $\hat{\boldsymbol{w}}^{(b)}$ for the minimization (2) with $\lambda \geq \lambda_T$ satisfies*

$$\lim_{T\to\infty} T^2\, P\big(\hat{w}_c^{(b)} \neq 0 \textit{ for some } c \in \mathcal{C} \setminus \hat{\mathcal{C}}_b\big) = 0.$$

The proofs of all the results stated in this section and additional discussion are given in the supplement. We next show that if we restrict ourselves to controls in $\hat{\mathcal{C}}_b$, then with very high probability, we will be considering only controls which has the same dampening sequence as the b^{th} treated unit.

Lemma 2. *Under assumptions A1–A6, for any $b \in \mathcal{B}$ and $\alpha < 1/2$*

$$\lim_{T\to\infty} T(\log T)^{\alpha}\, P\left(\sup_{c\in\mathcal{C}} \sum_{t=1}^{T} |\delta_{ct} - \delta_{bt}| \cdot I\{c \in \hat{\mathcal{C}}_b\} > 0 \right) = 0. \tag{6}$$

Equipped with the above the above results, we next establish a probabilistic upper-bound on the error of the proposed PSC estimator. It involves several steps. The essential part of the analysis is that when

$y_{bT} > 0$, then with very high probability, $y_{bT} - \hat{y}_{bT} = z_{bT} - \sum_c w_c z_{cT}$ for the proposed estimator, and we can concentrate on the estimation error for the non-truncated pay-offs from (3). For any weight $\boldsymbol{w}$, this error decomposes into three constituents corresponding to the factor model, the dampening sequence and the noise respectively:

$$R_f(\boldsymbol{w}) + R_\delta(\boldsymbol{w}) + R_\epsilon(\boldsymbol{w}), \text{ where, } R_f(\boldsymbol{w}) = \sum_{k=1}^{K} \mu_{kT}\big(\phi_{kb} - \sum_c w_c \phi_{kc}\big),$$
$$R_\delta(\boldsymbol{w}) = \sum_c w_c(\bar{\delta}_{h(b),T} - \bar{\delta}_{h(c),T}), \text{ and } R_\epsilon(\boldsymbol{w}) = \epsilon_{bT} - \sum_c w_c \epsilon_{cT}. \tag{7}$$

By lemma 2, R_δ is well-controlled for the proposed PSC estimator. Also, as any weight vector $\boldsymbol{w}$ that is trained on the pretreatment period as in (1) is independent of $\{\epsilon_{cT} : c = 1, \ldots, C\}$, $R_\epsilon(\boldsymbol{w})$ is stochastically dominated by $N(0, v)$ where, $v = \sigma^2(1 + ||\boldsymbol{w}||^2) \leq 2\sigma^2$ as $||\boldsymbol{w}_b||_1 = 1$. R_f in (7) can be well-controlled if the PSC method learns the factor model coefficients $\{\phi_{kb} : 1 \leq k \leq K\}$ pertaining to treatment b well, i.e., $\sum_c \hat{w}_{bc}\phi_{kc} \approx \phi_{kb}$. For any weight vector $\boldsymbol{w}$ define $\Phi(b; \boldsymbol{w}) = (\phi_{kb} - \sum_c w_c \phi_{kc} : 1 \leq k \leq K)$. Then, $|R_f(\boldsymbol{w})| \leq ||\boldsymbol{\mu}_T||_2 ||\Phi(b; \boldsymbol{w})||_2$. In the latent factor model where $\delta_{it} = 0$ for all i, t in (3), it follow directly from Appendix B of Abadie et al. (2010), that $||\Phi(b; \boldsymbol{w})||_2$ is upper bounded by a multiple of the imbalance between the treated unit and the PSC estimates in the pre-treatment period. The multiplier is proportional to the lowest eigenvalue of H where $H = \sum_{t=1}^{T-1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t'$. Using lemma 2 and applying similar derivations for SC estimates restricted to the class $\hat{\mathcal{C}}_b$ of controls, we obtain an analogous upper bound for $|R_f(\boldsymbol{w})|$. Combining these bounds on the three terms in (7) we arrive at our main result which provides an explicit upper-bound on the loss of the PSC estimator $\hat{y}_{bT}(\boldsymbol{w})$ for the bth treated unit based on weight $\boldsymbol{w}$. The bound depends on the sum of squared imbalances from positive time points $\text{Imp}(\boldsymbol{w}, b) = \sum_{t=1}^{T-1}(y_{bt} - \sum_{c=1}^{C} w_c y_{ct})^2 I\{y_{bt} > 0\}$ as well as on the factor model parameters in (3).

Theorem 1. *Under assumptions A1–A6, for any treated unit $b \in \mathcal{B}$ and for any weight $\boldsymbol{w} \in \mathcal{W}_b := \{\boldsymbol{w} \in \mathcal{W} : w_i = 0$ for $i \notin \hat{\mathcal{C}}_b\}$ and $\psi > 0$, we have,*

$$|y_{bT} - \hat{y}_{bT}(\boldsymbol{w})| \leq m_b^{-1/2} ||\boldsymbol{\mu}_T||_2 \left(\kappa_b \left\{ s_b^{-1} \textit{Imp}(\boldsymbol{w}, b) \right\}^{1/2} + 8\sigma \sqrt{s_b^{-1} \log T} \right) + 2\sigma \sqrt{\log T}, \tag{8}$$

with probability at least $1 - 1/T$ where m_b and κ_b are respectively the smallest eigenvalue and condition number of $s_b^{-1} \sum_{t=1}^{T-1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t' I\{y_{bt} > \psi^{-1}\}$ with $s_b = \sum_{t=1}^{T-1} I\{y_{bt} > \psi^{-1}\}$.

Note that in Theorem 1, $\mathcal{W}_b$ is the weight space with support concentrated on the control subset $\hat{\mathcal{C}}_b$. For moderately large T, the non-coverage probability of (8) is very small. Also, with K fixed as $T \to \infty$ when the factor loadings are well-regulated, we have $m_b = O(1), \kappa_b = O(1)$. Thus, in this case (8) gives

$$|y_{bT} - \hat{y}_{bT}(\boldsymbol{w})| \leq K||\boldsymbol{\mu}_T||_\infty \{s_b^{-1}\text{Imp}(\boldsymbol{w}, b)\}^{1/2} + 2\sigma\sqrt{\log T}. \tag{9}$$

By assumptions A5 and A6, the right side above for the optimal weighted PSC estimate is $O(\sqrt{\log T})$.

Now, to illustrate the importance of the penalty $\text{Pen}_{\texttt{sep}}(\lambda)$ we show that the SC estimator based on minimizing criterion (1) with $\lambda = 0$ have extremely high maximal risk as compared to the proposed PSC estimator. Consider the set Θ_T of all parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}_k,\ \phi_{ik},\ \bar{\Delta}_l : k = 1, \ldots, K; i = 1, \ldots, B + C; l = 1, \ldots, L)$ of (3) which along with assumptions A1–A4 also satisfy $\sup_k |\boldsymbol{\mu}_k| \leq \zeta$ for some prefixed $\zeta > 0$ and $\inf\{|\boldsymbol{\phi}_b - \boldsymbol{\phi}_c|_\infty : c \in \mathcal{C} \text{ and } \Delta_c = \Delta_b\} \leq \log T$. The following asymptotic result shows that with probability $1 - 1/T$, the worst case risk over Θ_T of the PSC estimate is $O(\sqrt{\log T})$ whereas the worse case risk of the SC estimate is higher than T.

Lemma 3. *Consider the PSC estimator $\hat{y}_{bT}(\hat{\boldsymbol{w}}_{psc})$ and the SC estimator $\hat{y}_{bT}(\hat{\boldsymbol{w}}_{sc})$ where the two weight vectors are selected by the minimization problem (2) with $\lambda \geq \lambda_T$ and with $\lambda = 0$ respectively. For any $a > 1/2$, there exists $C > 0$ such that,*

$$\lim_{T\to\infty} \min_{\boldsymbol{\theta}\in\Theta_T} T\left[1 - P_{\boldsymbol{\theta}}\left((\log T)^{-a}\left|y_{bT} - \hat{y}_{bT}(\hat{\boldsymbol{w}}_{psc})\right| < C\right)\right] = 0.$$

$$\lim_{T\to\infty} \max_{\boldsymbol{\theta}\in\Theta_T} T\left[1 - P_{\boldsymbol{\theta}}\left(T^{-1}\left|y_{bT} - \hat{y}_{bT}(\hat{\boldsymbol{w}}_{sc})\right| > C\right)\right] = 0.$$

3.4 Implementation of the method for analysis of physician payments data

We discuss the implementation details of the method, specifically our calculations of ν, λ and confidence interval. First, we follow Ben-Michael et al. (2021b)'s guide for the calculation of ν. We calculate W separately by minimizing $\text{Im}_{\texttt{sep}}$ and $\text{Im}_{\texttt{pool}}$. Then ν is set as $\sqrt{\text{Im}_{\texttt{sep}}}/(\sqrt{\text{Im}_{\texttt{pool}}} - \sqrt{\text{Im}_{\texttt{sep}}})$. Next, we use a cross-validation method to calculate λ. The cross-validation method leaves one of the pre-treatment time periods out at a time and fits the penalized synthetic control for a given λ. The λ is chosen as the one that minimizes the penalized partial sum of squared imbalance of the left-out time period to the synthetic control fit. Our simulation comparisons evaluate the proposed PSC where ν and λ are calculated this way.

In our payments data analysis, we analyze the average treatment effect on the treated (ATT) and the overall average treatment effect (ATE). The ATE calculation finds a synthetic control physician for each of the physicians from the states passing an MML as well as a synthetic counterpart physician from the pool of treated physicians for each of the physicians from the control states. We fit two W matrices for this purpose by solving two minimization problems, one finding a vector of weights of length equal to the number of control physicians for each treated state physician and the second finding a vector of weights of length equal to the number of treated physicians for each control state physician.

Finally, the confidence interval calculations use a leave-one-state-out calculation as in Rubinstein et al. (2021). For each state we create a new data set, leaving out all the physicians in that state, and calculate the ATE for the remaining physicians using our PSC method. We then estimate the standard error of our ATE by taking the squared root of the Jackknife variance formula ((17) of Rubinstein et al.). The calculation of standard errors for the ATT of FL leaves out control state physicians one state at a time. In an alternative approach, Keele et al. (2023) address correlated observations by estimating the correlation from the residuals from an outcome model fit. Our PSC method does not use an outcome model.

4 Simulation

We evaluate the relative performance of the proposed penalized synthetic control method to the existing methods. Our simulation model generates 30 units observed for $T = 55$ periods. Three of these 30 units are exposed to a treatment at time 45 and the rest 37 units remain unexposed. The original synthetic control method of Abadie and Gardeazabal (2003) and Abadie et al. (2010) is developed for a single exposed unit. This method is adapted when there are multiple exposed units by separately calculating the synthetic controls for each of the exposed units from the pool of all control units. Ben-Michael et al. (2021b) show that this method of separate calculations of the synthetic controls can be inefficient and propose a new method for simultaneous calculation of synthetic controls. We compare our method, with the proposed selection of the ν and λ, to these two state-of-the-art methods for synthetic control analysis.

In the notation introduced in the previous section, we generate data for unit i at time t as

$$y_{it} = \max(z_{it} + \tau_i W_{it}, 0); \text{ where } z_{it} = a_i + (t-1)/4 + \lfloor (t-1)/4 \rfloor + \delta_{it} + \epsilon_{it},$$

where a_i are iid uniform on $[10, 60]$, W_{it} are the treatment indicator which is 1 only when $t > 45$ and unit i is exposed, and τ_i is the treatment effect. The noise ϵ_{it} are independently drawn for each i, t from a normal distribution with mean 0 and standard deviation 5. We consider three clusters of the units and each cluster is specified by its units' common dampening sequence δ_{it}. How similar or different these dampening sequences are determines how similar or different these clusters are. One unit from each cluster is selected to be treated where the probability of treatment for unit i is proportional to $\sum_{t=20}^{45} y_{it}$.

We specified three models for the δ_{it}s in the three clusters in our simulations, varying the similarity of the clusters. We let $\delta_{it} = -80 \times q_{it}$ where q_{it} is 0 or 1, indicating the time when the process is dampened. The first two models are probabilistic and use exponential waiting time processes. Specifically, in cluster c, starting at time $t = 0$, after an exponential time with the rate θ_c a dampening starts. After that, the process is dampened for an exponential time length with the rate η_c. Following this, the first process restarts to find the next starting time for dampening. This model can be thought of as physicians and drug manufacturers following a similar exponential waiting process for deciding when they would interact.

Our first dampening model sets $\theta_c = 1/10$ for all $c = 1, 2, 3$ and $\eta_1 = \eta_2 = \eta_3 = 1/3$; the second model sets $\theta_1 = 1/10$, $\theta_2 = 1/15$, $\theta_3 = 1/7$ and $\eta_1 = \eta_2 = \eta_3 = 1/3$. The last model sets deterministic $q_i = 1$ for $i = 20, \ldots, 24, 40, \ldots, 44$ in cluster 1, for $i = 30, \ldots, 34, 45, \ldots, 49$ in cluster 2, and for $i = 40, \ldots, 44, 50, \ldots, 54$ in cluster 3. Figure 4 provides plots of data generated from these models.

