A Deep Learning Assessment of the Right to Counsel

Patrick Power, Shomik Ghosh and Markus Schwedeler
DRAFT

November 5, 2022

Abstract

Drawing from the Deep Learning Literature and in the language of Category Theory, we introduce a unified estimation framework that generalizes ordinary least squares, allows for nonparametric cluster effects,[1] and is inherently compositional, even under regularization. With this framework, we examine the effects of the Right to Counsel: a policy which ensures that low-income households facing eviction have access to free legal representation. Specifically, we consider the extent to which the policy increases housing instability by making it harder for low-income households to secure housing in the first place. Exploiting the ongoing roll-out of the policy across the state of Connecticut, our preliminary results suggest that housing is harder to secure in areas with higher eviction rates, indicating that the policy is less effective at scale than generally understood.

Keywords: Right to Counsel, Evictions, Deep Learning

---

[1]To clarify this point, in many applied microeconomic contexts, individuals grouped together in some fashion. In our context, this grouping is done via a shared zip code. We indicate this grouping via an indicator variable. The nonparametric component is that we allow these the effect of this group memebership on the outcome of interest to vary across the other controls/features. That is, we don't place a functional form assumption on the way the cluster indicator variable interacts with the other controls.

In this paper, the term *Right to Counsel* refers to a policy initiative which ensures that tenants have access to free legal representation in eviction cases.

1   Introduction

The 2 million evictions that occur each year across the United States are costly to individuals, landlords, courts, and the general public. Unlike criminal cases in the U.S., defendants in an eviction case are not entitled to a public attorney. A gap in legal representation, therefore, exists between landlords and tenants which is as large as $95\% - 1\%$ in some areas in favor of the landlord (Collinson et al. [2022]). Given the severity of the costs associated with eviction, the multitude of factors which contribute to its occurrence, and the typical manner in which a eviction case is settled, many in the U.S. believe that legal counsel should be provided to low income households. Over the past five years, 15 cities and 3 states have introduced a Right to Counsel with the aim of closing this gap and with the hope that doing so will improve downstream outcomes.

To some extent, this hope has been empirically justified when considering the direct outcomes of eviction cases. Both in the context of small scale randomized control trials as well as in city-wide roll-outs, researchers have generally found positive results (Seron et al. [2001], Greiner et al. [2012], Cassidy and Currie [2022]). The pressing unknown, though, is whether the indirect effects associated with the policy diminish or outweigh these positive legal outcomes. To date, one of the most consistent findings is that Legal Aid increases the length of the eviction process.[2] By increasing the cost of evicting a tenant, the policy may disincentive landlords from renting to low-income households in the first place. As Gunn [1995] writes, "By increasing landlords' costs of doing business, legal services attorneys may enrich their clients at the expense of all other similarly situated poor tenants."

In this paper, we draw on recent advances in deep learning to empirically assess whether this concern is justified. Specifically, we use the ongoing rollout of the Right to Counsel across the state of Connecticut to estimate its effect on the ease with which low-income households, who are currently experiencing homelessness, are able to find housing. A key challenge in our setting is that the treatment (the policy) is assigned a level (the zip code) above the unit of interest (the individual). Building on recent deep learning papers (Finn et al. [2017] and Kelly et al. [2020]), we implement an estimation framework that allows us to partial out the zip code effects in a nonparametric fashion.

While preliminary, our results suggest that across zip codes with a high eviction rate (as measured in 2019), the length of the housing search increases in response to the policy. Using a 30-day threshold, the probability that a household is still searching for housing after this cutoff increases by 5 percentage points. While the sign of this effect is relatively constant across a range of estimators, we emphasize that there are several important concerns/issues that cannot be addressed in our context. This paper should be therefore be understood

---

[2]Cassidy and Currie [2022] writes: "The number of days between a case filing and a judgment is also significantly longer in the UA zip codes after program implementation."

as the first empirical microeconomic work to consider the unintended consequences of this policy, but given the importance of this topic, it certainly should not be the last.

There are two distinct reasons for why there has been little empirical work on this issue previously. As Abramson [2021] notes, there hasn't been a suitable context to empirically assess the indirect effects of the policy. Prior works on the Right to Counsel either focus on small scale randomized control trials, such as (Greiner et al. [2012], Seron et al. [2001]), which by construction only allow for the study of direct outcomes, or as in the case of (Cassidy and Currie [2022] and Ellen et al. [2021]) study the city wide roll-out of New York City which is a very unique housing market.[3] Tackling a similar research question to ours, but written one year prior, Abramson [2021] resorts to studying the effects of the Right to Counsel via a counterfactual analysis.[4]

Only within the past year, as Washington and Connecticut adopted the Right to Counsel at the state level has the possibility of empirically assessing the policy at scale emerged. And only in the case of Connecticut, which adopted a meaningful and well documented staggered roll-out has this opportunity been realized. Due to supply constraints of legal services, low-income individuals in only a subset of Connecticut zip codes currently receive free legal aid under the policy. Meaningfully for our paper, several high eviction rate zip codes remain untreated as highlighted in figure 3a.

The second reason this topic remains understudied is that the indirect effects of the Right to Counsel are difficult to capture. Evictions themselves are often informal. Mathew Desmond writes in his New York Times Best Seller, *Evicted*, that informal evictions in Milwaukee account for 48% of forced moves while formal evictions account for 24%. It seems reasonable to expect landlords to respond to the Right to Counsel through a number of informal or hard to detect ways. Indeed. from conversations with lawyers who specialize in evictions, the question at times seemed not to be whether the costs of the Right to Counsel would be passed on to the tenants, but rather whether they would be observable in the data. As one lawyer explained, "The thing to remember is that higher costs for landlords always get passed on to the tenants in some form (higher rent, deposits, fees, etc.)."

To empirically assess the roll-out of the policy across the state of Connecticut, we therefore make use of individual level data related to local Rapid Rehousing Programs within the state. Tracked by the U.S. Department of Housing and Urban Development, theses programs assist households experiencing homelessness secure housing. Importantly for our work, these programs (1) are restricted to households who don't face significant barriers to housing, (2) provide limited short-term financial assistance and (3) require that the rental agreements that households sign have "the same rights and responsibilities as a typical lease holder."[5] For each household, we observe both pre-treatment characteristics

---

[3]Evans et al. [2019] writes, "Not all areas have experienced a decline in homelessness. The trends in New York City (CoC NY600) and Los Angeles County have been particularly different from the rest of the country . . . Note that since 2012, these rates have increased by 57 and 31 percent, respectively."

[4]As Abramson [2021] writes, "I quantify the model to match data on evictions, homelessness, and rents in San Diego County, and use it for counterfactual analysis."

[5]Reference: "It is imperative that any lease agreement provides the tenant with **the same rights and responsibilities as a typical lease holder** and that the financial terms of the lease are such that the household

(race, gender, family structure) as well as their length of their housing search, which is the main outcome variable of interest.
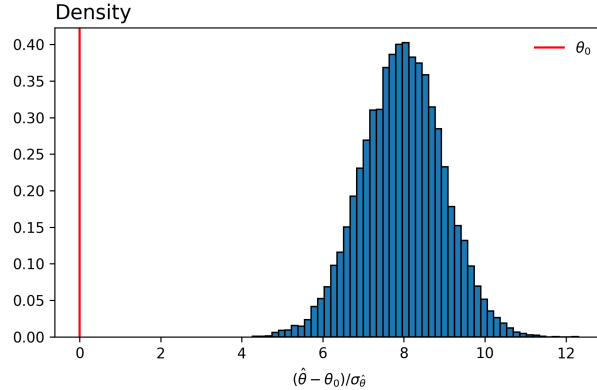

Figure 1: Under a conditional parallel trend assumption, linear models can fail (reproduced here). In this figure, we plot the sampling distribution of the standardized estimate of a linear difference-in-difference estimator. We use this plot to motivate our preference for fitting deep learning based difference-in-difference estimators.

