A Deep Learning Assessment of the Right to Counsel

Patrick Power, Shomik Ghosh and Markus Schwedeler
DRAFT

October 31, 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, which suggests that the policy is less effective at scale than previously understood.

Keywords: Right to Counsel, Evictions, Deep Learning

---

[1]To elaborate on this point, indivduals in the data are 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 provided with a public attorney. A gap in legal representation, therefore, exists between landlords and tenants which Collinson et al. [2022] has documented to be as large as $95\% - 1\%$ in some areas in favor of the landlord. Given the severity of these costs, the multitude of factors which contribute to an eviction filling, and the typical manner in which eviction cases are settled, many in the U.S. believe that free legal counsel should be provided to households facing eviction. Over the past five years, 15 cities and 3 states have attempted to close this gap in representation by initiating a Right to Counsel for low-income households in the hope improving 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 concern, in this context, though is that the indirect effects associated with the policy might diminish or possibly 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] The concern, therefore, is that by increasing the cost of evicting s 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 that are *not currently housed* 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 of (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 housing search duration 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

---

[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."

that cannot be addressed in this context. This paper should be understood as the first empirical microeconomic work to consider the unintended consequences of this policy, but given the importance of this topic, it certainly will not be the last.

There are two distinct reasons that explain why there has been little empirical work on this issue. The first is that, as Abramson [2021] writes, until recently, 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 same research question to ours, but written one year prior, Abramson [2021] studies the effects of the Right to Counsel via a counterfactual analysis.[4]

Only within the past year, as Washington and Connecticut adopted the policy 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 a subset of zip codes currently receive free legal aid under the policy in Connecticut. Meaningfully for our paper, several high eviction rate zip codes remain untreated as highlighted in figure 1a.

The second reason the question remains open 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, therefore,that landlords might respond to the Right to Counsel in an informal manor. Indeed from conversations with lawyers who specialize in evictions, the question seemed not to be whether the costs of the Right to Counsel would be passed on to the tenants, but rather would we be able to capture it with data. As one lawyer wrote to us, "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 find and maintain permanent 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-

---

[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

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

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 household level). In a cross-sectional setting, this presents a problem because there is no within cluster variation to exploit. Instead of fitting within clusters, the econometric models must partial out the cluster effect. 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 the deep learning model.

Deep Learning models are a subset of machine learning models that can be constructed by composing parameterized functions and trained via gradient descent. Typical models consists of 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 the gradients are often computed "automatically" via reverse mode automatic differentiation (Griewank and Walther [2008]) and the entire training process can be compiled and scaled across hardware accelerators like GPUs and TPUs.

Taken these models as given, this paper seeks to introduce an empirical estimation framework, that can seemlessly account for the presence of clustered data. As Belkin [2021] writes,

Our deep learning framework works with functions defined at a higher level of abstraction. We want a model that can account for the presence of clustered data, that has a clear, interpretable inductive bias, and scales with computation.

We compose regularized neural ordinary differential equation models with a regularized bi-level optimization procedure. Doing so our framework nests linear regression, and seemlessly allows for nonparametric cluster effects. The functions that make up our framework are defined at a higher level of absraction than the maps that define the forward pass of the a feed-forward neural network.

Importantly, though, they retains this interpretation of composition even under regularization. This is important because, 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.[6] 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

---

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)"

[6]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]

approach within the kernel method framework which in contrast to deep learning is well developed and well understood.

The key ideas of the framework 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 composition, even under regularization. This last point is best understood by considering figure 7. Note for visual clarify we assume that composition of functions is of higher precedence than function application.

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

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

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

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

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

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

For visual clarity we assume that composition has a higher precendance than function application.

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 D_i, C_i, X_i, A denote treatment status, cluster, covariates, and the set of treated clusters respectively, then in a cross sectional setting, our approach allows us to estimate the following conditional expectation function.

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

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.
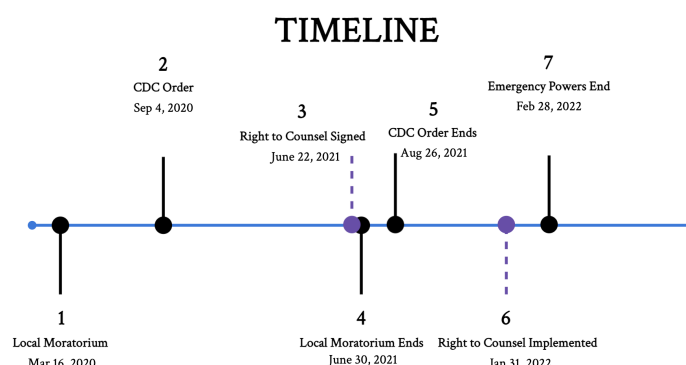
---

[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.