Table 2: Simulation comparison for different synthetic control methods when $\tau_i = 0$ for all: best performance in each row is in bold. Results are based on averaging over 500 simulations; standard errors are in the parentheses

	Synthetic control	Pooled SC	Proposed Penalized SC
clusters are probabilistically similar			
l_2 Imbalance	19.14 (0.26)	5.50 (0.08)	**4.04 (0.03)**
RMSE for ITT	20.69 (0.36)	5.87 (0.09)	**5.35 (0.11)**
RMSE for ATT	13.16 (0.39)	3.61 (0.11)	**2.36 (0.11)**
clusters are probabilistically different			
l_2 Imbalance	19.40 (0.27)	5.58 (0.10)	**4.07 (0.03)**
RMSE for ITT	21.33 (0.37)	5.81 (0.10)	**5.81 (0.18)**
RMSE for ATT	14.03 (0.37)	3.63 (0.13)	**2.89 (0.16)**
clusters are deterministic and different			
l_2 Imbalance	16.65 (0.17)	6.61 (0.08)	**4.50 (0.02)**
RMSE for IIT	26.30 (0.31)	6.56 (0.08)	**5.38 (0.04)**
RMSE for ATT	13.89 (0.22)	4.01 (0.10)	**2.54 (0.05)**

The simulation results are summarized in Tables 2 and 3, which report three performance measures. The 'l_2 imbalance' is the average of the three Euclidean distances of the pre-treatment outcomes of the three treated units and their synthetic controls. The 'RMSE for ITT' is the average of simulation-based root mean squared errors for estimating the individual treatment effect τ_i for each of the three treated units. Finally, the 'RMSE for ATT' is the simulation-based root mean squared error for estimating the average treatment effect $\sum_i W_{iT}\tau_i/3$ of the three treated units.

Table 3: Simulation comparison for different synthetic control methods when $\tau_i = 15, 25$ and -10 in the three exposed units respectively: best performance in each row is in bold. Results are based on averaging over 500 simulations; standard errors are in the parentheses

	Synthetic control	Pooled SC	Proposed Penalized SC
clusters are probabilistically similar			
l_2 Imbalance	19.50 (0.26)	5.50 (0.08)	**4.07 (0.03)**
RMSE for ITT	21.18 (0.36)	5.80 (0.09)	**5.33 (0.11)**
RMSE for ATT	13.74 (0.39)	3.52 (0.11)	**2.42 (0.11)**
clusters are probabilistically different			
l_2 Imbalance	19.54 (0.27)	5.64 (0.10)	**4.07 (0.03)**
RMSE for ITT	21.24 (0.37)	5.96 (0.10)	**5.81 (0.18)**
RMSE for ATT	13.41 (0.37)	3.81 (0.13)	**2.81 (0.16)**
clusters are deterministic and different			
l_2 Imbalance	16.33 (0.17)	6.79 (0.08)	**4.53 (0.02)**
RMSE for IIT	25.88 (0.31)	6.67 (0.08)	**5.38 (0.04)**
RMSE for ATT	13.70 (0.22)	4.15 (0.10)	**2.55 (0.05)**

These simulation results show that the original synthetic control method adapted to this situation performs very poorly in all measures. Comparatively, the pooled SC method performs better than the original method. Still, the proposed method has the best performance among all the methods in better fit and estimation. Further, the performance of pooled SC becomes progressively worse with three structures of the clusters that create increasing distinctions between the latent structures in the clusters. By contrast, the proposed method provides consistently good performance across different cluster structures.

Figure S2 in the supplement, which shows that the RMSEs are lowest for the proposed λ over a grid of choices of λs, provides an empirical justification of the cross-validated choice of λ. As in Kern et al. (2016), we further conduct a simulation study "calibrated" to the physician payments data set to judge our method's performance in practice. Results reported in the supplement show that the proposed method performs better in this calibrated simulation study; see Figure S3.

5 Results

5.1 Synthetic control analysis of the MML passage on physician payments

Our primary analysis considers all pain medicine physicians from 13 states, of which three (PA, OH, LA) were 'treated' states that passed an MML in the second quarter of 2016. The method, described in Section 3.4, produces synthetic controls for each physician in the treated states using physicians in the control states, and likewise produces synthetic counterparts for each physician in the control states using physicians in the treated states.

The accompanying Figure 5 in its left panel shows the average of the differences in the payments of 190 pain-medicine physicians against their synthetic counterparts. The difference in payments is nearly 0, with a confidence interval between $-\$27.0$ and $\$0.3$, in the pre-treatment periods. Thus, the match provides a good fit, which is an important requirement to draw causal conclusions from the calculated differences during the post-treatment period (Abadie, 2021, Abadie et al., 2010). Recently, Parast et al. (2020) proposed a new type of measure that is similar to the commonly used standardized differences in matched studies for balance diagnosis of a synthetic control analysis. Using this measure, Figure S4 in the supplement shows that the synthetic counterparts are good matches for the target physicians.

Figure 5 shows the average treatment effect (ATE) of MML on payments to physicians, estimating the effect on physicians from treated as well as control states where an MML might be enacted in the future. In the left panel, a significant and negative ATE is observed for pain medicine physicians, indicating declining payments after MML passage. The negative impact on payments persists, except for a minor shift from Q1 to Q2 of 2017. This reflects evolving dynamics in the interaction between these physicians and opioid manufacturers. The estimated payment decrease is substantial, around $\$1217.41$ in Q3 2017.

Unlike the treated states mentioned earlier, Florida enacted an MML in the final quarter of 2016. We explore whether actions in other states (PA, OH, and LA) during the quarter before Florida's MML passage influenced payment activities in Florida. Such potential spillover effects from MML in other states in Q2 2016 could bias the estimated treatment effect for Florida's pain medicine physicians using the synthetic control method (Schuler et al., 2021). We conducted two analyses on Florida physician payments. In the first, we created synthetic controls from physicians in the 10 control states, considering

payment history up to Q2 2016. In the second, we used payment history up to Q4 2016. Figure 6 contrasts payments to Florida physicians with their synthetic controls from these two analyses. Estimates from the first analysis are similar to estimates from the earlier analysis shown in left panel of Figure 5. This analysis reveals a negative effect on Florida physicians' payments as early as Q3 and Q4 of 2016, possibly due to spillover from MML passage in other states. However, this spillover only accentuates the negative MML effect. Controlling for concurrent MML passage in other states in our second analysis, the significant negative MML effect on payments to Florida physicians persists. This suggests that opioid manufacturers could have anticipated marijuana legalization, thus reducing direct payments with physicians.

5.2 Mechanism

We now elaborate on the possible mechanism behind the declining payments to pain medicine physicians due to MML passage. We attribute that this decline in direct payments from opioid manufacturers to the evolution of marijuana as a superior substitute in states that have legalized medical marijuana. Scherer (1980) theorizes that upon entry of a substitute, existing competitors in the market respond in one of the following ways: by counterattacking (increased marketing spending), retreating (by cutting back on marketing spending), or remaining passive (no reaction) (Hanssens, 1980, Lambin et al., 1975). Gatignon et al. (1989) suggest that existing players may reduce marketing spending as a profit-maximizing move when they are unable to effectively counter new entrant(s) (Oxenfeldt and Moore, 1978). In our context, research shows a shift in consumption from opioids (exiting players) to marijuana (the new entrant) (Boehnke et al., 2019, Cooper et al., 2018, Hollenbeck and Uetake, 2021, NIDA, 2021). Thus, opioid manufacturers may curtail marketing spending where the substitution to marijuana is inevitable.

However, the declining opioid payments in the treated states can potentially happen due to factors other than MML passage. For example, some states might have stricter laws to curb opioid usage, such as stricter PDMPs or Pill Mill laws (Moyo et al., 2017). In those states, the pressure on physicians to move away from opioids can lead to opioid manufacturers strategically decreasing their presence and thus investing less on financial inducements to physicians. And if those states pass MML, it might be difficult to identify the substitution effect that can be attributed to MML passage separately from the effect induced by opioid-restricting policies of those states. However, such identification issue will arise

only if the MML and opioid-curbing policies were implemented simultaneously. The opioid-restricting policies in our treatment states were passed before 2016 (i.e., not simultaneously with MML), e.g., Ohio passed the MML in April 2016 while the pill mill law was passed in 2011. Thus, our analysis would be able to adjust for these systematic differences in the pre-treatment period. Therefore, we can cleanly attribute the decrease in payment to opioid physicians to substitution effect arising due to MML passage.

Among the potentially other factors that could explain the estimated decline, it could also be that the treated states were able to pass the MML because of potentially weak lobbying power of the opioids manufacturers in those states (Frances, 2021). This could have led to the opioid manufacturers selectively reducing their activities in the treated states and hence a spurious negative effect in our analysis is manifested. If this conjecture is valid, we would see a negative effect on payments to physicians agnostic of their specializations. To test such a theory, we also analyzed direct payments to anesthesiologists, who received the second highest payments after pain medicine physicians.[9]

Both specializations, pain medicine and anesthesiology, regularly prescribe opioid. Pain medicine physicians primarily address chronic pain, while anesthesiologists manage both chronic and acute pain, often necessitating intravenous interventions (INTEGRIS, 2020). Despite increasing medical marijuana acceptance, its lack of FDA approval and its limited efficacy in treating acute pain make it unsuitable for anesthesiologists' intravenous procedures (Corliss, 2022, FDA, 2020). Therefore, pain medicine physicians are more likely to switch to marijuana compared to anesthesiologists. Consequently, post-MML, opioid manufacturers may reduce the payments to pain medicine physicians in response to the evitable substitution to marijuana, but may not decrease payments to anesthesiologists. Alternatively, if reduced payments to pain medicine physicians were solely due to opioid manufacturers losing ground in blocking MML passage, then a similar payment decline would be seen for anesthesiologists. We analyzed payments to anesthesiologists in MML and non-MML states, akin to the analysis on pain medicine physician (left panel of Figure 5), using our synthetic control method. The corresponding ATE plot for anesthesiologist payments is shown in the right panel of Figure 5. Results indicate that though anesthesiologist payments initially dropped post-MML passage, they rebounded within a quarter to pre-treatment levels.

It is probable that there could be another mechanism that explains the observed effects. The decline in

[9] All the other specializations, e.g., internal medicine and family medicine, had less than 10% of opioids related payments.

Table 4: Regression analysis results in cities where physicians practice across Florida

Dependent Var.:	Log of Payment Rec.	Log of Payment Rec.	Log of Payment Rec.	Percent Change Mari. Patient
Lag of % Change in Mari. Patients	−0.3626. (0.1886)		−0.3701** (0.1882)	
Whether City has Marijuana Dispensary		−0.1435*** (0.0243)	−0.1439*** (0.0243)	
Log of Payment Rec.				-0.0001 (0.0001)
Fixed-Effects: Physician Specialty	Yes	Yes	Yes	Yes
S.E. type	Heteroskeda.-rob.	Heteroskedast.-rob.	Heteroskedast.-rob.	Heteroskedast.-rob.
Observations	8,467	8,467	8,467	9,453
R2	0.04408	0.04725	0.04757	0.00280
Within R2	0.00044	0.00375	0.00421	0.00001

payments might not directly result from perceived medical marijuana influence, but could be due to lower payments prompting reduced opioid prescriptions and subsequently an increase in medical marijuana recommendation. We aim to distinguish between these mechanisms using medical marijuana patient registration data. Florida's provides bi-weekly updates on medical marijuana activity in the state, detailing active dispensary locations and number of registered patients over time.[10] Figure 7 depicts a rolling average of physician payments linked to active dispensaries in their regions. By merging marijuana patient registration data with payment data, we performed two regressions. First, we regressed log of payment against lagged bi-weekly change in marijuana patients and whether the physician's city had a marijuana dispensary. Second, we regressed bi-weekly change in marijuana patient against lagged payment activity. The results reported in Table 4 show a significant negative correlation between an increase in marijuana patients in preceding period as well as presence of a marijuana dispensary with opioid-prescribing physician payment. However, no significant association exists between physician payment and subsequent period change in marijuana patients. These findings further support that the dominant factor in reducing payments to pain medicine physician post-MML is the substitution effect of marijuana.

6 Heterogeneity

In Section 5, we found the passage of MML led to reduced direct payments to pain medicine physicians from opioid manufacturers. We argued this might arise from manufacturers recognizing increased opioid-

[10]Information is available at https://knowthefactsmmj.com/2018/07/28/2017-ommu-updates-archive/

to-marijuana substitution, especially for chronic pain management. Exploring distinct data patterns can inspire further studies to extend understanding of this mechanism. Our synthetic control method offers insight into individual treatment effects (ITEs), i.e., how MML affects each physician's payments. Using ITEs, we conduct a secondary analysis to explore variations in treatment effects based on physician characteristics and practice demographics. Figures 8–10 display estimated ITEs for physicians in 13 states, averaged over four quarters following MML passages in PA, OH, and LA in 2016 Q2.

We first look at the two physician characteristics, namely gender and year of graduation, in Figure 8. While the ITEs for the year of graduation do not show any defined pattern, we find that the decreases in payments to male physicians are less pronounced than female physicians. Historically, empirical research involving physician care has under-studied female physicians (Kimball and Crouse, 2007). However, significant differences exist in practice patterns between male and female physicians. Research shows that female physicians are more patient-centric, more open to patient concerns, have longer visits, and ask more questions (Hall et al., 1994, Roter et al., 2002). Our current finding emphasizes that in the context of pain relief, there is significant heterogeneity in how the introduction of medical marijuana affects physicians based on gender and, consequently, the population they serve.