To this data set, we fit a series of estimators all within a single, unified estimation framework. As previously mentioned, a key challenge in our context, and in many policy oriented papers, is that treatment is assigned at a level above the unit of observation (treatment is assigned at the zip code level, as the policy is rolled out across zip codes, while observations are the the individual level). In a cross-sectional setting, this presents a challenge because there is no within cluster variation to exploit. Instead of fitting within clusters, econometric models, in this setting, must partial out the cluster effects. Building off the deep learning papers of Finn et al. [2017] and Kelly et al. [2020], we introduce a framework that partials out the nonparametric effects in a controllable manner by *regularizing the forward pass* of model.

Deep Learning models, which we base our framework on, are a subset of machine learning models that can be constructed by composing parameterized functions and trained via gradient descent-like methods. Typical models involve composing linear maps with nonlinear activation functions which can be shown to be universal approximators (Hornik et al. [1989]). In recent years, such models have become easier to train as gradients are computed "automatically" often via reverse mode automatic differentiation (Griewank and Walther [2008]) and the entire training process can be compiled and scaled across hardware accelerators like GPUs.

While often touted for their ability to drastically improve their performance with the size of the data, we leverage them because they allow us to encode meaningful priors in a controllable fashion. This quasi-Bayesian viewpoint of learning/estimation in the finite sample is in line with modern deep learning works (Jacot et al. [2018], Nagarajan and Kolter [2019], Wilson [2020], Belkin [2021], Zhang et al. [2021]) which emphasize the importance

---

has a reasonable ability to assume rental costs once financial support ends (keeping in mind that in the majority of cases, even households with no income at move-in retain their housing)"

of "understanding the nature of the inductive bias"[6] of the estimator and differentiates this work from recent theoretical econometric work that seeks to reduce the influence of regularization in the estimation process (Chernozhukov et al. [2018]). We provide further clarification and justification for our approach in section 7.2.

Our high level aim is an empirical estimation framework that can seamlessly account for clustered data in a flexible way, has a clear inductive bias, can be easily extended, and is conceptually simple. **TIES**: **T**argeted (handle cluster data), **I**nterpretable (clear inductive bias), **E**xtendable and **S**imple. While the details of the framework are further explained in section 3, the key ideas are captured by the equations below. As indicated in the first line, the framework generalizes ordinary least squares (fitting a linear model to the data). Furthermore, as represented by the clusterMap, the model allows for nonparameteric cluster effects by allowing the parameters of the neural network to vary across clusters during the forward pass.[7] That is by equating clusters of the data to tasks in a meta-learning context, we partial out the cluster effects by allowing the parameters to adopt in the forward pass to the clusters, thereby "locally" correcting for the presence of clustered data in a global way. And finally, as indicated by the fish operator '>=>', the model remains inherently compositional, even under regularization. This is important for at least three reasons. First, it keeps the model conceptually simple. When we train the model we use one form of composition, when we predict with the model we use the standard form.[8] Second, it enables us to understand the adjustment for the presence of clustered data as a gradient correction that favors early stopping at the cluster level. And third, via this gradient correction, and the work of Domingos [2020], we can understand our approach within the kernel method framework which in contrast to deep learning is well developed and well understood.

$$\text{linearModel data} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(OLS)}$$

$$\text{linearModel} \circ \text{ identityMap data}$$

$$\text{linearModel} \circ \big(\text{featureMap data}\big) \text{ params}$$

$$\text{linearModel} \circ \big(\text{featureMap data}\big) \circ \text{identityMap params}$$

$$\text{linearModel} \circ \big(\text{featureMap data}\big) \circ \big(\text{clusterMap data}\big)\text{params}$$

$$\text{linearModel} >=> \big(\text{featureMap data}\big) >=> \big(\text{clusterMap data}\big)\text{params}$$

---

[6] Belkin [2021] "the properties that make some solutions preferable to others despite all of them fitting the training data equally"

[7] The clusterMap is the cluster specific map that takes the parameters of the neural network that define the conditional expectation function and adopts them to better fit the individual cluster.

[8] Indeed this language is helpful not only for understanding the model, but how it can be seemlessly evaluated and trained with little effort as highlighted in Frostig et al. [2018]

For visual clarity we assume that composition has a higher precendance than function application. Using Haskell-like notation, the fish operator denotes the composition of 'embellished' functions. The functions are embellished in the sense that not only do they return the transformed data in the case of the featureMap, or the cluster specific parameters as in clusterMap, but to each function call they also return a data dependent penalty value which is the regularization term in *regularizing the forward pass*.

While our framework is introduced and well-motivated in a cross-sectional setting with clustered treatment assignment, we apply it in this paper in a repeated cross-section setting. The advantage of our approach in this context, in contrast to a more traditional non-parametric difference-in-difference estimator, is that we explicitly correct for the presence of clustered data in a nonparametric way using pre and post-treatment information. If we let Z_i, X_i, A denote zip code, covariates, and the set of treated clusters respectively, then in a cross sectional setting, our approach allows us to estimate the following conditional expectation functions.

$$E[Y_i|X_i, C_i \in A], E[Y_i|X_i, C_i \in A^c]$$

Given this function, it is straightforward to write down sufficient conditions for identification in a repeated cross-sectional setting following Angrist and Pischke [2009] without introducing any functional form assumptions.

The remainer of the paper is organized as follows. Section 2 describes the essential institutional background of the Right to Counsel and Rapid Rehousing. Section 3 introduces the empirical framework. Section 4 sketches out a toy model, clarifying the estimand of interest in this context. Section 5 contains results and further details of the empirical framework as it relates to the repeated cross-sectional setting. Section 6 concludes.

2   Background & Policy Details

2.1   Evictions

- Grace Period

- Serve notice to quit

- Summons & Complaints

- Judgement & Resitution (Control of property is returned to the landlord

2.2 Implementation

Signed in June of 2021 by Governor Lamont, the Right to Counsel went into effect on January 31, 2022, as rental relief services in response to Covid-19 we're coming to an end, well after the expiration of the CDC's eviction moratorium for nonpayment of rent (August 26, 2021). Because the expected demand for legal services under the Right to Counsel exceed the level of legal support, state representatives rolled out the policy in phases. In the first stage, the policy was made available to a subset of the zip codes (the 14 shown in yellow) in figure 3b which make up 30% of evictions and 20% percent of the renter population (Connecticut Bar Foundation). The expectation was that 2,000 households would receive legal aid in the first phase out of the 20,000 households expected. Individuals and families within these zip codes who made 80% or less than the area median income were eligible (Reference).
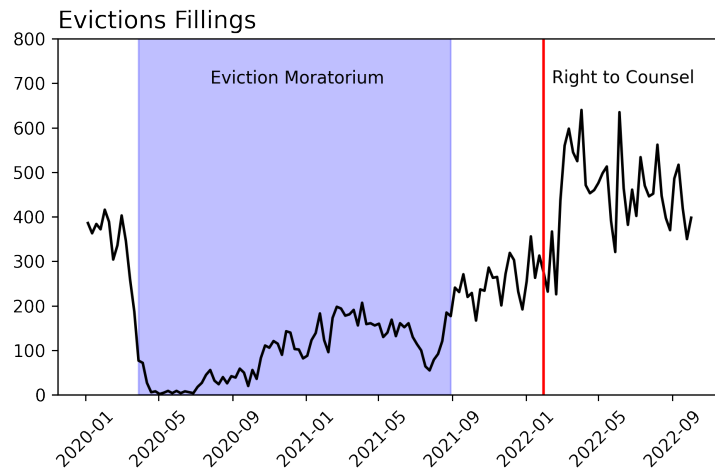

Figure 2: Weekly eviction fillings within the state of Connecticut as reported by the Eviction Lab at Princeton
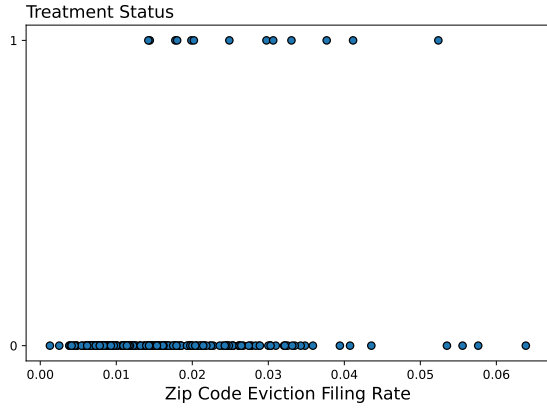
2.3 Treatment