We organize the paper in the following way. 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 Policy Details

2.1 Timeline

In June of 2021, Governor Lamont of Connecticut signed into law the Right to Counsel (3). As highlighted in the timeline below, this event coincides with the end of the local moratorium on evictions (4) in Connecticut. Six months later, on January 31, 2022, as Covid-19 relief was winding down, the Right to Counsel went into effect. A detailed description of the timeline is provided in the appendix, but key dates of interest, including the start and end of local and national moratoriums, are illustrated below.

TIMELINE

2	7
CDC Order	Emergency Powers End
Sep 4, 2020	Feb 28, 2022

3
Right to Counsel Signed
June 22, 2021

5
CDC Order Ends
Aug 26, 2021

1
Local Moratorium
Mar 16, 2020

4
Local Moratorium Ends
June 30, 2021

6
Right to Counsel Implemented
Jan 31, 2022

2.2 Implementation

Because the expected demand for legal services under the Right to Counsel exceed the current level of legal support, state representatives decided to implement the policy in phases.[8] In the first stage, the policy was made available to a subset of the zip codes (the 14 shown in yellow) which correspond to 30% of evictions and 20% percent of the renter population.[10] Individuals and families within these zip codes who made 80% or less than the area median income were eligible for legal support starting on January 31, 2022.[11] Beginning earlier, on

---

[8]During the two years leading up to the pandemic, Connecticut saw 20,000 eviction fillings on average[9]

[10]CBF Assignment Sheet– received via email on August 26th –Expected to assist 2,000 households in the first phase, in a state where 20,000 residents faced eviction

[11]("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

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.[12][13] Given that the lack of enforcement with respect to landlords and that 37% of 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.[14]
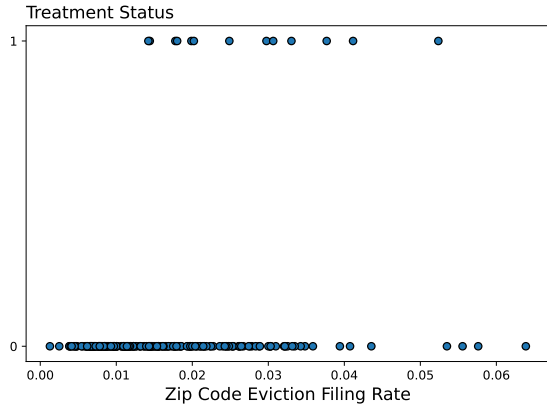
(a) Eviction Filing Rates
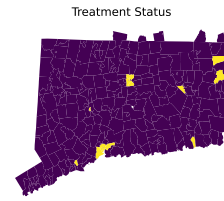
(b) Treatment status for each zip code (yellow indicates treated zip code): Reproduced Here

Figure 1: 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

2.3   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 popu-

---

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);

[12] "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"

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

[14] 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.

lation.[15] Its aim is to help "families exit shelters and get back into permanent housing quickly."[16] 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."[17] Second, the program does not target people who might need long-term assistance. Those individuals and families are helped by permanent supportive housing programs.[18][20] Third, the lease agreement households sign come with "the same rights and responsibilities as a typical lease holder."[22]

- 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).
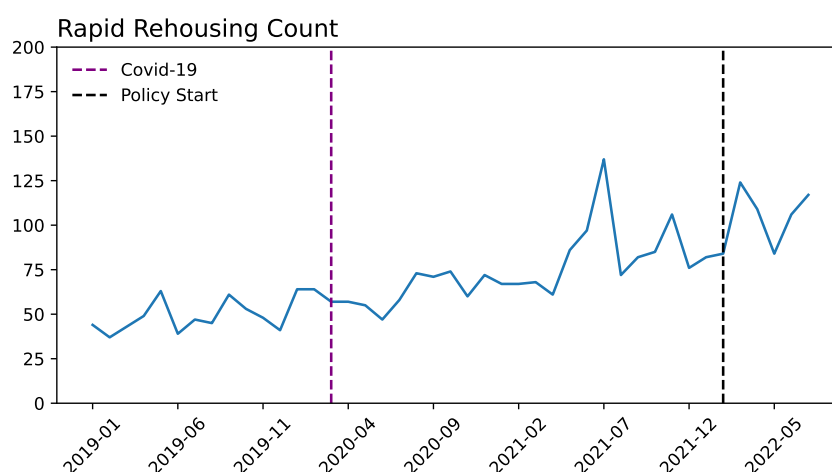