Analyzing the ITEs based on demographics of where the physicians practice, in Figure 9, we find that the payments show a greater decrease in low-income areas. Low income areas would have a higher proportion of people who engage in blue-collar jobs and, therefore, have more requirements for chronic pain management (Jacobsen et al., 2013). Also, poorer regions have shown higher instances of opioid misuse (Ghertner and Groves, 2018). Arguably, substituting into marijuana will help these communities (Compton et al., 2017). The recent governmental intervention to reduce opioid misuse has focused on these vulnerable communities (The White House, 2016, USDA, 2019). The observed larger decrease in payments to physicians practicing in lower income communities could be a manifestation of a combination of governmental efforts and societal awareness regarding the potential harmfulness of opioids.

In Figure 9, ITE patterns with respect to the median age of the population reveal that the decrease in payments is the largest in the regions with a median age between 30-40 years. Populations younger than this bracket will have lesser requirements for chronic pain management; therefore, opioid manufacturers might feel less threatened by the entry of marijuana and may not decrease payments significantly. How-

ever, regions with older populations, who are more prone to chronic pain, also show lesser decrease in payments to physicians, suggesting lesser threat to substitution to medical marijuana. Focusing on racial dispersion, in Figure 10, ITEs show a sharp decline for physicians practicing in populations with a higher percentage of blacks. Recent studies have shown Blacks/African-Americans are significantly less likely to be prescribed opioids for pain by medical providers than white patients (SAMHSA, 2020). In Figure 10, we also observe that areas above the national average in White population demonstrate a greater MML passage effect than those exceeding the national average in Black/African American population.[11]

Our heterogeneity analyses reveal decline in direct payments vary based on physicians' characteristics and demographics of where they practice, possibly due to inherent differences in physicians' and patients' substitution patterns. Motivated by our results, hypothesis-driven studies would be helpful to (a) establish whether decrease in payments to opioid prescribing physicians is causally related to substitution to medical marijuana and (b) subsequently identify the modifiers of the substitution effect on treating pain.

7 Discussion

In the wake of opioid epidemic in the US (Feinberg, 2019), several measures were instituted at the federal and state levels to regulate proper management of opioid consumption. Some states also passed laws legalizing medical marijuana consumption partly in response to the opioid epidemic. While adoption of medical marijuana is on the rise (Geluardi, 2016), the FDA notes that, to date, "[it] has not determined that cannabis is safe and effective for any particular disease or condition;"[12] leading to opioids continuing to be potent treatment for mitigating pain. Physicians are the primary gatekeepers for deciding medication to patients needing pain management. Our research links these three key factors: medical marijuana laws, opioid manufacturers, and their engagement with prescribing physicians. Specifically, we study how the introduction of medical marijuana as a transformative competitor impacts direct payments to physicians from opioid manufacturers. We show that MML leads opioid manufacturers to reduce direct payments to opioid-prescribing physicians. Our mechanism analyses attribute this change to a growing adoption of marijuana for pain management, suggesting that opioid manufacturers perceive marijuana as a superior

[11] By April, 2020 estimates, the demography of the United States has about 59.3% non-Hispanic white and 13.6% black/African American population. https://www.census.gov/quickfacts/fact/table/US/PST040221

[12] https://www.fda.gov/news-events/public-health-focus/fda-regulation-cannabis-and-cannabis-derived-products-including-cannabidiol-cbd

substitute and subsequently respond by reducing these direct payments (Gatignon et al., 1989).

While our research focuses on opioid manufacturers and physicians, it is worth considering MML's effect on patient pain management. To set the context, pain medicine physicians tend to prescribe more opioid than non-opioid medication. They typically prescribe a certain amount in 30 days' fills but also medication for the number of days of use. The annual prescription data (mentioned in Section 2) shows that, from 2015 to 2017, in the states not passing an MML, 30 days' fill of opioid vs non-opioid remained flat at a 1.38:1 ratio. However, in the states passing an MML, from 2015 to 2017, 30 days' fill as well as the number of days of prescription of opioid vs non-opioid decreased from a 1.57:1 ratio to a 1.52:1 ratio. In particular, the pattern of opioid vs non-opioid prescriptions did not change in the control states, while there was a relative decrease in opioid prescriptions in the MML states from 2015 to 2017. We leave further analysis of the possible effect of MML passage on patient care for future research.

Methodologically, we develop a novel penalized synthetic control method that estimates an average treatment effect from a longitudinal dataset on multiple treated and control individuals. Using the pooled synthetic control strategy, we create a synthetic counterpart of each treated and control unit by closely matching on the target unit's and their groups' average pre-treatment outcome history. Further, we use a novel penalty to adapt the resulting estimators to the latent groups in the data whose members have similar quarterly non-payment patterns. The penalty reduces interpolation bias by closely matching individuals and their synthetic counterparts on their non-payment patterns. Finally, under an additive mixture model appropriate for our study, we show that an unpenalized synthetic control method will have uncontrolled maximal risk while the proposed method produces efficient SC estimates. In the future, developing penalized synthetic control methods for more complex latent structures in the data will be useful.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature 59*(2), 391–425.

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 the American Statistical Association 105*(490), 493–505.

Abadie, A., A. Diamond, and J. Hainmueller (2015). Comparative politics and the synthetic control method. *American Journal of Political Science 59*(2), 495–510.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American economic review 93*(1), 113–132.

Angell, T. (2018). Senator calls out big pharma for opposing legal marijuana. Forbes.com. https://www.forbes.com/sites/tomangell/2018/02/23/senator-calls-out-big-pharma-for-opposing-legal-marijuana/?sh=6ba9e9161bac.

Bachhuber, M. A., B. Saloner, C. O. Cunningham, and C. L. Barry (2014). Medical cannabis laws and opioid analgesic overdose mortality in the united states, 1999-2010. *JAMA Internal Medicine 174*(10), 1668–1673.

Ben-Michael, E., A. Feller, and J. Rothstein (2021a). The augmented synthetic control method. *Journal of the American Statistical Association 116*(536), 1789–1803.

Ben-Michael, E., A. Feller, and J. Rothstein (2021b). Synthetic controls with staggered adoption. Technical report, National Bureau of Economic Research.

Black, L. (2022). Is it legal for doctors to prescribe medical cannabis? GoodRx Health. https://www.goodrx.com/classes/cannabinoids/can-doctors-prescribe-medical-marijuana.

Blanco, C., D. Alderson, E. Ogburn, B. F. Grant, E. V. Nunes, M. L. Hatzenbuehler, and D. S. Hasin (2007). Changes in the prevalence of non-medical prescription drug use and drug use disorders in the united states: 1991–1992 and 2001–2002. *Drug and Alcohol Dependence 90*(2-3), 252–260.

Boehnke, K. F., S. Gangopadhyay, D. J. Clauw, and R. L. Haffajee (2019). Qualifying conditions of medical cannabis license holders in the united states. *Health Affairs 38*(2), 295–302.

Carey, C., E. M. Lieber, and S. Miller (2021). Drug firms' payments and physicians' prescribing behavior in medicare part d. *Journal of Public Economics 197*, 104402.

Chou, R., G. J. Fanciullo, P. G. Fine, J. A. Adler, J. C. Ballantyne, P. Davies, M. I. Donovan, D. A. Fishbain, K. M. Foley, J. Fudin, et al. (2009). Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. *The journal of pain 10*(2), 113–130.

CMS (2013). Medicare, medicaid, children's health insurance programs; transparency reports and reporting of physician ownership or investment interests. https://www.federalregister.gov/documents/2013/02/08/2013-02572/medicare-medicaid-childrens-health-insurance-programs-transparency-reports-and-reporting-of.

Cohn, E. R. and J. R. Zubizarreta (2022). Profile matching for the generalization and personalization of causal inferences. *Epidemiology 33*(5), 678–688.

Compton, W. M., N. D. Volkow, and M. F. Lopez (2017). Medical marijuana laws and cannabis use: intersections of health and policy. *JAMA Psychiatry 74*(6), 559–560.

Cooper, Z. D., G. Bedi, D. Ramesh, R. Balter, S. D. Comer, and M. Haney (2018). Impact of co-administration of oxycodone and smoked cannabis on analgesia and abuse liability. *Neuropsychopharmacology 43*(10), 2046–2055.

Corliss, J. (2022). Does cannabis actually relieve pain — or is something else going on? https://www.health.harvard.edu. https://www.health.harvard.edu/blog/does-cannabis-actually-relieve-pain-or-is-something-else-going-on-202212082863.

DeJong, C., T. Aguilar, C.-W. Tseng, G. A. Lin, W. J. Boscardin, and R. A. Dudley (2016). Pharmaceutical industry–sponsored meals and physician prescribing patterns for medicare beneficiaries. *JAMA internal medicine 176*(8), 1114–1122.

Donoho, D. L. and I. M. Johnstone (1994). Minimax risk over lp-balls for lp-error. *Probability Theory and Related Fields 99*, 277–303.

Donohue, J. M., M. Cevasco, and M. B. Rosenthal (2007). A decade of direct-to-consumer advertising of prescription drugs. *New England Journal of Medicine 357*(7), 673–681.

Engelberg, J., C. A. Parsons, and N. Tefft (2014). Financial conflicts of interest in medicine. *Available at SSRN abstract no. 2297094*. doi:10.2139/ssrn.2297094.

Evans, W. N., E. M. Lieber, and P. Power (2019). How the reformulation of oxycontin ignited the heroin epidemic. *Review of Economics and Statistics 101*(1), 1–15.

FDA (2015). Abuse-deterrent opioids — evaluation and labeling. https://www.fda.gov/media/84819/download.

FDA (2020, October). Fda and cannabis: Research and drug approval process. https://www.fda.gov/news-events/public-health-focus/fda-and-cannabis-research-and-drug-approval-process.

Feinberg, J. (2019). Tackle the epidemic, not the opioids. *Nature 573*(7773), 165.

Frances, A. (2021). Opioid companies lobby against medical marijuana. Rehabs.com. https://rehabs.com/pro-talk/opioid-companies-lobby-against-medical-marijuana/.

Gatignon, H., E. Anderson, and K. Helsen (1989). Competitive reactions to market entry: Explaining interfirm differences. *Journal of Marketing Research 26*(1), 44–55.

Geluardi, J. (2016). *Cannabiz: The explosive rise of the medical marijuana industry*. Routledge.

Ghertner, R. and L. Groves (2018). The opioid crisis and economic opportunity: geographic and economic trends. *ASPE Research Brief*, 1–22.

Hall, J. A., J. T. Irish, D. L. Roter, C. M. Ehrlich, and L. H. Miller (1994). Gender in medical encounters: an analysis of physician and patient communication in a primary care setting. *Health Psychology 13*(5), 384.

Hanssens, D. M. (1980). Market response, competitive behavior, and time series analysis. *Journal of Marketing Research 17*(4), 470–485.

Hollenbeck, B. and K. Uetake (2021). Taxation and market power in the legal marijuana industry. *The RAND Journal of Economics 52*(3), 559–595.

INTEGRIS (2020). What does a pain management doctor do? https://integrisok.com/. https://integrisok.com/resources/on-your-health/2020/september/what-does-a-pain-management-doctor-do.

Jacobs, L. A., Z. Branson, C. G. Greeno, J. L. Skeem, and T. Labrum (2022). Community behavioral health service use and criminal recidivism of people with mental, substance use, and co-occurring disorders. *Psychiatric Services 73*(12), 1397–1400.

Jacobsen, H. B., A. Caban-Martinez, L. C. Onyebeke, G. Sorensen, J. T. Dennerlein, and S. E. Reme (2013). Construction workers struggle with a high prevalence of mental distress and this is associated with their pain and injuries. *Journal of Occupational and Environmental Medicine 55*(10), 1197.

Jones, R. G. and C. Ornstein (2016). Matching industry payments to medicare prescribing patterns: an analysis. *ProPublica*.

Keele, L., E. Ben-Michael, and L. Page (2023). Approximate balancing weights for clustered observational study designs. *arXiv preprint arXiv:2301.05275*.

Kern, H. L., E. A. Stuart, J. Hill, and D. P. Green (2016). Assessing methods for generalizing experimental impact estimates to target populations. *Journal of research on educational effectiveness 9*(1), 103–127.

Kimball, E. B. and B. J. Crouse (2007). Perspectives of female physicians practicing in rural wisconsin. *WMJ : official publication of the State Medical Society of Wisconsin 106*(5), 256.

Korenstein, D., S. Keyhani, and J. S. Ross (2010). Physician attitudes toward industry: a view across the specialties. *Archives of Surgery 145*(6), 570–577.

Lambin, J.-J., P. A. Naert, and A. Bultez (1975). Optimal marketing behavior in oligopoly. *European Economic Review 6*(2), 105–128.

Levy, M., J. Webster, and R. A. Kerin (1983). Formulating push marketing strategies: A method and application. *Journal of Marketing 47*(1), 25–34.

Moyo, P., L. Simoni-Wastila, B. A. Griffin, E. Onukwugha, D. Harrington, G. C. Alexander, and F. Palumbo (2017). Impact of prescription drug monitoring programs (pdmps) on opioid utilization among medicare beneficiaries in 10 us states. *Addiction 112*(10), 1784–1796.

Nam, Y. H., W. B. Bilker, F. J. DeMayo, M. D. Neuman, and S. Hennessy (2020). Incidence rates of and risk factors for opioid overdose in new users of prescription opioids among us medicaid enrollees: a cohort study. *Pharmacoepidemiology and Drug Safety 29*(8), 931–938.

Neuman, M. D., S. Hennessy, D. S. Small, C. Newcomb, L. Gaskins, C. M. Brensinger, D. N. Wijeysundera, B. T. Bateman, and H. Wunsch (2020). Drug enforcement agency 2014 hydrocodone rescheduling rule and opioid dispensing after surgery. *Anesthesiology 132*(5), 1151–1164.