[9] Beginning earlier, on October 1, 2021, landlords were to notify individuals of the existence of this policy when serving tenants with a notice to quit (the first step in an eviction proceeding – the entire process is described in more detail in the appendix). In addition to landlords, courts were expected to inform tenants of the policy when and if tenants appeared in court.[10][11] Given that the lack of enforcement with respect to landlords and that 37% of

---

[9]CBF Assignment Sheet– received via email on August 26th –

[10] "On and after October 1, 2021, an owner, lessor, landlord, legal representative or agent of an owner, lessor or landlord, a housing authority or a housing subsidy program administrator, as applicable, shall attach a copy of the notice described under subdivision (1) of this subsection, to (A) a notice to quit delivered to a covered individual pursuant to chapter 832 or chapter 412 of the general statutes; (B) a summons and complaint for a summary process action pursuant to chapter 832 or chapter 412 of the general statutes; (C) a lease termination notice for a public or subsidized housing unit; and (D) a notice to terminate a state or federal housing subsidy"

[11] "Any court notice scheduling a mediation or hearing that is sent to a self-represented party in a covered matter shall include plain language information about the availability of legal representation through the

Connecticut tenants fail to appear in court during eviction proceedings, it's quite possible that many eligible tenants are not yet aware of the policy's existence.[12]

(a) Eviction Filing Rates
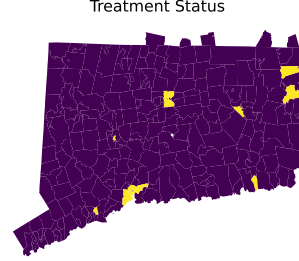
(b) Treatment status for each zip code (yellow indicates treated zip code): Reproduced Here

Figure 3: The general pattern that these figures try to highlight is that in order to fit the the 'v'-shaped valley in the function, the model overfits the tails of the function – Reproduced Here

2.4 Data & Outcomes

As referenced prior, the outcome of interest is the search length for households who are currently experiencing homelessness but face limited barriers to housing. Rapid Re-housing, a federally tracked program, provides limited short term assistance to exactly this population.[13] Its aim is to help "families exit shelters and get back into permanent housing quickly."[14] While different in nature than an independent housing search, the search length of individuals in a Rapid Rehousing program a reasonable proxy for the following three reasons. First, Rapid Rehousing programs "serve people experiencing homelessness with no preconditions such as employment, income, absence of criminal record, or sobriety."[15] Second, the program does not target people who might need long-term assistance. Those individuals and families are helped by permanent supportive housing programs.[16][18] Third, the lease agreement households sign come with "the same rights and responsibilities as a

---

right to counsel program and a phone number for accessing information and applying for assistance."

[12] According to article 37% of Connecticut tenants fail to appear in court during eviction proceedings, and their cases end in default, according to a 2021 report by the Connecticut Advisory Council on Housing Matters.

[13] As Evans et al. [2019] explains, "The Homeless Management Information System (HMIS) is a nationwide network of local IT systems that collect client-level data for people entering sheltered services. These systems are organized at the CoC level and most homeless shelters across the country are part of the HMIS system"

[14] Missing Source

[15] Reference

[16] Very different from permanent supportive housing which is as Rosanne Haggerty writes in the NyTimes, "is ideal for those with serious health challenges who have been homeless for long periods of time".[17]

[18] **Cost**: at $6,678 per family, it is cheaper than transitional housing at $32,557 per family.[19]

typical lease holder."[20]

- Evans et al. [2019]: The Family Options Study provides the primary experimental test of the effectiveness of rapid rehousing (RRH), but the results for this treatment arm are mixed. There were no statistically significant differences in housing outcomes between the control and short-term subsidy group by the three-year followup. (Gubits et al. 2016).
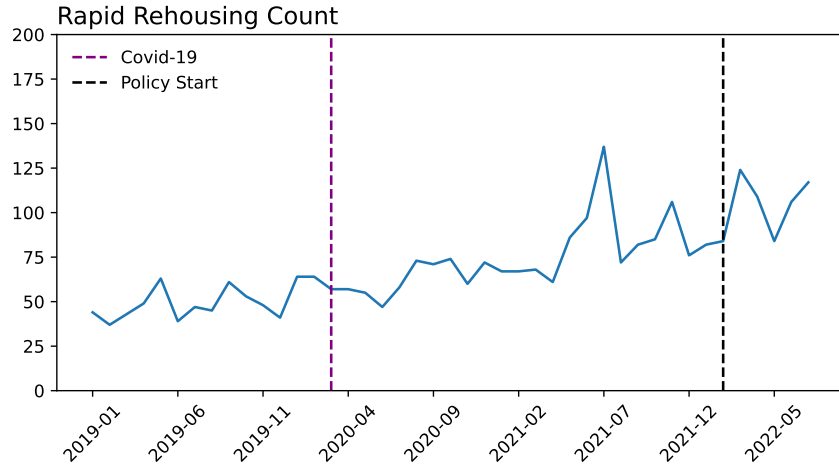

Figure 4: Treatment status for each zip code (yellow indicates treated zip code): Reproduced Here

3 Framework

3.1 Context

To keep things simple, we describe our approach in the specific context of cluster-level randomized control trials where we're interested in estimating treatment heterogeneity. Cluster-level randomized control trials are randomized control trials where treatment varies at a level above the unit of interest. Such experiments are common in development, education, and health settings because they are (A) generally easier to implement (Gerber and Green [2012]), (B) better adhere to the potential outcome framework (Imbens and Rubin [2015]) and perhaps most importantly (C) allow us to understand the the effects of scaling the treatment (List [2021]).

---

[20]It is imperative that any lease agreement provides the tenant with **the same rights and responsibilities as a typical lease holder** and that the financial terms of the lease are such that the household has a reasonable ability to assume rental costs once financial support ends (keeping in mind that in the majority of cases, even households with no income at move-in retain their housing)"

3.2 Challenge ([The Tragic Triad](#))[21]

In the absence of cluster data, under the selection on observable and stable unit treatment value assumptions, such an estimation problem is equivalent to solving the following optimization problem over the treatment and controls groups separately.

$$\inf_{f \in \sigma(X)} \mathbb{E}\big[(Y - f)^2\big]$$

Under the potential outcome framework, clustered level treatment assignment poses a few challenges because in each group: (1) we observe only a subset of the clusters, (2) the distribution of covariates can differ across clusters, and (3) the distribution of outcomes conditional on covariates may differ across clusters. These issues are perhaps only magnified as we increase the dimensionality of the data. As shown in figure 5, we extended the work of Balestriero et al. [2021] to the situation of clustered sampling. As illustrated in figure 5a, clustered sampling doesn't change the fundamental issue of learning in high dimensions (extrapolation) so much as it motivates us to reconsider how we go about it, as highlighted in figure 5b.

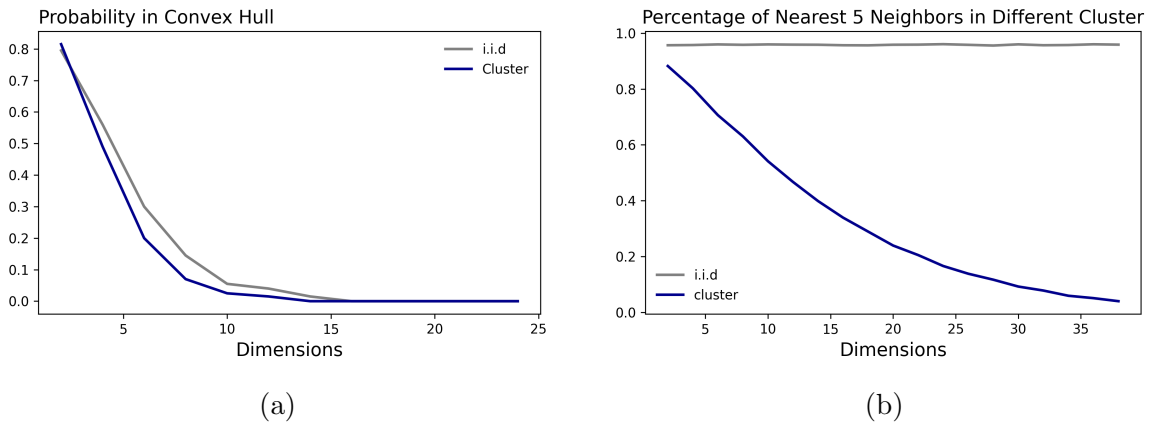(a)   (b)

Figure 5: (a) Probability that a observation in the test set is in the convex hull formed by the training set (b) Fraction of the five nearest neighbors in a different cluster. Data consists of 25 cluster and 25 observations per cluster. Clusters differ only in the mean which is drawn from multivariate normal distribution with an identity covariance matrix: Reproduced Here and Here

We can use the following motivating example to illustrate the pratical challenge. The key issue is how to locally correct for the presence of clustered data. Intuitively, when there are a lot of clusters present, we would prefer a small bandwidth. When there are few clusters present, we would prefer a larger bandwidth. Figure 6 highlights why this local correction is necessary in that in order to fit the the 'v'-shaped valley in the function, the
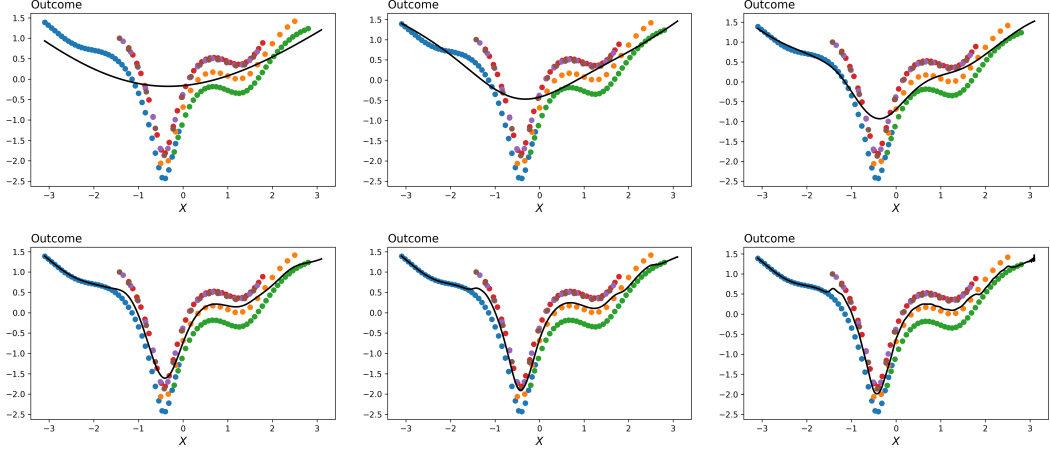
Figure 6: Reproduced Here

model overfits the tails of the function. Our approach is motivated in part by Domingos [2020] which illustrates how neural networks can be understood within the framework of a kernel machine. The insight of the work is that the kernel, the similarity measure $k(x, x')$ can be understood as the inner product of the gradient of the partially evaluated neural network a the corresponding point. While such a relationships holds exactly only in the gradient flow regiem, it suggests that we can locally correct our model for the presence of clustered data in a global manner by adjusting how we take the gradients of the model.

Path Kernel

$$K(x, x') = \int \Big\langle \nabla_\theta f_x\big(\theta(t)\big), \nabla_\theta f_{x'}\big(\theta(t)\big) \Big\rangle dt$$

Model

$$K_\theta(x, x') = \int_0^1 \Big\langle \nabla_\theta f_x\big(\theta(t)\big), \nabla_\theta f_{x'}\big(\theta(t)\big) \Big\rangle dt$$

Cluster Specific Model

$$K_{\theta_c^*(\theta)}(x, x') = K_\theta(x, x') + \int_1^{1+\varepsilon} \Big\langle \nabla_\theta f_x\big(\theta(t)\big), \nabla_\theta f_{x'}\big(\theta(t)\big) \Big\rangle dt$$

illustrates that as the learning rate of a neural network approaches the gradient flow

---

[21] The expression "tragic triad" is taken from Yu et al. [2020]

regime, the learned model can be represented by a kernel machine with the path kernel.[22]

In contrast, subject to the appropriate hyperparameters, our method produces a reasonable estimate.
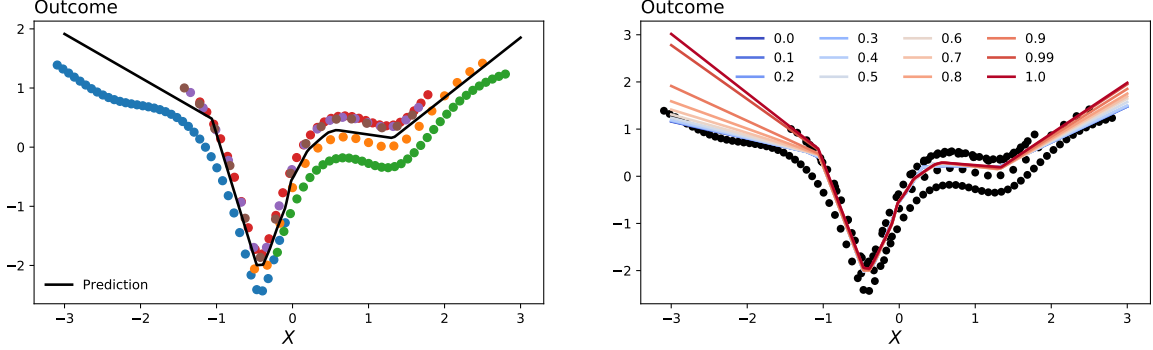

Figure 7: The general pattern that these figures try to highlight is that in order to fit the the 'v'-shaped valley in the function, the model overfits the tails of the function – Reproduced Here

3.3   Implicit Cluster Map

- In contrast to linear regression, we do it implicitly

- In contrast to econometric, machine-learning agnostic, we're trying to leverage the success of deep learning models to define what's local.

As mentioned in the introduction, our approach to modern statistical learning theory places less emphasis on understanding the statistical properties of the method (which are difficult to capture in a meaningful way) and more emphasis on the understanding the inductive bias of the algorithm. **TL;DR** In this section, we define our implicit cluster map as a regularized version of the popular gradient based meta-learning MAML.[23] While follow-up work such as Raghu et al. [2019] has shown that the "the meta-initialization already providing high quality representations", such an analysis assumes a cluster specific head of the network. We show that empirically, without this cluster specific head, the meta-initialization is not done in the right space which motivates our need to add an additional form of regularization.

As applied microeconomists, we are accustomed to writing our problem as a bi-level optimization problem so as to better distinguish between the parameters of interest and the

---

[22]Kernel Machines

$$y = g\Big(\sum a_i K(x, x_i) + b \Big)$$

[23]Model Agnostic Meta-Learning (MAML), a method that consists of two optimization loops, with the outer loop finding a meta-initialization, from which the inner loop can efficiently learn new tasks.– Raghu et al. [2019]

nuisance parameters. In this context, the nuisance parameters are cluster specific parameters that are "fit" during the inner optimization process. Below, we follow the language used by Belkin [2021] in distinguishing between classical and modern estimation techniques.

$$\mathcal{L}(\theta) := \sum_c \mathcal{L}_c(\theta), \quad \mathcal{L}_c(\theta) := G(\theta, \theta_c^*(\theta)), \quad \theta_c^*(\theta) := \operatorname*{argmin}_{\theta_c} F(\theta, \theta_c)$$

With "Classical" under-parameterized models, as in the case of linear regression, F, the clustere-specific empirical loss function is exactly what you would expect.

$$F(\theta, \theta_c^*(\theta)) := \sum_{i \in c} \left(y_i - \theta^T d_i - \theta_c^*(\theta)^T x_i \right)^2$$

With "Modern" over-parameterized models, though, like the ones that we target in this paper, we make the following adjustments. We restrict the objective function that is used to implicitly define the cluster specific maps. First, we generalize the above set-up by allowing the cluster specific parameters to be in one-to-one correspondance with the parameters of interest. Second, we constrain the implicit cluster function $\theta_c^*(\theta)$. Without some form of regularisation, given the flexibility of neural network models as illustrated in Zhang et al. [2021], the jacobian of this function can easily be zero. We illustrate this concern in figure **??**. And finally, we add a penalty term to the cluster specific loss function to ensure that adaptation happens in the right space.