NIDA (2021). Is marijuana safe and effective as medicine? Cannabis (Marijuana) Research Report, National Institute on Drug Abuse. https://nida.nih.gov/publications/research-reports/marijuana/marijuana-safe-effective-medicine.

NIDA (2022, Jun). Overdose death rates. National Institutes of Health, U.S. Department of Health and Human Services. https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates.

Oxenfeldt, A. R. and W. L. Moore (1978). Customer or competitor: which guideline for marketing? *Management Review 67*(8), 43–48.

Parast, L., P. Hunt, B. A. Griffin, and D. Powell (2020). When is a match sufficient? a score-based balance metric for the synthetic control method. *Journal of Causal Inference 8*(1), 209–228.

Powell, D., R. L. Pacula, and M. Jacobson (2018). Do medical marijuana laws reduce addictions and deaths related to pain killers? *Journal of health economics 58*, 29–42.

Prochaska, J. J., E. A. Vogel, A. Chieng, M. Kendra, M. Baiocchi, S. Pajarito, and A. Robinson (2021). A therapeutic relational agent for reducing problematic substance use (woebot): development and usability study. *Journal of Medical Internet Research 23*(3), e24850.

Richardson, E., R. Saver, R. Lott, et al. (2014). *The physician payments sunshine act*. Health Affairs. doi:10.1377/hpb20141002.272302.

Rosenbaum, L. (2015). Beyond moral outrage—weighing the trade-offs of coi regulation. *New England Journal of Medicine 372*(21), 2064–2068.

Roter, D. L., J. A. Hall, and Y. Aoki (2002). Physician gender effects in medical communication: a meta-analytic review. *JAMA 288*(6), 756–764.

Rubinstein, M., A. Haviland, and D. Choi (2021). Balancing weights for estimated region-level data: the effect of medicaid expansion on the uninsurance rate among states that did not expand medicaid. *arXiv e-prints*, arXiv–2105.

SAMHSA (2020). The opioid crisis and the black/african american population: An urgent issue. *HHS Publication No. PEP20-05-02-001*. Substance Abuse and Mental Health Services Administration.

Scherer, F. (1980). M. industrial market structure and economic performance.

Schuler, M. S., B. A. Griffin, M. Cerdá, E. E. McGinty, and E. A. Stuart (2021). Methodological challenges and proposed solutions for evaluating opioid policy effectiveness. *Health Services and Outcomes Research Methodology 21*(1), 21–41.

Schwartz, L. M. and S. Woloshin (2019, 01). Medical Marketing in the United States, 1997-2016. *JAMA 321*(1), 80–96.

Shi, Y. (2017). Medical marijuana policies and hospitalizations related to marijuana and opioid pain reliever. *Drug and alcohol dependence 173*, 144–150.

Szalavitz, M. (2023). 'Entire body is shaking': Why americans with chronic pain are dying. New York Times, Opinion, Guest Essay. https://www.nytimes.com/2023/01/03/opinion/chronic-pain-suicides.html.

The White House (2016). Fact sheet: President Obama proposes $1.1 billion in new funding to address the prescription opioid abuse and heroin use epidemic. https://obamawhitehouse.archives.gov/the-press-office/2016/02/02/president-obama-proposes-11-billion-new-funding-address-prescription.

USDA (2019). Rural community action guide: Building stronger healthy, drug-free rural communities. https://www.usda.gov/sites/default/files/documents/rural-community-action-guide.pdf.

Zhang, P., C.-W. Chiang, S. Quinney, M. Donneyong, B. Lu, L. F. Huang, and F. Cheng (2020). The concurrent initiation of medications is associated with discontinuation of buprenorphine treatment for opioid use disorder. *medRxiv*, 1–19. doi:10.1101/2020.01.15.20017715.

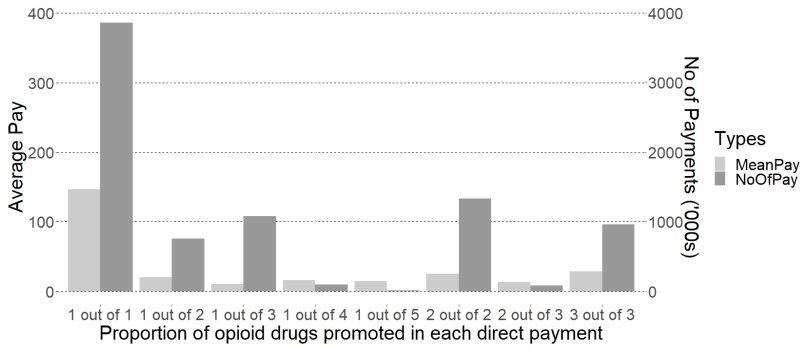

Figure 1: Distribution of average payment (in US dollars) and the number of payments related to opioid, categorized by the ratio of opioid to non-opioid drugs promoted during each payment.

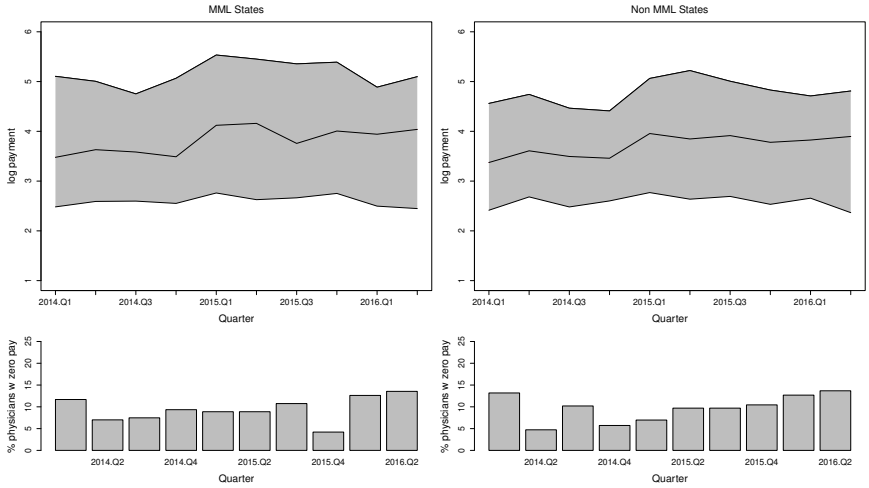

Figure 2: Summary of the payments to physicians in different quarters of pre-treatment period by the states that did and did not pass medical marijuana laws. The MML states are 'FL', 'LA', 'OH' and 'PA'; the non MML states are 'AL', 'GA', 'IN', 'NC', 'NE', 'SC', 'TX', 'UT', 'VA' and 'WI'. The plots on the top two panels show the 85th, 50th and 15th percentiles of log payments.

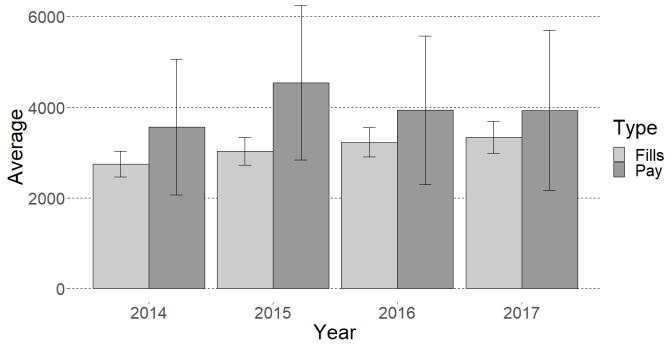

Figure 3: Distribution of average annual payments (in US dollars) to pain-medicine physicians and the corresponding average number of prescriptions (in '000s) written by those physicians across our analysis window (i.e., 2014-2017).

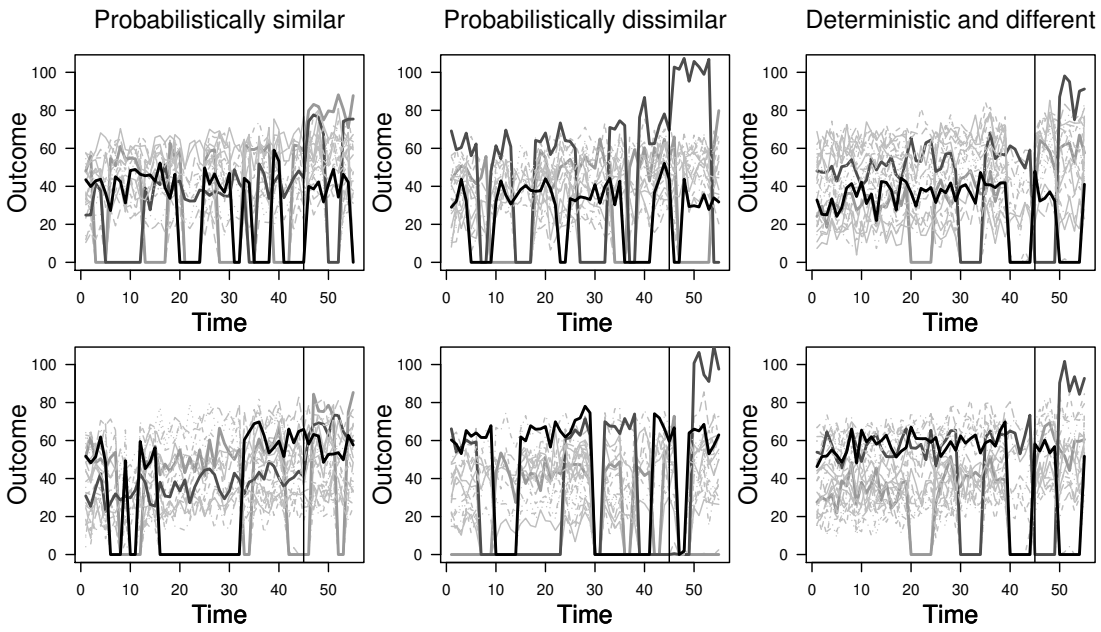

Figure 4: Two sets of simulated data in each the tree columns from the three simulation models. The three treated units from the three clusters are in colors 'black', 'dark gray' and 'gray' respectively; the vertical line shows the treatment adoption time.

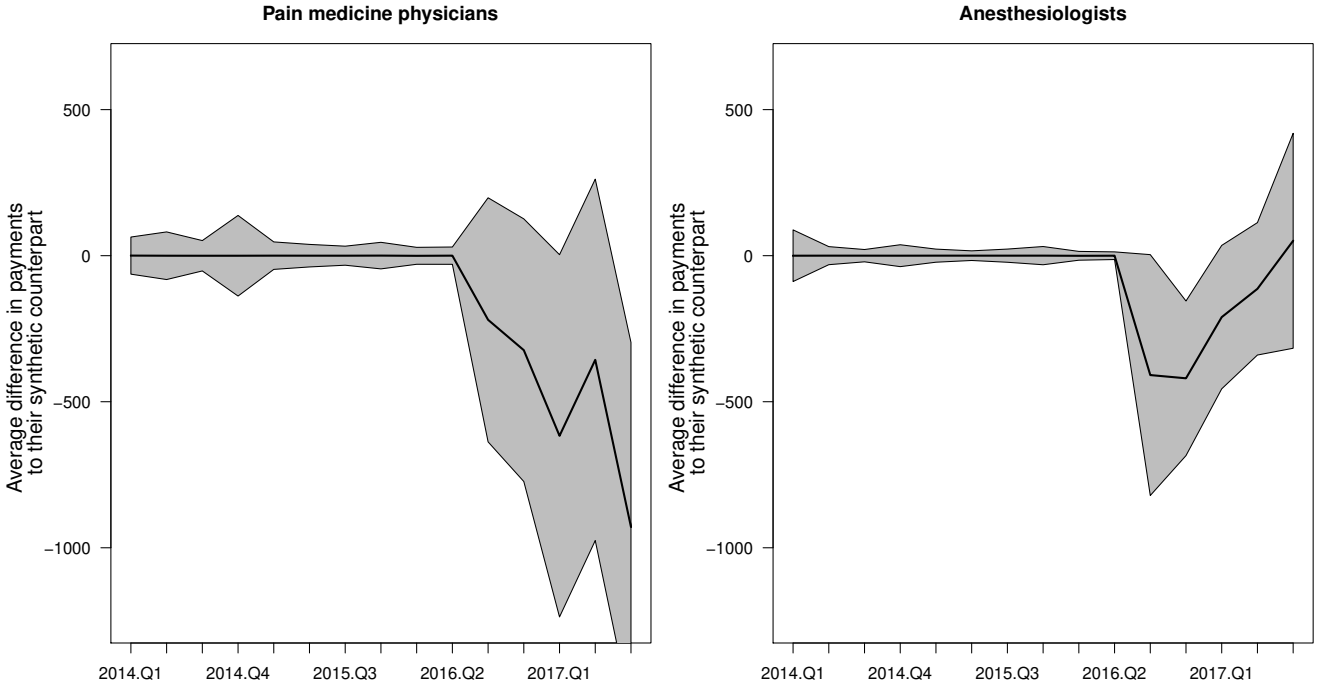

Figure 5: Synthetic counterpart analysis for MML passage on payments to physicians from 13 states in the US.