Define our empirical cluster parameters in relation to the parameters of interest

$$G(\theta) := \sum_{i \in c} \left(y_i - f(\theta_c^*(\theta), x_i) \right)^2$$

Implicit Cluster Map

$$\theta_c^*(\theta) := \theta^t - \alpha \nabla_\theta G(\theta^{t-1}), \quad \theta^0 = \theta$$

As well as introduce an auxilliary term to the cluster specific loss function

$$\mathcal{L}_c(\theta) := G(\theta) + H(\mathrm{Path}(\theta, \hat{\theta}_c^*(\theta)))$$

We illustrate the relative importance of our design choices in figure 15. In figure 15a we see that supverized learning overfits to the clusters in the tails of their distributions. In figure 15b, we see that MAML meta initialization is in the wrong space. And finally in figure 15c, we see that a regularized version of MAML appears to learn something reasonable.
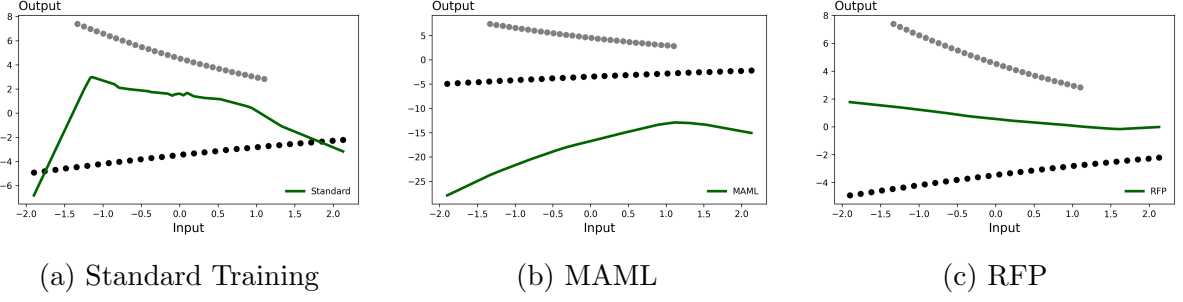
(a) Standard Training   (b) MAML   (c) RFP

Figure 8: Reproduced Here: The grey and black dots represent data from separate clusters. Each figure corresponds to fitting a neural network to this data under different training algorithms
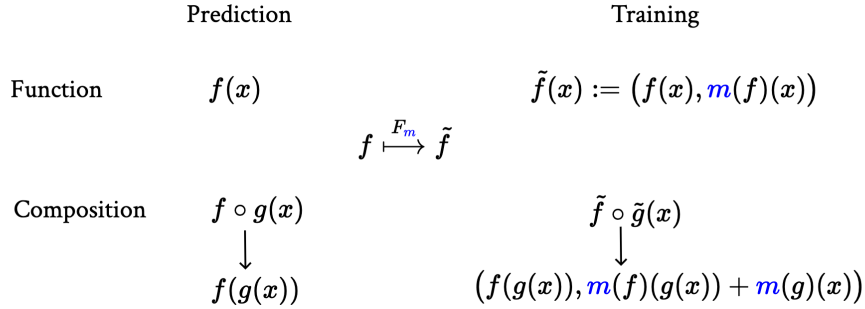

Figure 9: High Level Summary

Such an approach is inspired/related by proximal methods (Parikh et al. [2014]).[24] Which can be thought of as generalized projections, as well as gradient steps.

4 Toy Model

We write down the following model in order to clarify our estimand of interest. The key detail is that we parameterize the "consumer's" constraint function with a variable that denotes the status of the Right to Counsel (where 1 means in the policy is in effect, and 0 means it is not). We allow the parameters of the constraint and utility functions to differ across individuals, and define our estimand as the expected value taken with respect to these the probability law induced by these random variables. **Notationally**, we denote the partial evaluation of a function by subscripting the function with the argument.

We define an element of the choice set as a bundle of housing related features. The first component we include, which is the most relevant to our dataset, is housing search length. It's important to acknowledge that individuals may be substituting across a wide range of housing aspects (such as the price of the house and the quality of the house) among other features. For the purposes of this model, we omit these, as we do not observe these features

---

[24]Note that the codomain of f is taken to be $\mathcal{R} \cup \{\infty\}$, while the domain are set of inputs that take finite values.

in the data.

$$x := (\text{search period}, \text{quality}, \text{price}, \dots) \in \mathcal{X}$$

As indicated by its signature, we introduce a parameterized constraint function where we allow the parameters to vary across individuals. This constraint function not only incorporates the macroeconomic features of the housing market, but also the effects of the landlords actions/policies towards individual tenants.

$$F :: \text{Params} \rightarrow \text{RTC} \rightarrow \mathcal{X} \rightarrow \{0, 1\}$$

The utility function then captures the value of each housing option.

$$U :: \text{Params} \rightarrow \mathcal{X} \rightarrow \mathcal{R}_+$$

With the essential components of the model defined, we can express the individual choice problem as follows. The following problem would model an individual not in a treated zip code or prior to the policy implementation as the second parameter of the constraint function takes the value 0. The pre-image of 0 under the constraint function defines the feasibility set.

$$\underset{x \in F_{p,0}^{-1}(0)}{\text{maximize}} \; U_\alpha(x)$$

Via-this parameterized optimization problem, we can then define the implicit function of interest which captures the effects of the individual-level parameters and the policy on the search length.

$$s^*(\text{rtc}, p, \alpha) := \underset{x \in F_{p,\text{rtc}}^{-1}(0)}{\text{argmax}_0} \; U_\alpha(x)$$

As we are only interested in the later effects, we integrate over the individual-level parameters leading to the estimand defined below.

$$\theta := \mathbb{E}[s^*(1, p, \alpha) - s^*(0, p, \alpha)]$$
$$= \int_\Omega s^*(1, p, \alpha) - s^*(0, p, \alpha) d\mathbb{P}$$

5 Empirical Strategy

The following notation defines the key variables used in the estimators defined below.

$$Y_i : \text{Search Duration}$$
$$X_i : \text{Age, Gender, Race, Family Size}$$
$$Z_i : \text{Zip Code}$$
$$D_i : \text{Treated Zip Code}$$

Note: Subscripts on the outcome corresponding to subsets of individuals who are observed in that corresponding time period.

5.1 Difference-in-Difference

We fit the following difference-in-difference estimator. In this specification and for the ones that follow, Y_i is a binary variable that indicates whether the search length was less than some specified number of days. We run multiple regressions where we vary what this cutoff is, be it ten, twenty, or sixty days. These results are illustrated in figure 10, where we run a number of Difference-in-Difference regressions by changing the cutoff point and graphing the estimates. As we can see the percentage point effect of the Right to Counsel (the y-axis) is relatively constant as we vary the acceptable move-in-date threshold.

$$\beta_0 = \mathbb{E}[Y_1 - Y_0 \mid D = 1] - \mathbb{E}[Y_1 - Y_0 \mid D = 0]$$

Importantly, when we restrict the sample to only high eviction zip codes in figure 10b, we see a relatively consistent negative effect on the probability of moving into housing by a certain date.
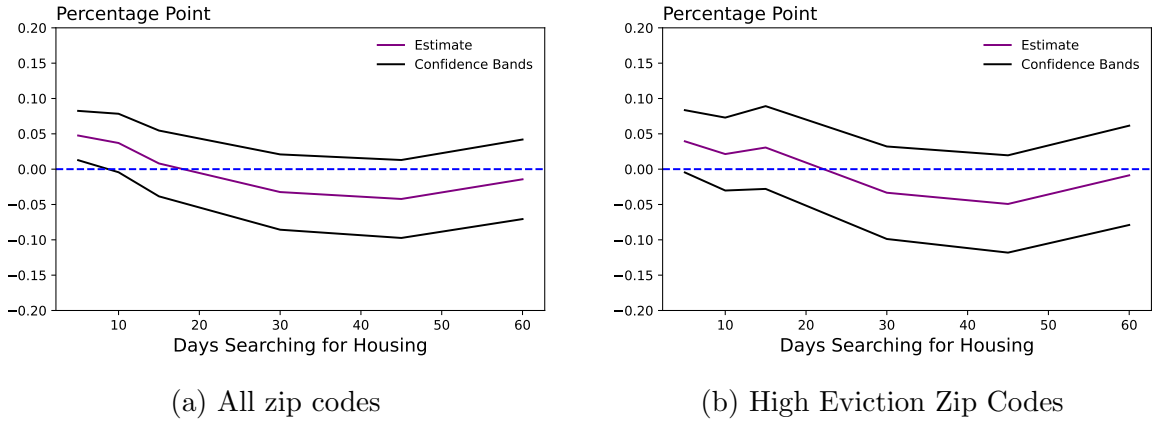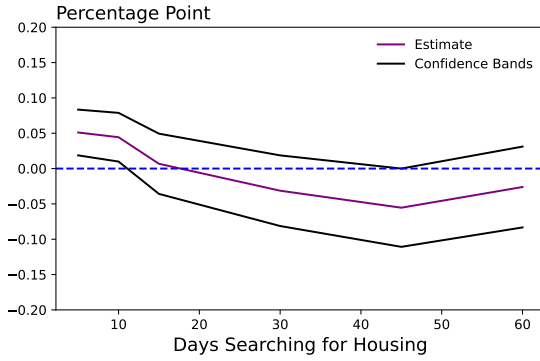

(a) All zip codes (b) High Eviction Zip Codes

Figure 10: Reproduced Here: Confidence Bands formed via stratified bootstrapped sampling without replacement with a 75% sampling rate.)
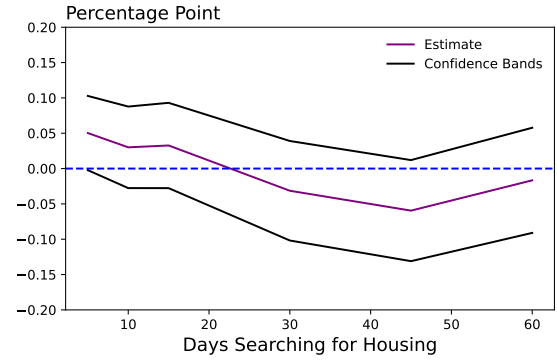
5.2 Difference-in-Difference with Controls

Adding individual level controls as well as zip code fixed effects makes a material difference for the magnitude of the effect but not the sign when we again restrict the sample to high eviction zip codes in figure 11b.

$$Y_i = \alpha_0 + \beta_0 \text{Post}_i \times \text{Treated}_i + \beta_1 \text{Post}_i + \beta_2 \text{Treated}_i$$
$$+ \beta_3 X_i + \beta_4 Z_i + \varepsilon_i$$

(a) All zip codes　　　　(b) High Eviction Zip Codes

Figure 11: Reproduced Here: Confidence Bands formed via stratified bootstrapped sampling without replacement (75%)

5.3 Regularizing the Forward Pass

It's possible to re-write the difference-in-difference model with controls as a linear version of the following estimator. Doing to we see that there are two potential concerns. The first is that the model fails to correct for the propensity score which as highlighted in figure ?? can be problematic in certain contexts.

$$\tilde{Y}_i - \mathbb{E}[\tilde{Y}_i | X_i] = \beta_0 \big(D_i - \mathbb{E}[D_i | X_i] \big) + \varepsilon_i, \quad \text{where } \tilde{Y}_i = Y_{1i} - Y_{0i}$$

The natural correction is to then deploy a fully-nonparametric model.

$$\beta_0 := \int \mathbb{E}[\tilde{Y}_i | X_i, D_i = 1] - \mathbb{E}[\tilde{Y}_i | X_i, D_i = 0] d\mathbb{P}_X$$

As previously illustrated, though, when treatment varies at the cluster level such estimators are susceptible to the tragic triad. We therefore implicitly partiall out the cluster effects by

fitting these conditional expectations using our regularized forward pass framework which generates the results captured in figure 13. Note, in order to limit the computation burden of these models we fit the following model. The only difference is that it removes the first step of estimating $\tilde{Y}$.

$$\beta_0 := \int d\mathbb{P}_X \Big(\big(\mathbb{E}[Y_1|X, D=1] - \mathbb{E}[Y_0|X, D=1]\big) - \big(\mathbb{E}[Y_1|X, D=0] - \mathbb{E}[Y_0|X, D=0]\big) \Big)$$
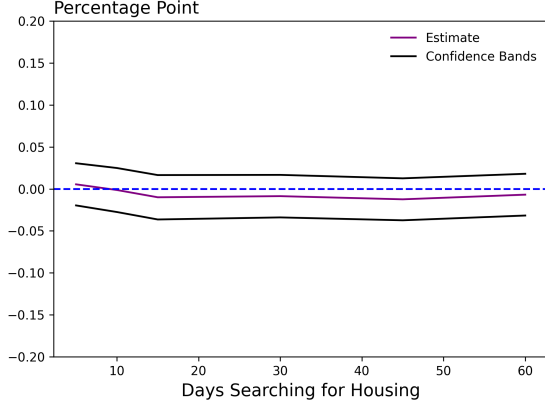
Cross-Sectional Setting

$$\begin{aligned}
\hat{\theta}_t(X) &:= \mathbb{E}[Y_{it}|X_{it}, D_{it}=1, C_{it} \in A] - \mathbb{E}[Y_{it}|X_{it}, D_{it}=0, C_{it} \in A^c] \\
&= \mathbb{E}[Y_{it}(1)|X_{it}, C_{it} \in A] - \mathbb{E}[Y_{it}(0)|X_{it}, C_{it} \in A^c]
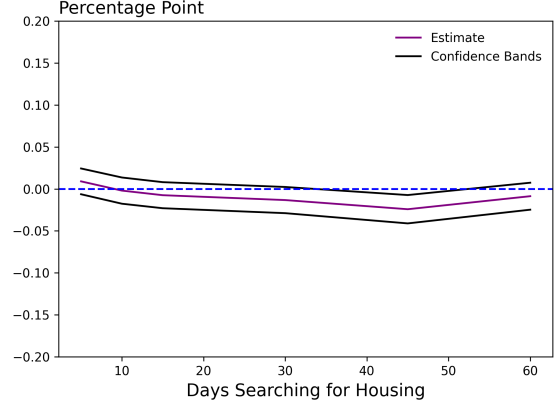\end{aligned}$$