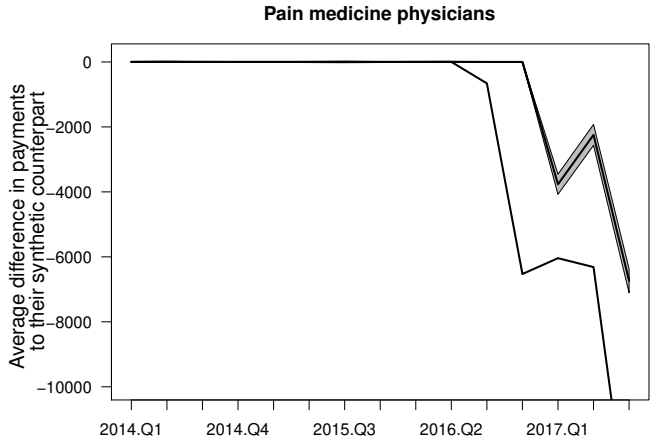

Figure 6: Synthetic control analysis for MML passage on payments to pain medicine physicians in Florida. The dashed line pretends MML passage in FL happened in the second quarter of 2016.

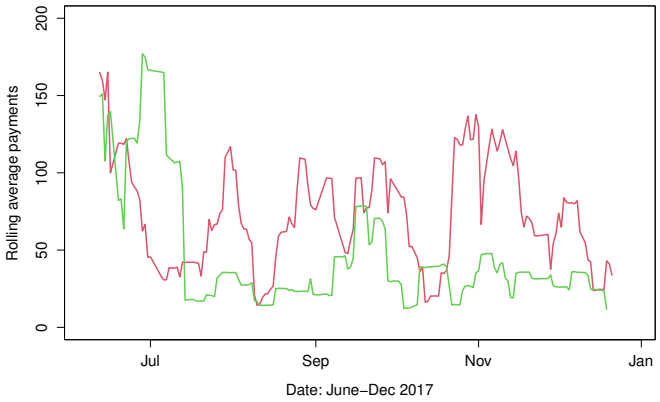

Figure 7: Payments to physicians in Florida between June and Dec 2017 in the cities without any marijuana dispensary at that time in red and with a marijuana dispensary at that time in green.

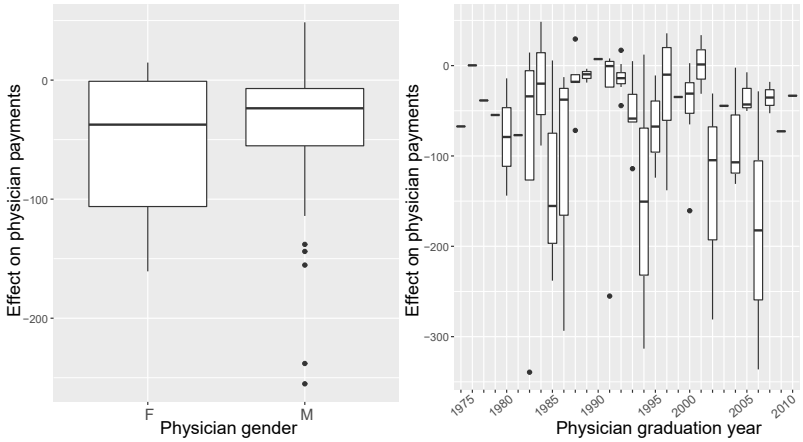

Figure 8: Effect heterogeneity by physician gender and year of graduation.

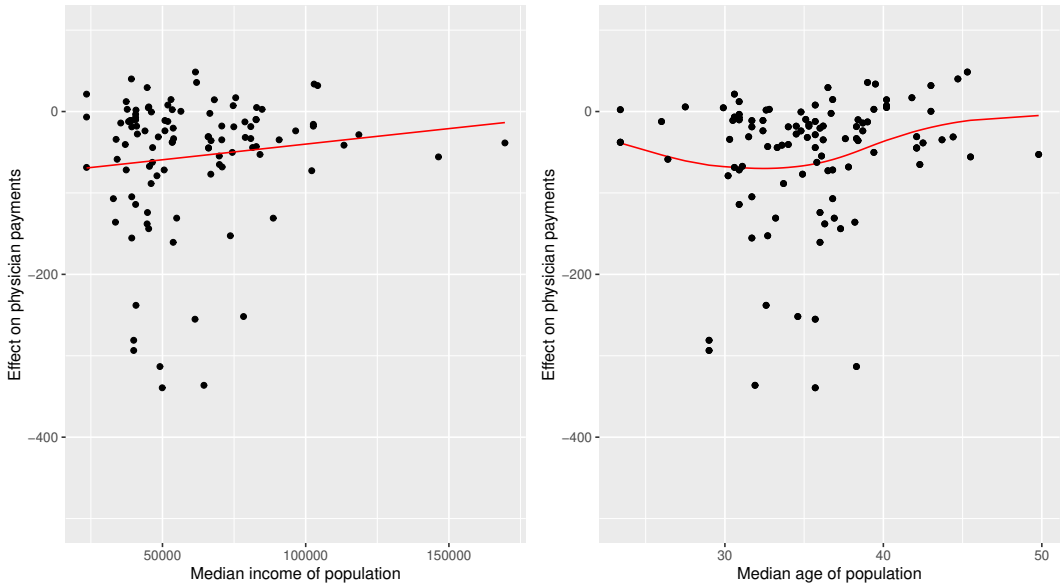

Figure 9: Effect heterogeneity by median income and the age of the zip code.

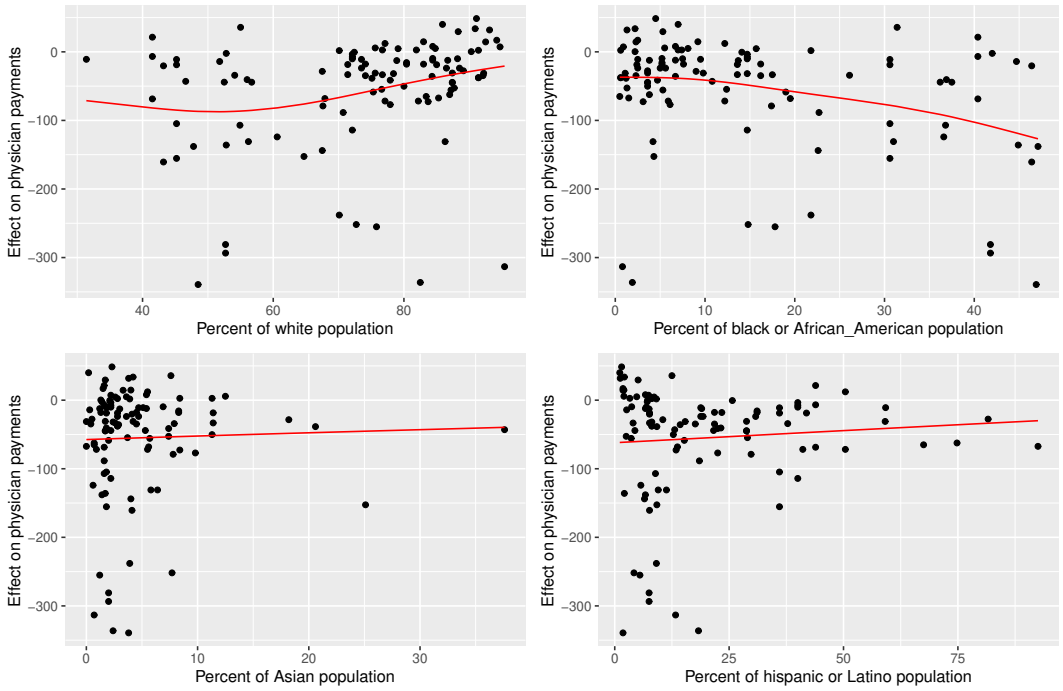

Figure 10: Effect heterogeneity by racial composition of the zip code.

Supplement to Using Penalized Synthetic Controls on Truncated data: A Case Study on Effect of Marijuana Legalization on Direct Payments to Physicians by Opioid Manufacturers

S1 Proofs and Other Section 3.3 Details

An example of the additive mixture model with illustration of the inferential challenges. Compared to the latent factor model analyzed in Abadie et al. (2010), it is more challenging to construct efficient SC estimates in (3). In figure S1, we show an example of (3). We consider $K = 3$, $\Phi_{1b} \sim \texttt{Uniform}(10, 60)$, $\Phi_{2b} = 1$, $\Phi_{3b} = 2$, $\mu_{1t} = 1$, $\mu_{2t} = t(\text{mod}4)$, and $\mu_{2t} = \lfloor t/4 \rfloor$. Considering each time period as a quarter, the second and third factors represent seasonal changes and yearly trends respectively. We consider $T = 100$ and Gaussian noise of variance 25. We also consider 3 unique Δ sequences with $\bar{\Delta}_h = -80 \sum_{j \in J_h} \boldsymbol{e}_j$ for $h = 1, 2, 3$ where $\boldsymbol{e}_j$ is the jth canonical basis vector of $\mathbb{R}^{100}$ and $J_1 = \{20, \ldots, 24\} \cup \{70, \ldots, 74\}$, $J_2 = J_1 + 10$ and $J_3 = J_2 + 10$. Consider one of the blue curves in the bottom most sub-plot of figure S1 as the treated unit and the others as controls. Then, minimizing $\text{Im}_{\texttt{sep}} + \nu\, \text{Im}_{\texttt{pool}}$ can lead to a PSC estimate with positive weights from green or blue controls. This however can create a problem in prediction at time $t = T$ as the red, blue and green curve have different dampening sequences. We would like to estimate blue treated units by only blue controls so that we do not have controls with different Δs than the treated unit. In this aspect, $\text{Pen}_{\texttt{sep}}$ in (1) helps us in correctly learning the coefficients $\boldsymbol{\phi}_b$ for any treated unit b.

Notations. With $\nu = 0$, the optimization of (1) decouples in optimization for each treated unit separately. For $b \in \mathcal{B}$, denote the unit specific imbalance and penalty:

$$\text{Im}(\boldsymbol{w}, b) = \sum_{t=1}^{T-1} \left(y_{bt} - \sum_{c=1}^{C} w_c\, y_{ct} \right)^2 \text{ and}$$

$$\text{Pen}_\lambda(\boldsymbol{w}, b) = \sum_{t=1}^{T-1} \sum_{c=1}^{C} w_c \exp\left\{ \lambda \left(y_{ct}\, I\{y_{bt} = 0\} + y_{bt} I\{y_{ct} = 0\} \right) \right\}.$$

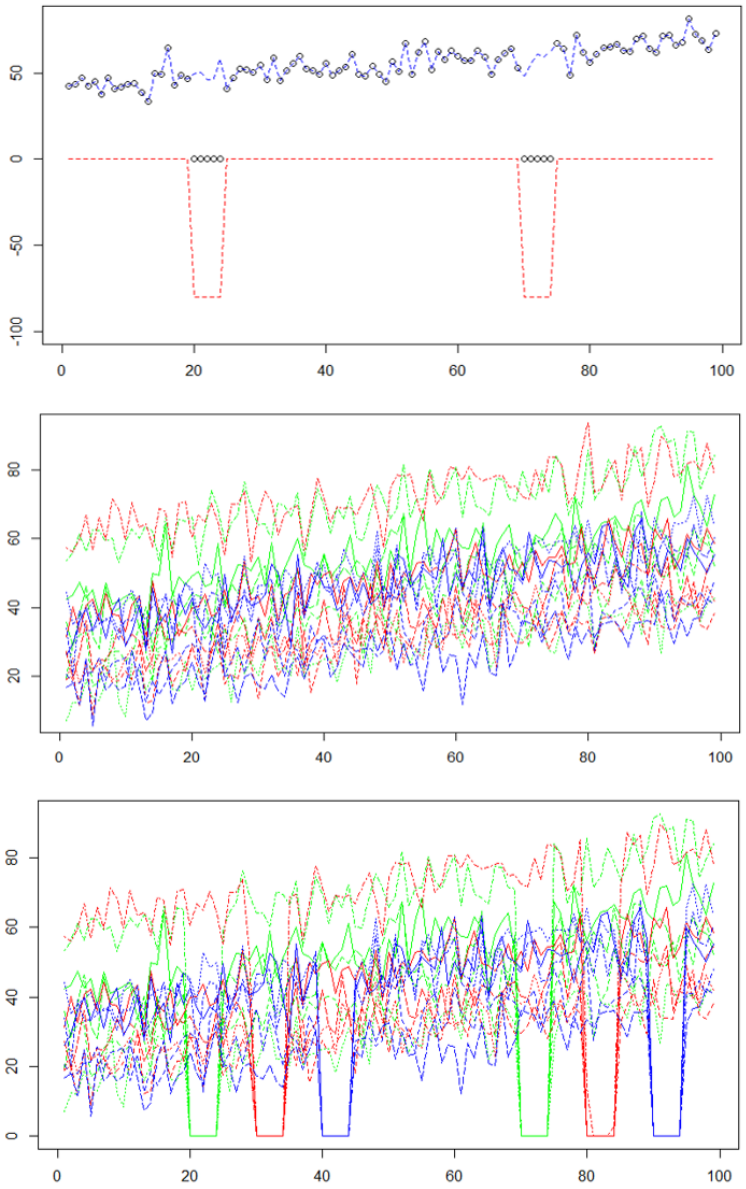

Figure S1: Simulated illustration of the additive mixture model. Top panel is based on the first unit and plots the observations y_{1t}, the factor model and noise $f_{1t} + \epsilon_{1t}$ and Δ_{1t} in black points, dotted blue line and dotted red line respectively for $t = 1, \ldots, 100$ time points. In the middle and bottom panel, we plotted observations from 15 units over the 100 time-points. The middle panel shows the latent pay-off z_{it}s and the bottom panel shows the truncated observed payments y_{it}. The three different colors in the middle and bottom plots represent the three different groups based on latent Δ sequences.