Repeated Cross-Sectional Setting

$$\begin{aligned}
\hat{\theta}(X) &:= \hat{\theta}_t(X) - \hat{\theta}_{t-1}(X) \\
&= \mathbb{E}[Y_t(1) - Y_{t-1}(0)|X, C \in A] - \mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, C \in A^c] \\
&= \mathbb{E}[Y_t(1) - Y_t(0)|X, C \in A] \\
&\quad + \underbrace{\mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, C \in A] - \mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, C \in A^c]}_{\approx 0} \\
&= \mathbb{E}[Y_t(1) - Y_t(0)|X, C \in A], \qquad \text{Local Treatment Effect on the Treated}
\end{aligned}$$

$$\hat{\theta} := \int \hat{\theta}(X) d\mathbb{P}_X, \qquad \text{Average Treatment Effect on the Treated}$$
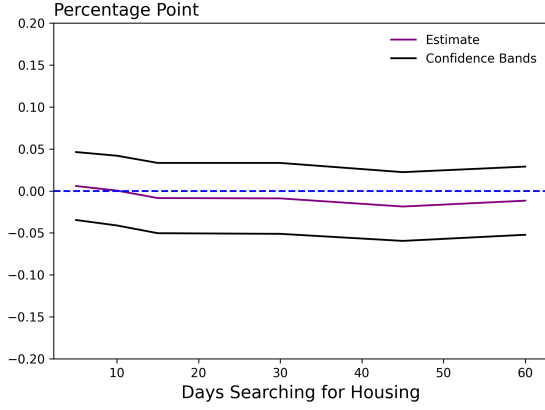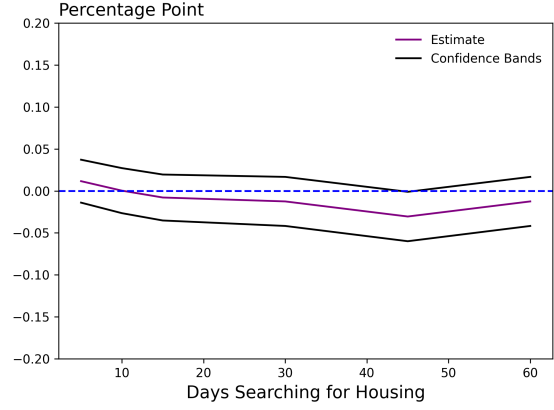
(a) All zip codes   (b) High Eviction Zip Codes

Figure 12: Reproduced Here: Confidence Bands formed sampling across random parameter initializations

(a) All zip codes   (b) High Eviction Zip Codes

Figure 13: Reproduced Here: Confidence Bands formed sampling across random parameter initializations

While the magnitude of the effect shrinks, the sign of the effect is relatively constant and negative when restricting the sample to high eviction zip codes. We also note that the difference in the estimated effects when considering the all zip codes and only high eviction zip codes decrease.

Over a long time frame, it would be interesting to assess whether the tail of the distribution increased. As Evans et al. [2019] notes in summaring previous homelessness related work, the costs are often associated with those in the tail.[25]

$$\tilde{Y}_i - \mathbb{E}[\tilde{Y}_i|X_i] = \beta_0(D_i - \mathbb{E}[D_i|X_i]) + v_i, \quad \tilde{Y}_i = Y_i - \mathbb{E}_{\text{pre}}[Y_i|X_i, Z_i]$$

---

[25]Evans et al. [2019] notes that In Spellman et al. (2010), the costliest 10 percent of people incur up to 83 percent of the total costs of shelter during homelessness. In Flaming et al. (2015), individuals with costs in the top 5 percent of the public costs incur 47 percent of all costs. Hence, whether a program targets the costliest cases matters significantly for the mean costs reported above."

Extending the regression model in a nonparametric way is straightforward: simply integrate the transformed outcome across the features/covariates within each treatment group and take the difference. Presented in this way, it's clear that this approach only corrects for the presence of clustered data by using pre-treatment information and leaves to the post-treatment information on the table. Our framework can leverage both sources of information. While we leave a technical description of the framework to section 3 of the paper

$$\beta_0 := \int \mathbb{E}[\tilde{Y}_i|X_i, D_i = 1] - \mathbb{E}[\tilde{Y}_i|X_i, D_i = 0]d\mathbb{P}_X$$

6 Conclusion

- Reference Additional renter protections, such as right to counsel, expungement of eviction records, and just-cause eviction standards, are needed to help protect renters now and in the long term.

The recent events of Covid-19 and rising inflation have magnified the importance and fragility of housing for low-income individuals. Indeed homelessness itself has become a staple in the news and political conversations (NyTimes). In response to this, we empirically assess the effectiveness of an initiative, growing in popularity across the U.S., known as the Right to Counsel (RTC). Currently adopted in 15 cities and 3 states, the initiative aims to combat the 3.6 million eviction fillings that occur each year in the U.S. by closing the gap in legal representation between landlord and tenant. Closing such a gap entails providing legal representation for tenants which as the case of Boston in 2019 which saw 39,594 eviction cases filed only with only 8.7% of tenants represented.[26] Lastly, while no pre-analysis plan accompanies this paper, the only source of heterogeneity explored is the effects of the policy on Black and female tenants. As well documented in the eviction literature, these subgroups share the greatest likelihood and costs of evictions, as Desmond [2019] writes, "Low-income women, especially poor black women, are at high risk of eviction", and Collinson et al. [2022] notes that with regards to the costs of evictions, "We find particularly sharp negative impacts for female and Black tenants,[27] who drive the effects on labor market outcomes, residential mobility, and interactions with homelessness." It seems likely, therefore, that if landlords respond in an adverse way to The Right to Counsel, it would be towards this sub-population in particular. Hence, all regression specifications are fit both over the entire sample and this sub-sample of interest. As a general statement, Economists are interested in understanding the effects of policies at scale. Almost by definition, though, these effects are not well identified. The aim, therefore, is to capture a particular effect of the policy as it's in the process of being deployed with the hope that this intermediate measurement might be informative about the effect under the new equilibrium.

---

[26](Samuelson et a. [2020])

[27]Evans et al. [2019] writes '

In this paper, we take as our intermediate measurement the housing search length for low-income individuals. As explained in the body of the paper, this is a far from perfect or comprehensive outcomes variable. That said, we regard as a meaningful signal in the sense that if there are adverse effects of the policy – if landlords decide to re-optimize in an adverse fashion in response to the Right to Counsel – we would likely see it via the housing search channel.

And indeed, while extremely preliminary, we see some indication that there may be adverse effects along this channel. Across a series of estimators, each of which fits within a general empirical estimation structure of "Regularizing the Forward Pass" which we introduce in this paper, we see that search length increases.

References

Boaz Abramson. The welfare effects of eviction and homelessness policies. 2021.

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press, 2009.

Randall Balestriero, Jerome Pesenti, and Yann LeCun. Learning in high dimension always amounts to extrapolation. *arXiv preprint arXiv:2110.09485*, 2021.

Mikhail Belkin. Fit without fear: remarkable mathematical phenomena of deep learning through the prism of interpolation. *Acta Numerica*, 30:203–248, 2021.

Michael T Cassidy and Janet Currie. The effects of legal representation on tenant outcomes in housing court: Evidence from new york city's universal access program. Technical report, National Bureau of Economic Research, 2022.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018.

Robert Collinson, John Eric Humphries, Nicholas S Mader, Davin K Reed, Daniel I Tannenbaum, and Winnie van Dijk. Eviction and poverty in american cities. Technical report, National Bureau of Economic Research, 2022.

M Desmond. Unaffordable america: poverty, housing, and eviction. fast focus 22-2015, 2019.

Pedro Domingos. Every model learned by gradient descent is approximately a kernel machine. *arXiv preprint arXiv:2012.00152*, 2020.

Ingrid Gould Ellen, Katherine O'Regan, Sophia House, and Ryan Brenner. Do lawyers matter? early evidence on eviction patterns after the rollout of universal access to counsel in new york city. *Housing Policy Debate*, 31(3-5):540–561, 2021.

William N Evans, David C Philips, and Krista J Ruffini. Reducing and preventing homelessness: A review of the evidence and charting a research agenda. 2019.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017.

Roy Frostig, Matthew James Johnson, and Chris Leary. Compiling machine learning programs via high-level tracing. *Systems for Machine Learning*, 4(9), 2018.

Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation.* WW Norton, 2012.

D James Greiner, Cassandra Wolos Pattanayak, and Jonathan Hennessy. The limits of unbundled legal assistance: a randomized study in a massachusetts district court and prospects for the future. *Harv. L. rev.*, 126:901, 2012.

Andreas Griewank and Andrea Walther. *Evaluating derivatives: principles and techniques of algorithmic differentiation.* SIAM, 2008.

Steven Gunn. Eviction defense for poor tenants: Costly compassion or justice served. *Yale L. & Pol'y Rev.*, 13:385, 1995.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press, 2015.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, 31, 2018.

Jacob Kelly, Jesse Bettencourt, Matthew J Johnson, and David K Duvenaud. Learning differential equations that are easy to solve. *Advances in Neural Information Processing Systems*, 33:4370–4380, 2020.

John List. The voltage effect in behavioral economics. Technical report, The Field Experiments Website, 2021.

Vaishnavh Nagarajan and J Zico Kolter. Uniform convergence may be unable to explain generalization in deep learning. *Advances in Neural Information Processing Systems*, 32, 2019.

Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239, 2014.

Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. *arXiv preprint arXiv:1909.09157*, 2019.

Carroll Seron, Martin Frankel, Gregg Van Ryzin, and Jean Kovath. The impact of legal counsel on outcomes for poor tenants in new york city's housing court: results of a randomized experiment. *Law and Society Review*, pages 419–434, 2001.

Andrew Gordon Wilson. The case for bayesian deep learning. *arXiv preprint arXiv:2001.10995*, 2020.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in Neural Information Processing Systems*, 33:5824–5836, 2020.

Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3):107–115, 2021.

7 Appendix

7.1 Regularized ODE

$$\text{prox}_{\lambda,f}(v) := \underset{x}{\text{argmin}} \left(f(x) + \frac{1}{2\lambda}\|x - v\|^2 \right)$$

7.2 Inductive Bias

7.3 Implementation Details

- ("Income-eligible" means (A) having household income at or below eighty per cent of the state median income adjusted for family size, as determined by the United States Department of Housing and Urban Development, at the time of the request for representation; or (B) receiving one of the following types of public assistance: (i) Temporary Assistance for Needy Families, (ii) Supplemental Nutrition Assistance Program benefits, (iii) Medicaid, (iv) Supplemental Security Income, (v) refugee resettlement benefits, (vi) rental assistance under chapter 138a of the general statutes, or (vii) the federal Housing Choice Voucher Program, 42 USC 1437f(o);
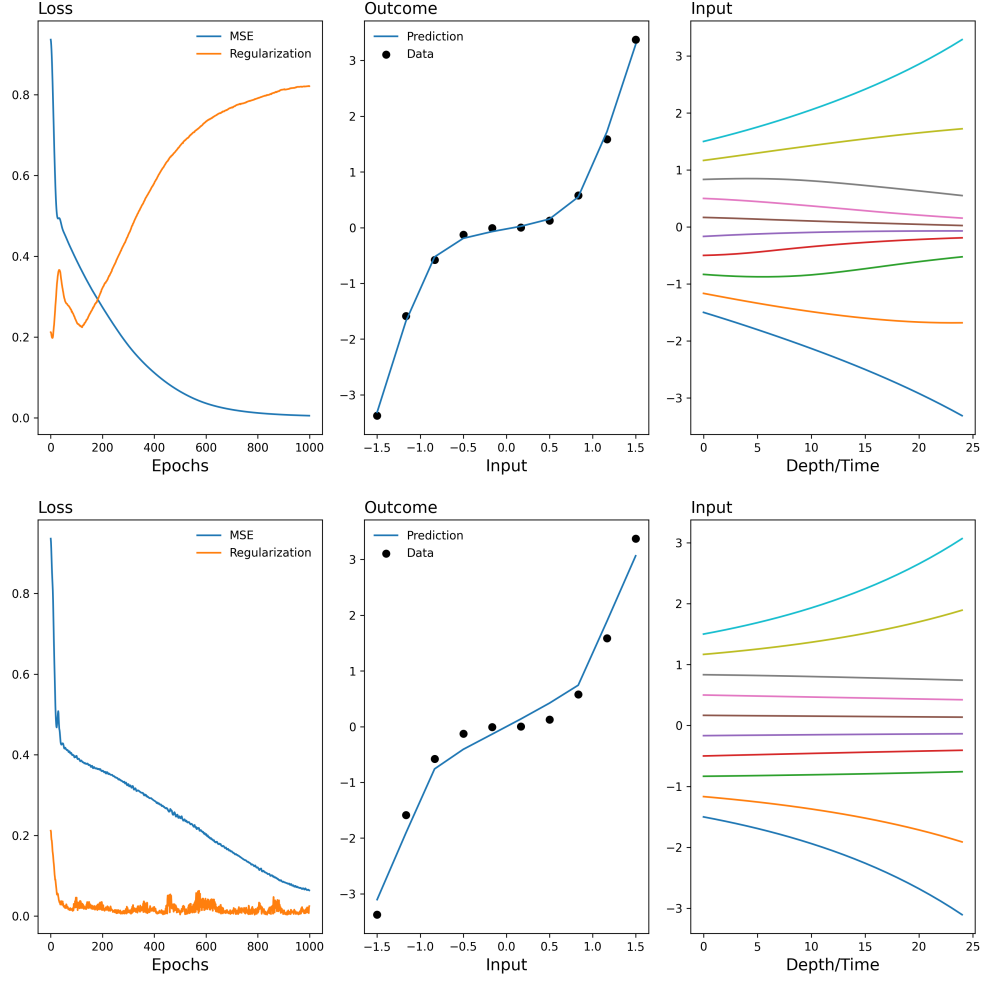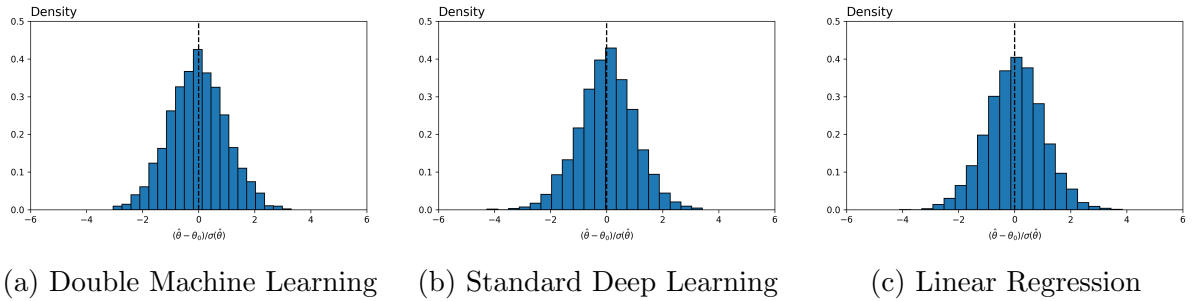
Figure 14: High Level Summary

(a) Double Machine Learning
(b) Standard Deep Learning
(c) Linear Regression

Figure 15: Reproduced Here: The grey and black dots represent data from separate clusters. Each figure corresponds to fitting a neural network to this data under different training algorithms

23

7.4 Detailed Timeline

- Reference: The CDC eviction moratorium took effect September 4 and was initially set to expire on December 31. Congress extended the moratorium through January 2021 – and President Biden further extended it through March, June, and July – and provided a total of $46.5 billion for emergency rental assistance (ERA). The eviction moratorium lapsed on July 31, but the CDC announced on August 3 a limited eviction moratorium through October 3 for renters living in communities experiencing a surge in COVID-19 cases, covering an estimated 80day after the Biden administration announced additional steps it will take to protect renters and prevent evictions during the pandemic, including those recommended by NLIHC and the National Housing Law Project.

- Reference The Centers for Disease Control and Prevention (CDC) took unprecedented action on September 1 by issuing a temporary national moratorium on most evictions for nonpayment of rent to help prevent the spread of coronavirus. Citing the historic threat to public health posed by coronavirus, the CDC declared that an eviction moratorium would help ensure people are able to practice social distancing and comply with stay-at-home orders. The CDC issued on October 9 guidance creating new burdens for renters seeking moratorium protections. While the new guidance does not rescind the moratorium on most evictions for nonpayment of rent, it states that landlords may challenge tenant declarations and initiate eviction proceedings at any time. The new guidance undermines the intent of the CDC's order by eroding protections for renters and making it more difficult for struggling renters to remain stably housed.

- CDC argues coronavirus presents a historic threat to public health and eviction moratoriums can be an effective public health measure to prevent the spread of coronavirus. An eviction moratorium is necessary to ensure people are able to follow best practices recommended by the CDC to cut down on coronavirus transmission, including quarantine, isolation, and social distancing.

- Evictions – even a single eviction filing – creates a longterm mark on a renter's record, hindering their access to future housing opportunities.