Henceforth, for the remainder of this section we consider minimizing the following criterion:

$$\text{Im}(\boldsymbol{w}, b) + \text{Pen}_\lambda(\boldsymbol{w}, b) \; . \tag{S1}$$

Recall,

$$\hat{\mathcal{C}}_b = \{c \in \mathcal{C} : y_{ct} \leq \psi^{-1} \text{ if } y_{bt} = 0 \text{ and } y_{ct} > 0 \text{ if } y_{bt} \geq \psi^{-1} \text{ for all } t = 1, \ldots, T-1\}.$$

Note that, non-asymptotically there are differences between the proposed penalty in (1) and a penalty that restricts the control set to $\hat{C}_b$. In the proposed penalty, when $y_{bt} = 0$ and $y_{ct} \neq 0$ the magnitude of y_{ct} also plays a role in setting its weight based on the optimization in (1). A penalty that restricts the control set to $\hat{C}_b$, would treat all non-zero y_{ct}s as same until they are below some threshold ψ^{-1}. We concentrate on an asymptotic analysis in this section. Let $\tau_b = \{1 \leq t \leq T-1 : y_{bt} > \psi^{-1}\}$. Partition $\text{Im}(\boldsymbol{w}, b)$ into $\text{Imp}(\boldsymbol{w}, b, \psi)$ and $\text{Imz}(\boldsymbol{w}, b, \psi)$ based on τ_b and its complement. Let $\hat{\mathcal{W}}_b = \{\boldsymbol{w} \in \mathcal{W} : w_i = 0 \text{ for } i \notin \hat{\mathcal{C}}_b\}$ be the weight space with support concentrated on the control subset $\hat{\mathcal{C}}_b$. We present the proofs for the arbitrary bth treated unit. For notational convenience, henceforth the dependence on b will be kept implicit unless it is needed to be mentioned.

Proof of Lemma 1. By definition of $\hat{\mathcal{C}}_b$, for any $\boldsymbol{w} \in \mathcal{W} \setminus \hat{\mathcal{W}}_b$ we have

$$\text{Pen}_\lambda(\boldsymbol{w}; b) \geq (CT)^{-1} \exp(\lambda \psi^{-1}) = CT(\log T)^2. \tag{S2}$$

Define the subset $\hat{\mathcal{D}}_b$ of $\mathcal{C}$ as $\hat{\mathcal{D}}_b = \{c \in \mathcal{C} : \sum_{t=1}^{T-1} I\{y_{ct} = 0\} I\{y_{bt} = 0\} + I\{y_{ct} \neq 0\} I\{y_{bt} \neq 0\} = T-1\}$. Let $\hat{\mathcal{W}}_{\mathcal{D}}$ be the set of all weights in $\mathcal{W}$ that have support contained in $\hat{\mathcal{D}}_b$. Note, that for any weights $\boldsymbol{w} \in \hat{\mathcal{W}}_{\mathcal{D}}$, $\text{Pen}_\lambda(\boldsymbol{w}; b) = 0$ and thus, $\hat{\mathcal{W}}_{\mathcal{D}} \subseteq \hat{\mathcal{W}}_b$.

By assumptions A1 to A4, we have: $P(\hat{\mathcal{C}}_b \subseteq \hat{\mathcal{D}}_b) \leq 1 - T^{-2}$. Let $S_b = \{1 \leq t \leq T-1 : \delta_{bt} \neq 0\}$. Decompose the sum of squared imbalances $\text{Im}(\boldsymbol{w}, b)$ into $\text{Im}_1(\boldsymbol{w}, b)$ and $\text{Im}_2(\boldsymbol{w}, b)$ based on S_b and its complement. Then, by (4) of Assumption A5, it follows that with probability $1 - T^{-2}$ we have $\min_{\boldsymbol{w} \in \hat{\mathcal{W}}_{\mathcal{D}}} \text{Im}_1(\boldsymbol{w}, b) \leq O(T \log T)$. Also, by A1 to A4, we have $\min_{\boldsymbol{w} \in \hat{\mathcal{W}}_{\mathcal{D}}} \text{Im}_2(\boldsymbol{w}, b) = O(\log T)$ with probability $1 - 1/T^2$. Thus, the minimal value of the optimization criterion $\text{Im}(\boldsymbol{w}, b) + \text{Pen}_\lambda(\boldsymbol{w}, b)$ over $\boldsymbol{w} \in \hat{\mathcal{W}}_{\mathcal{D}}$ is $O(T \log T)$ with probability $1 - T^{-2}$ which is less than the penalty in (S2). Thus, the result follows.

Proof of Lemma 2. First note that by construction of $\hat{\mathcal{C}}_b$ we have for all $c \in \hat{\mathcal{C}}_b$: $\quad (a) \quad y_{ct} \leq \psi^{-1} \Leftrightarrow$

$y_{bt} = 0$, and (b) $y_{ct} > 0 \Leftrightarrow y_{bt} \geq \psi^{-1}$. The following two identities directly follow from these relations:

$$P(y_{bt} < \psi^{-1}) + P(y_{bt} \geq \psi^{-1} \text{ and } \inf_{c \in \mathcal{C}_b} y_{ct} > 0) = 1 \tag{S3}$$

$$P(\inf_{c \in \mathcal{C}_b} y_{ct} = 0) + P(y_{bt} \geq \psi^{-1} \text{ and } \inf_{c \in \mathcal{C}_b} y_{ct} > 0) = 1. \tag{S4}$$

Now, for some $c \in \mathcal{C}_b$ if $\Delta_c \neq \Delta_b$ then by A4 there exists some t such that $|\delta_{ct} - \delta_{bt}| \geq f^* + \gamma$. Consider the following two sets: $A_t = \{\delta_{bt} = 0 \text{ and } \delta_{ct} \leq -f^* - \gamma \text{ for some } c\}$ and $B_t = \{\delta_{bt} \leq -f^* - \gamma \text{ and } \delta_{ct} = 0 \text{ for some } c\}$. Note that, $P(\Delta_c \neq \Delta_b \text{ for some } c) \leq \sum_{t=1}^{T} P(A_t) + P(B_t)$.

We first concentrate on set A_t. Note that on A_t, we have $P(y_{bt} < \psi^{-1}|A_t) \leq P(\epsilon_{bt} < -f_* + \psi^{-1}) \leq P(\epsilon_{bt} < -\gamma + \psi^{-1})$, where the last inequality follows by assumption A1. Also, in this case note that,

$$P(\inf_{c \in \mathcal{C}_b} y_{ct} > 0|A_t) \leq P(\sup_{c \in \mathcal{C}_b} \epsilon_{ct} + f^* \leq -\delta_{ct}|A_t) \leq P(\sup_{c \in \mathcal{C}_b} \epsilon_{ct} \leq \gamma).$$

By (S3) we have:

$$P(A_t) \leq P(\inf_{c \in \mathcal{C}_b} y_{ct} > 0|A_t) + P(y_{bt} \geq \psi^{-1} \text{ and } \inf_{c \in \mathcal{C}_b} .y_{ct} > 0|A_t)$$

The right side above is again upper bounded by $P(\epsilon_{bt} < -\gamma + \psi^{-1}) + P(\sup_{c \in \mathcal{C}_b} \epsilon_{ct} \leq \gamma) \leq 2(\log T)^{-1/2} T^{-2}$.

On B_t, we compute $P(\inf_{c \in \mathcal{C}_b} y_{ct} = 0|B_t)$ and $P(y_{bt} \geq \psi^{-1}|B_t)$. Again as before using (S4) we have

$$P(B_t) \leq P(\inf_{c \in \mathcal{C}_b} \epsilon_{ct} \leq -\gamma) + P(\epsilon_{bt} > \gamma + \psi^{-1}) \leq 2(\log T)^{-1/2} T^{-2}.$$

Thus, we have, $P(\Delta_c \neq \Delta_b \text{ for some } c) \leq \sum_{t=1}^{T} P(A_t) + P(B_t) \leq 4(\log T)^{-1/2} T^{-1}$.

Proof of Theorem 1. First, note that when $y_{bT} = 0$, then by Assumption A3 and lemma 2 it follows that with probability $1 - T^{-1}(\log T)^{-\alpha}$, for $\alpha < 1/2$, we have $y_{cT} = 0$ for all $c \in \hat{\mathcal{C}}_b$. Thus, in this case $\hat{y}_{bT} - y_{bT} = 0$, with probability $1 - T^{-1}(\log T)^{-\alpha}$.

Henceforth, we concentrate on the case where $y_{bT} > 0$. Then, $y_{bT} = z_{bT}$. As $y_{bT} > 0$, using

Assumption A1, A3 and Lemma 2 again, it follows that

$$P\Big(\inf_{c\in\hat{\mathcal{C}}_b} Y_{cT} > 0\Big) \geq 1 - T^{-1}(\log T)^{-\alpha}.$$

Thus, in this case with the aforementioned probability $y_{bT} - \hat{y}_{bT}$ reduces to $z_{bT} - \sum_{c\in\hat{\mathcal{C}}_b} w_c z_{cT}$. We focus on providing probabilistic control on this difference: $z_{bT} - \sum_c w_c z_{cT}$ for weights $\boldsymbol{w} \in \mathcal{W}_b$. For any $\boldsymbol{w} \in \mathcal{W}_b$, define $\hat{z}_{bt}(\boldsymbol{w}) = \sum_c w_c z_{ct}$ for all $t \in \tau_b$. As all weights in $\mathcal{W}_b$ have support in $\hat{\mathcal{C}}_b$, $\hat{z}_{bt}(\boldsymbol{w})$ is well-defined for all $t \in \tau_b$. When, $t = T$, we define $\hat{z}_{bT}(\boldsymbol{w}) = \sum_c w_c z_{cT}$ when $\inf_{c:w_c>0} y_{cT} > 0$; else it is set to $-\infty$. From (3), we know that for all $t \in \tau_b$,

$$z_{bt} - \hat{z}_{bt} = R_f(t) + R_\delta(t) + R_\epsilon(t), \text{ where, } R_f(t) = \sum_{k=1}^{K} \mu_{kt}\big(\phi_{kb} - \sum_c w_{bc}\phi_{kc}\big),$$
$$R_\delta(t) = \sum_c w_{bc}(\bar{\delta}_{h(b),t} - \bar{\delta}_{h(c),t}), \text{ and } R_\epsilon(t) = \epsilon_{bt} - \sum_c w_{bc}\epsilon_{ct}.$$

Note that, unlike (7) we have kept the dependence on $\boldsymbol{w}$ implicit here. We only consider weights $\boldsymbol{w} \in \mathcal{W}_b$. Consider the set $\mathcal{A} = \{\sup_{c\in\mathcal{C}_b} |\Delta_c - \Delta_b| = 0\}$. By Lemma 2, $P(\mathcal{A}) \geq 1 - 4T^{-1}(\log T)^{-1/2}$.

Consider the following vector $\mathcal{R}_z = \{z_{bt} - \hat{z}_{bt} : t \in \tau_b\}$. It is of dimension $|\tau_b| := \text{size}(\tau_b) = s_b$. Stacking the above equation in vector format for $t \in \tau_b$ we have:

$$\mathcal{R}_z = \mathbb{M}_b \cdot \mathcal{R}_\phi + \mathcal{R}_\delta + \mathcal{R}_\epsilon \,, \tag{S5}$$

where, $\mathcal{R}_\delta = (R_\delta(t) : t \in \tau_b)$, $\mathcal{R}_\epsilon = (R_\epsilon(t) : t \in \tau_b)$ are s_b dimensional vectors and $\mathcal{R}_\phi$ is the K dimensional vector $(\phi_{kb} - \sum_c w_{bc}\phi_{kc} : 1 \leq k \leq K)$. The matrix $\mathbb{M}_b$ is of $s \times K$ dimension and $(\mathbb{M}_b)_{ij} = \mu_{ji}$. Note that if $\tau_b = 1, \ldots, T-1$ then $\mathbb{M}_b = \mathbb{M}$ which is the matrix of all latent factors and thus, $\mathbb{M}_b$ depends on b only through τ_b. Note that, $\mathbb{M}_b$ does not depend on $\boldsymbol{w}$. Further note that,

$$\mathbb{M}'\mathbb{M} = \sum_{t=1}^{T-1} \boldsymbol{\mu}_t' \boldsymbol{\mu}_t = H \text{ and } \mathbb{M}_b'\mathbb{M}_b = \sum_{t\in\tau_b} \boldsymbol{\mu}_t' \boldsymbol{\mu}_t = H_b,$$

where, $\boldsymbol{\mu}_t = (\mu_{kt} : 1 \leq k \leq K)$ is a K dimensional vector. Next, note that,

$$z_{bT} - \hat{z}_{bT} = \boldsymbol{\mu}_T' \, \mathcal{R}_\phi + R_\delta(T) + R_\epsilon(T) = \boldsymbol{\mu}_T' \, \bar{H}_b(\mathcal{R}_z - \mathcal{R}_\delta - \mathcal{R}_\epsilon) + R_\delta(T) + R_\epsilon(T), \tag{S6}$$

where, $\bar{H}_b = (\mathbb{M}_b'\mathbb{M}_b)^{-1}\mathbb{M}_b'$ and the second equation follows by inverting (S5). On the set $\mathcal{A}$, we have $\mathcal{R}_\delta = \mathbf{0}$ and $R_\delta(T) = 0$. And so, on the set $\mathcal{A}$, we have

$$|z_{bT} - \hat{z}_{bT}| \leq |\boldsymbol{\mu}_T'\bar{H}_b\mathcal{R}_z| + |\boldsymbol{\mu}_T'\bar{H}_b\mathcal{R}_\epsilon| + |R_\epsilon(T)| \quad \leq ||\boldsymbol{\mu}_T||_2 \, ||\bar{H}_b\mathcal{R}_z||_2 + ||\boldsymbol{\mu}_T||_1||\bar{H}_b\mathcal{R}_\epsilon||_\infty + |R_\epsilon(T)|.$$

The bounds on the right side above can be further simplified. Note that,

$$||\bar{H}_b \, \mathcal{R}_z||_2 \leq \{\sigma_1(\bar{H}_b'\bar{H}_b)\}^{1/2} \cdot ||R_z||_2 \leq \kappa_b \, s_b^{-1/2} \, m_b^{-1/2} \, \{\text{Imp}(\boldsymbol{w}, b)\}^{1/2},$$

where, $\sigma_1(\bar{H}_b'\bar{H}_b)$ is the largest eigenvalue of $\bar{H}_b'\bar{H}_b$. The second inequality follows by noting that $||R_z||_2^2 = \text{Imp}(\boldsymbol{w}, b, \psi) \leq \text{Imp}(\boldsymbol{w}, b, 0) = \text{Imp}(\boldsymbol{w}, b)$ and $\sigma_1(\bar{H}_b'\bar{H}_b) \leq \kappa_b m_b^{-1}$ where κ_b is the condition number and m_b is the lowest eigenvalue of $s_b^{-1}H_b$.

Now note that, $R_\epsilon(T) \stackrel{d}{=} N(0, v_T)$ where $v_T = \sigma^2(1 + \sum_c w_{bc}^2) \leq 2\sigma^2$ as $\sum_c w_{bc} = 1$ and $w_{bc} \geq 0$. Next, we use the naive bound: $||\bar{H}_b\mathcal{R}_\epsilon||_\infty \leq ||\bar{H}_b\mathcal{R}_{b,\epsilon}||_\infty + ||\bar{H}_b\mathcal{R}_{\mathcal{C},\epsilon}||_\infty$. For any fixed $\boldsymbol{w}$, $R_\epsilon(T), \mathcal{R}_{b,\epsilon}$ and $\mathcal{R}_{\mathcal{C},\epsilon}$ are independent among themselves. Also, $\bar{H}_b\mathcal{R}_{b,\epsilon} \stackrel{d}{=} N(0, \sigma^2\bar{H}_bH_b') = N(0, \sigma^2 H_b^{-1})$. Thus, $||\bar{H}_b\mathcal{R}_{b,\epsilon}||_\infty$ is stochastically bounded by $\sigma\check{m}_b^{-1/2}X_1$ where $\check{m}_b = m_b s_b$ and X_1 is the maximum of s independent Chi-square random variables. Similarly, $||\bar{H}_b\mathcal{R}_{\mathcal{C},\epsilon}||_\infty$ is stochastically bounded by $2\sigma\check{m}_b^{-1/2}X_2$ where $X_2 \stackrel{d}{=} X_1$ and independent of X_1. Thus, we have: $||\bar{H}_b\mathcal{R}_\epsilon||_\infty \leq 2\sigma\check{m}_b^{-1/2}X_3$ where $X_3 = X_1 + X_2$. Accumulating the above bound we obtain the following bound for all events in the set $\mathcal{A}$:

$$|z_{bT} - \hat{z}_{bT}| \leq s_b^{-1/2} \, m_b^{-1/2}||\boldsymbol{\mu}_T||_2 \left(\kappa_b \, \{\text{Imp}(\boldsymbol{w}, b)\}^{1/2} + 2\sigma X_3\right) + \sqrt{2}\,\sigma X_4,$$

where $X_4 \stackrel{d}{=} \chi(1)$ is independent of X_3.

Now, let $l = 2\log s_b$ and $h = \log\log s_b$. Consider the set $\mathcal{D} = \{X_3 \leq 2\sqrt{l+h}$ and $X_4 \leq \sqrt{l}\}$. As X_3 and X_4 are independent using tails bounds on the Mill's ratio it follows that $P(\mathcal{D}) \geq (1 -$

$s_b^{-1}(\log s_b)^{-1/2})^2$. On $\mathcal{A} \cap \mathcal{D}$ we have:

$$|z_{bT} - \hat{z}_{bT}| \leq m_b^{-1/2} ||\boldsymbol{\mu}_T||_2 \left(\kappa_b \left\{ s_b^{-1} \text{Imp}(\boldsymbol{w}, b) \right\}^{1/2} + 4\sigma \sqrt{(l+h)/\tau_b} \right) + \sigma \sqrt{2l},$$

and the result follows as $P(\mathcal{A} \cap \mathcal{D}) \geq 1 - P(\mathcal{A}^c) - P(\mathcal{D}^c) \geq 1 - 4T^{-1}(\log T)^{-1/2} - \{1 - (1 - s_b^{-1}/(\log s_b)^2\} \geq 1 - 3s_b^{-1}/(\log s_b)^{-1/2}$,

$$|z_{bT} - \hat{z}_{bT}| \leq m_b^{-1/2} ||\boldsymbol{\mu}_T||_2 \left(\kappa_b \left\{ \tau_b^{-1} \text{Imp}(\boldsymbol{w}, b) \right\}^{1/2} + 8(s_b/T)^{-1} \sigma \sqrt{T^{-1} \log T} \right) + 2\sigma \sqrt{\log T}.$$

Proof of Lemma 3. Here, we first exhibit an instance of Θ_T over which the risk of the SC estimate is very high with high probability. The risk bound on the PSC estimate follows from the result in Theorem 1.

Consider an asymptotic set-up with C fixed and $T \to \infty$. Let $\sigma = 1$ and $\ell = (4 \log T)^{-1/2}$. With out loss of generality assume, $\delta_{b,T-1} = \delta_{b,T} = 0$. Consider a control a which satisfies $\boldsymbol{\phi}_a = \boldsymbol{\phi}_b = \ell$ and $\delta_{a,t} = \delta_{b,t}$ for $t = 1, \ldots, T-2$ and $\delta_{a,T-1} = \ell$ and $\delta_{a,T} = T^{\delta}$ for some $\delta > 1$. The, f_* and f^* are both $O(\ell)$. Let all other controls $c \in \mathcal{C}$ are based on parameters in (3) of the form: $\phi_{ck} = \phi_{bk} + \ell$ for $k = 1, \ldots, K$ and $\Delta_c = \Delta_b$ for $c \in \mathcal{C} \setminus \{a\}$.

Consider minimization of the criterion (2) when $\lambda = 0$. For this set of parameter $\boldsymbol{\theta}$ (defined above) it follows that the SC estimate $\hat{y}_{bT}(\boldsymbol{w}_{\text{SC}})$ satisfies $P_{\boldsymbol{\theta}}\big(\hat{y}_{bT}(\boldsymbol{w}_{\text{SC}}) = y_{aT}\big) \geq 1 - T^{-1}$, which implies that $P_{\boldsymbol{\theta}}\big(|\hat{y}_{bT}(\boldsymbol{w}_{\text{SC}}) - y_{bT}| \geq T^{\delta}\big) \geq 1 - T^{-1}$ as $\delta_{b,T} = 0$ and $\delta_{a,T} = T^{\delta}$.

On the other hand, minimization of the criterion (2) when $\lambda \geq \lambda_T$ produces PSC estimate $\hat{y}_{bT}(\boldsymbol{w}_{\text{PSC}})$ which satisfies: $P_{\boldsymbol{\theta}}\big(\hat{y}_{bT}(\boldsymbol{w}_{\text{PSC}}) = O(\log T)\big) \geq 1 - T^{-1}$. To derive the above results first note that the set-up satisfies all the assumption A1-A6 mentioned in the main paper. In particular, assumption A5 follows from the fast that for any $c \in \mathcal{C} \setminus \{a\}$ we have $P_{\boldsymbol{\theta}}(\sum_{t=1}^{T-1}(y_{bt} - y_{ct})^2 \leq T(\ell^2 + \ell^2 K^2 \zeta)) \geq 1 - 2^{-1}T^{-1}$, where, $\zeta = \sup_k ||\boldsymbol{\mu}_k||_{\infty}$. The result follow from (9) in the main paper.

S2 Additional empirical evaluations

Choice of λ

Our proposed penalized synthetic control method includes a penalty parameter λ. We proposed a cross-validation based method to select this penalty parameter. To assess the accuracy of this choice, we conducted an empirical evaluation. We used the simulation models in Section 4 and ran our method over a grid of choices of λs. The two panels of Figure S2 plot the relative RMSEs for ITT and ATT at these λ values, compared to the RMSEs for the cross-validated choice of λ. The plot shows that the RMSEs are the lowest at the proposed choice. While for λ larger than the proposed choice the RMSEs are much higher, for λs slightly smaller than the proposed choice, we see relatively smaller losses of accuracy. Overall, this empirical evaluation provides a justification for the proposed choice of λ.

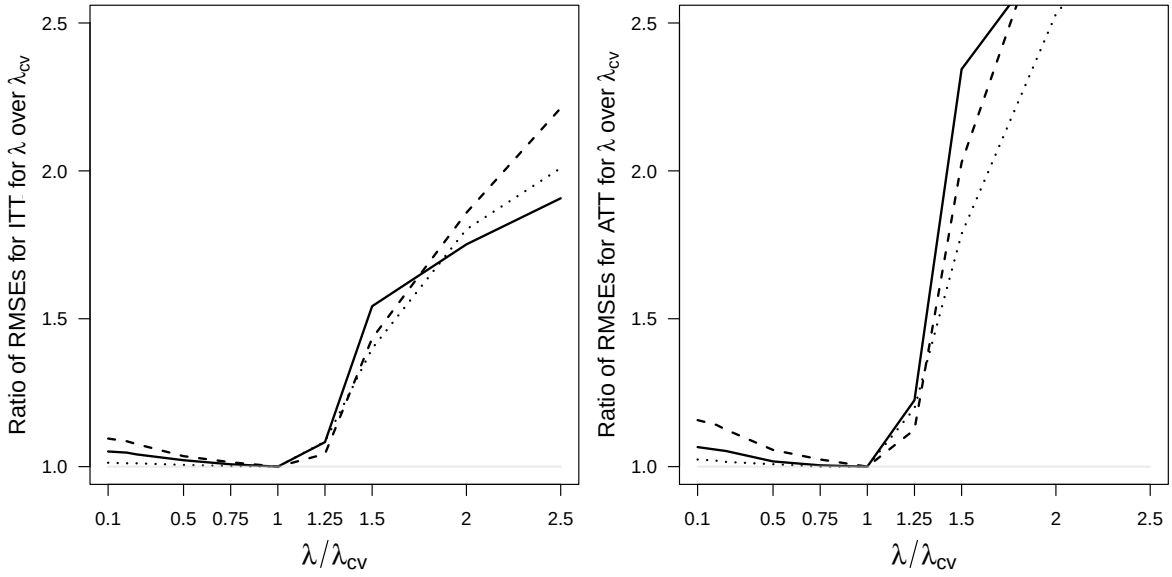

Figure S2: Ratio of RMSEs vs Ratio of λs. The solid, dashed and dotted lines correspond to simulation models in in Section 4 where clusters are probabilistically similar, probabilistically dissimilar, and deterministic and different, respectively. The cross-validated choice of λ results in the lowest RMSEs.

Comparison to the L_1 penalty

An alternative penalty to our proposed penalty would be to penalize the L_1 difference between the treated and control units' pre-treatment outcomes. However, unlike the proposed penalty, the weighted L_1 difference based penalty does not distinguish between non-zero and zero values of y_{bt}. As a consequence, penalized synthetic control estimates of y_{bT} based on the weighted L_1 difference based penalty can contain controls that are generated from a different dampening sequence in the censored additive factor model (1). Due to this, the component of the risk R_δ in (6) from the dampening sequencing is unbounded.

We conducted a detailed simulation to compare the proposed penalized method to this L_1 difference based penalty. In this simulation, we consider a simulation model where the factor model part only includes an intercept at 25, there is i.i.d. gaussian noise of standard deviation 20 and there is no treatment effect. Additionally, the dampening model is as in our simulation model in Section 4. Notice that the noise variance in this simulation is large compared to the simulation models there. Finally, we consider only 5 units per cluster, as compared to 10 units in our simulation in Section 4, to emphasize that the L_1 penalty will have a higher chance of making mistakes in picking the synthetic controls. Our simulation results are presented in Table S1. The results show a much better performance of the proposed penalized method compared to using L_1 penalty.

Table S1: Simulation comparison for different synthetic control methods when $\tau_i = 0$ for all: best performance in each row is in bold. Results are based on averaging over 500 simulations; standard errors are in the parentheses

	Synthetic control	Pooled SC	Proposed Penalized SC	L_1 Penalized SC
clusters are probabilistically similar				
l_2 Imbalance	25.51 (0.27)	18.79 (0.20)	**15.95 (0.13)**	18.05 (0.19)
RMSE for ITT	25.61 (0.36)	20.32 (0.29)	**18.76 (0.25)**	19.80 (0.28)
RMSE for ATT	14.57 (0.39)	10.63 (0.32)	**6.96 (0.20)**	9.19 (0.29)
clusters are probabilistically different				
l_2 Imbalance	26.14 (0.27)	18.53 (0.20)	**15.92 (0.13)**	17.93 (0.20)
RMSE for ITT	25.89 (0.33)	20.21 (0.29)	**18.63 (0.25)**	19.73 (0.22)
RMSE for ATT	13.35 (0.37)	10.46 (0.32)	**6.93 (0.20)**	9.15 (0.25)

A calibrated simulation study

For a more informative simulation study for our physician data, we conduct a simulation study "calibrated" to the main data set as in Kern et al. (2016). In particular, we consider the following simulation model.

There are 100 units observed over 13 time periods, among which 10 are pre-treatment time periods. The pre-treatment data for the units are sampled randomly from our physician payments data for the 10 pre-treatment time periods. Thus, the longitudinal observational in each simulated data set is representative of the data in our empirical study.

Next, we describe how the treated units are selected. The sampled control data are clustered using a k-means cluster on their vector of indicators if the observation is 0 or not. Thus, the k-means algorithm takes 100 binary vectors of size 10 as input, and we create 3 clusters by the clustering method. One unit each is selected from the three clusters as a treated unit randomly.

We consider two different post-treatment outcomes to capture different complexities in our empirical study. First, we use a linear outcome model that follows the same pattern in the three post-treatment periods as in time periods 1–3. More concretely, $y_{it} = \check{y}_{it} = y_{i(t-10)} - y_{i(t-10)} + y_{i10}$ for $t = 11, 12, 13$. Second, we use a quadratic outcome model where $\check{y}_{it} = \sqrt{y_{i(t-10)} y_{i(t-7)}}$ for $t = 11, 12, 13$. While, the post-treatment outcome for the treated units are $y_{i11} = \check{y}_{i11} \times .9$, $y_{i11} = \check{y}_{i11} \times .7$ and $y_{i11} = \check{y}_{i11} \times .8$. Thus, the treatment effects are heterogeneous, varies over time and exits only when there is a non-zero counterfactual payment.

We report the simulation methods in Figure S3 comparing the proposed penalized synthetic control method to the pooled synthetic control method. The plots show the l_2 imbalance and RMSEs for estimating ITT and ATT. The proposed method shows overall better performance across all the measures.

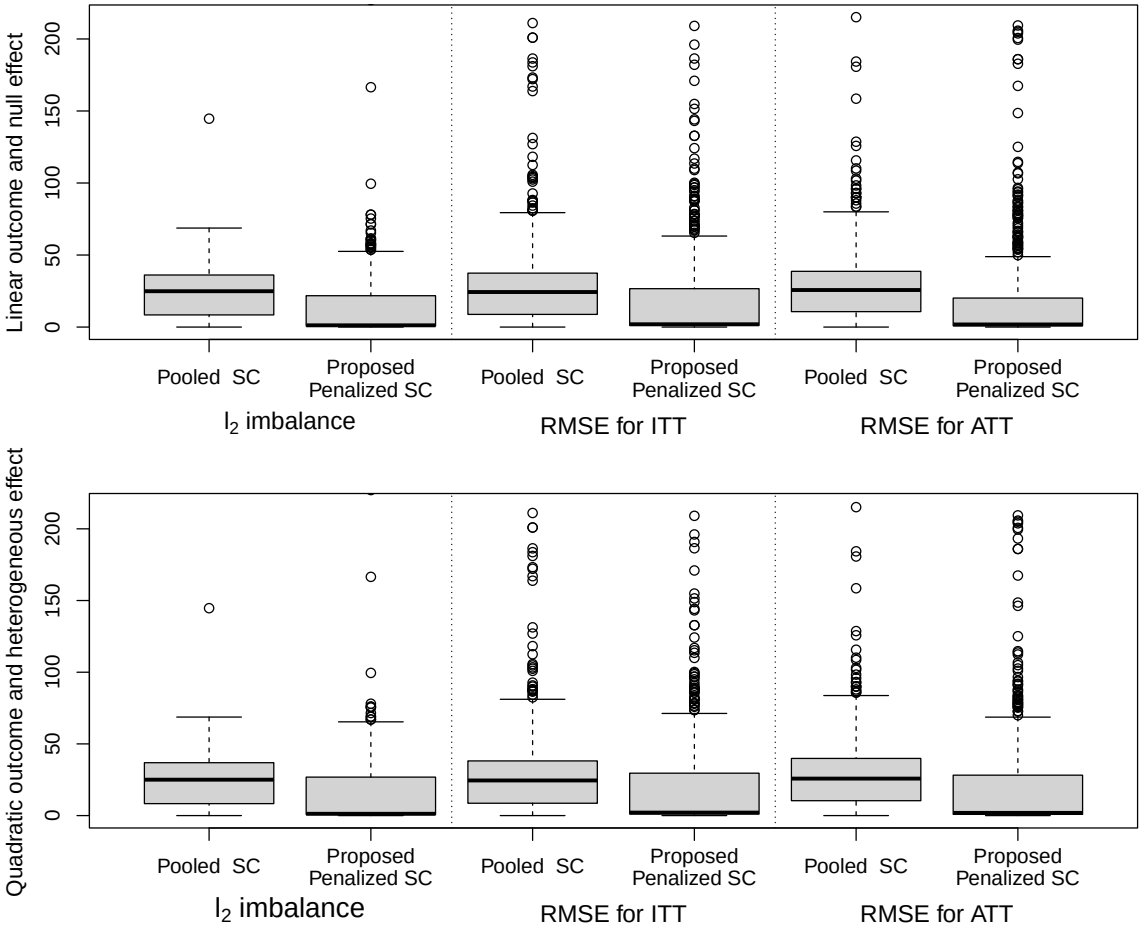

Figure S3: Result comparing pooled SC and the proposed penalized SC method on two calibrated data generating models. Based on 500 data sets.

S3 Additional findings and notes on physician payments study

Entry of new opioids during post-treatment period

In our current analysis we use the opioids drugs manufactured by five companies and their subsidiaries, namely, Depomed, INSYS, Janssen, Purdue Pharma, and Teva Pharmaceuticals. In 2017, our post-period, FDA approved Arymo ER (January), Vantrela ER (January), and RoxyBond (April).[1] Among these particular brands, it is pertinent to note that Vantrela ER is manufactured by Teva Pharmaceuticals. Our dataset encompasses the overall marketing activities associated with Teva Pharmaceuticals. Egalet Corporation, which manufactures Arymo ER, were involved in promoting this drug in the later part of 2017.[2] However, RoxyBond, distributed by Protega Pharmaceuticals, was not commercially launched in 2017.[3] Therefore, we expect limited interference due to new brand entry in our empirical analyses.

Balance diagnostics

For a synthetic control to be reliable, we need to make sure that the synthetic control matches are good. The standardized mean difference is commonly used to assess quality of a match in the traditional matching of a treated group to a control group. Recently, Parast et al. (2020) proposed a new type of measure that is similar to the original standardized differences for balance diagnosis of a synthetic control analysis. We use their balance measure to assess the quality of our synthetic control fit. They suggest that the synthetic control is reliable if the values of this measure are below the threshold of 0.10 on the absolute scale.

Figure S4 plots the standardized differences as suggested by Parast et al. (2020) for each physician and their synthetic counterpart. Except for a handful of cases only in the first two time periods, we see that all standardized differences are below the desired 0.10 on the absolute scale. The average of the absolute standardized differences is below 0.10 for all physicians in our data, which is recommended

[1] https://www.fda.gov/drugs/information-drug-class/timeline-selected-fda-activities-and-significant-events-addressing-substance-use-and-overdose

[2] https://www.sec.gov/Archives/edgar/data/1586105/000110465917019839/a17-10087_1ex99d1.htm

[3] https://www.globenewswire.com/news-release/2022/07/06/2474928/0/en/Protega-Pharmaceuticals-LLC-Announces-Commercialization-of-RoxyBond-oxycodone-hydrochloride-Tablets-CII-in-the-U-S.html

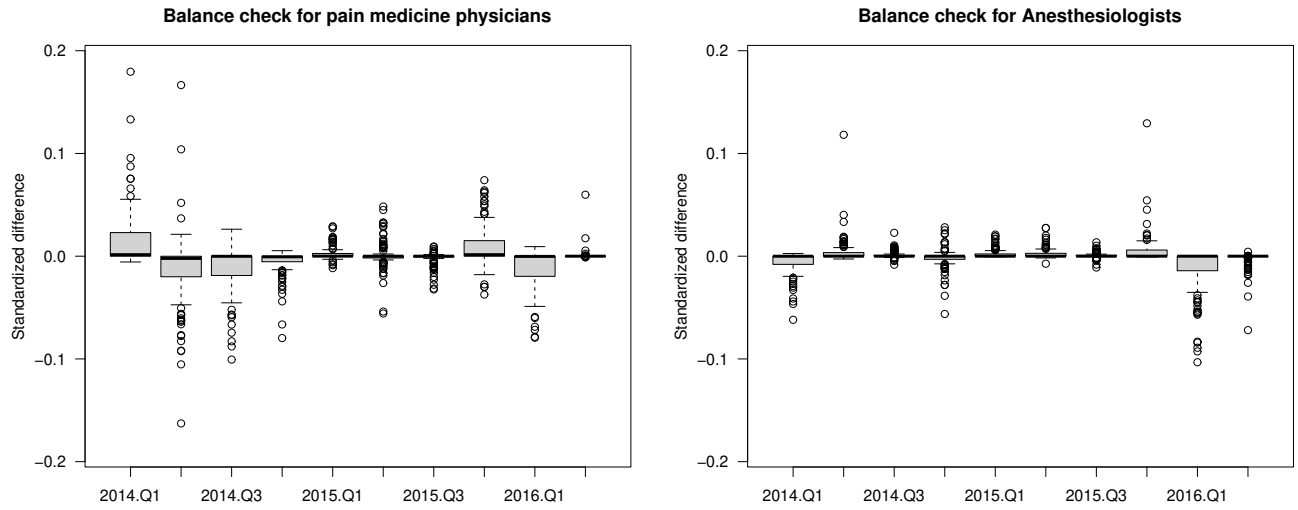

Figure S4: Standardized mean difference for the penalized synthetic control fit for pain-medicine physicians and anesthesiologists. This plot corresponds to Figure 5 of the main paper.

guide to judging a good match (Parast et al., 2020).

Two-way fixed effects model for the excluded units

Recall from Section 2 that about 8% of the units were removed for our synthetic control analysis because their outcomes had extreme values. Here, we try to assess if excluding these physicians could have severely changed our findings.

We have run a two-way fixed effects analysis with time and state fixed effects on the physicians who were removed because of unusually high payments. The analysis shows a significant negative effect of MML passage consistent with the findings from our SC analysis. We see a significant negative coefficient for the treatment. Note that the standard errors are based on cluster robust standard errors with states as clusters. A two-way fixed effects model would have been a typical choice for analyzing this data.

```
 Results for Two-Way Fixed Effects Model
==========================================
                    Dependent variable:
                ---------------------------
                     I(log(pay + 10))
------------------------------------------
```

```
MML              -1.362*** (-2.210, -0.513)
time               -0.023 (-0.073, 0.028)
as.factor(st)2      3.082*** (1.992, 4.172)
as.factor(st)3   -1.784*** (-3.061, -0.508)
as.factor(st)4     -0.348 (-2.016, 1.320)
as.factor(st)5     -0.040 (-1.427, 1.347)
as.factor(st)6      1.071 (-0.206, 2.348)
as.factor(st)7   -3.103*** (-4.620, -1.585)
as.factor(st)8     -0.984 (-2.797, 0.828)
as.factor(st)9    -1.727** (-3.203, -0.251)
as.factor(st)10 -2.733*** (-4.166, -1.300)
as.factor(st)11    -0.575 (-2.213, 1.064)
as.factor(st)12    -1.082 (-2.854, 0.690)
as.factor(st)13    -0.372 (-2.043, 1.299)
as.factor(st)14    1.882*** (0.547, 3.217)
as.factor(st)15    -0.886 (-2.663, 0.891)
as.factor(st)16   -1.476** (-2.921, -0.030)
as.factor(st)17     0.050 (-1.606, 1.706)
as.factor(st)18    2.078*** (0.951, 3.205)
as.factor(st)19    -1.068 (-2.593, 0.458)
as.factor(st)20   -1.936** (-3.551, -0.321)
as.factor(st)21    -0.970 (-2.543, 0.603)
as.factor(st)22    -0.500 (-1.907, 0.906)
as.factor(st)23 -2.104*** (-3.472, -0.736)
as.factor(st)24    -0.913 (-2.393, 0.568)
as.factor(st)25    -1.274 (-3.068, 0.520)
as.factor(st)26     0.617 (-0.751, 1.984)
as.factor(st)27 -2.782*** (-4.023, -1.540)
as.factor(st)28    -1.244 (-2.775, 0.287)
as.factor(st)29 -2.234*** (-3.789, -0.679)
as.factor(st)30     0.008 (-1.665, 1.682)
Constant           7.227*** (6.136, 8.318)
-----------------------------------------
R2                          0.300
Adjusted R2                 0.252
=========================================
Note:           *p<0.1; **p<0.05; ***p<0.01
```

However, it generally does not build on a causal framework and recent literature has argued that the two-way fixed effects results may be hard to give causal interpretation (Imai and Kim, 2021, Imbens et al., 2021). Thus, in our main analysis, we chose the synthetic control method that builds on a causal framework.

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 the American Statistical Association* **105 (490)**, 493–505.

Kern, H. L., E. A. Stuart, J. Hill, and D. P. Green (2016). Assessing methods for generalizing experimental impact estimates to target populations. *Journal of Research on Educational Effectiveness* **9 (1)**, 103–127.

Imai, K. and I. S. Kim (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis* **29 (3)**, 405–415.

Imbens, G., N. Kallus, and X. Mao (2021). Controlling for unmeasured confounding in panel data using minimal bridge functions: From two-way fixed effects to factor models. arXiv:2108.03849.