Figure 2: Treatment status for each zip code (yellow indicates treated zip code): Reproduced Here

---

[15]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"

[16]Missing Source

[17]Reference

[18]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".[19]

[20]**Cost**: at $6,678 per family, it is cheaper than transitional housing at $32,557 per family.[21]

[22]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 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.[23] Such experiments are common in development, education, and health settings because they are (A) generally easier to implement, (B) better adhere to the potential outcome framework[24] and perhaps most importantly[25] (C) allow us to understand the the effects of scaling the treatment.[26] With a binary treatment variable, such a problem can be decomposed into two separate problems where the objective function is minimized separately over the treatment and control groups.

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

3.2 Challenge (The Tragic Triad)[27]

Under the potential outcome framework, clustered level treatment assignment can be roughly thought of as forming the treatment and controls groups via random clustered sampling. From an estimation standpoint, this poses a few challenges because in each treatment group: We observe only a subset of the clusters; The distribution of covariates can differ across clusters; The distribution of outcomes conditional on covariates may differ across clusters. The above issues are perhaps only magnified as we increase the dimensionality of the data.

As shown in figure 3, we extended the work of Balestriero et al. [2021] to the situation of clustered sampling. As illustrated in figure 3a, clustered sampling doesn't change the fundamental issue of learning in high dimensions: extrapolation. It does, however, as indicated in figure 3b, suggest that we may need to reconsider how we go about learning in this context

The central challenge is how to incorporate a cluster indicator in the training phase so that the function adaptively pools information across clusters, without using the cluster indicator in the inference phase. To highlight this, we construct a toy data set where the average within cluster outcome value is zero (i.e. adding cluster specific fixed effects

---

[23]Cluster-level randomized control trials are randomized control trials where treatment varies at a level above the unit of interest

[24]Reduce the chance of spillover effects between treated and non-treated individuals.

[25]See John Lists's book, 'The Voltage Effect' which highlights this importance in great detail

[26]Many large scale studies such as HIE prefer to include many control variables in their regression specification: size of family, age categories, education level, income, self-reported health status, and use of medical care in the year prior to the start of the experiment, kind of insurance (if any) the person had prior to the experiment, whether family members grew up in a city, suburb, or town, and spending on medical care and dental care prior to the experiment

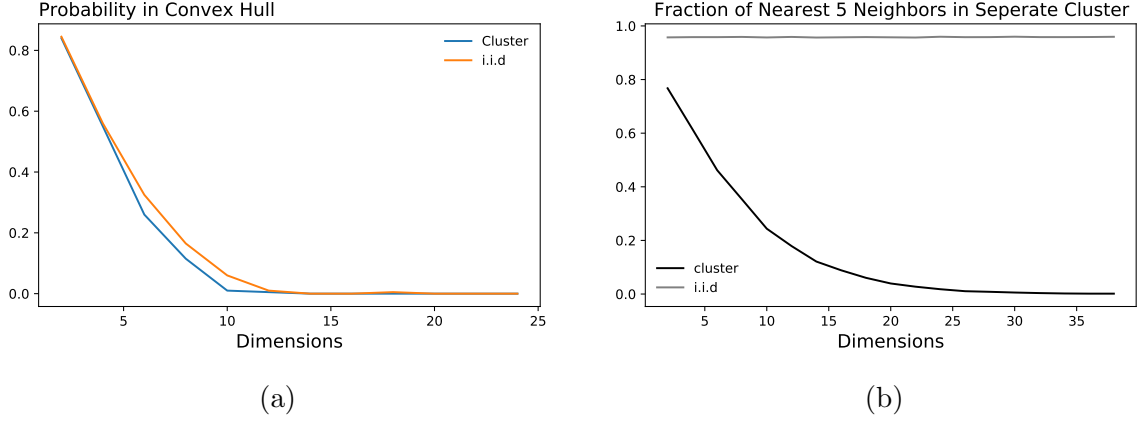[27]The expression "tragic triad" is taken from Yu et al. [2020]

Figure 3

would not improve the fit to the data). The central challenge is how "addaptively" share information across clusters. That is, when there are a lot of clusters present, we would intuitively prefer a small bandwidth. When there are few clusters present, we would prefer a larger bandwidth. And of course, we would like to extend this to high dimensions.
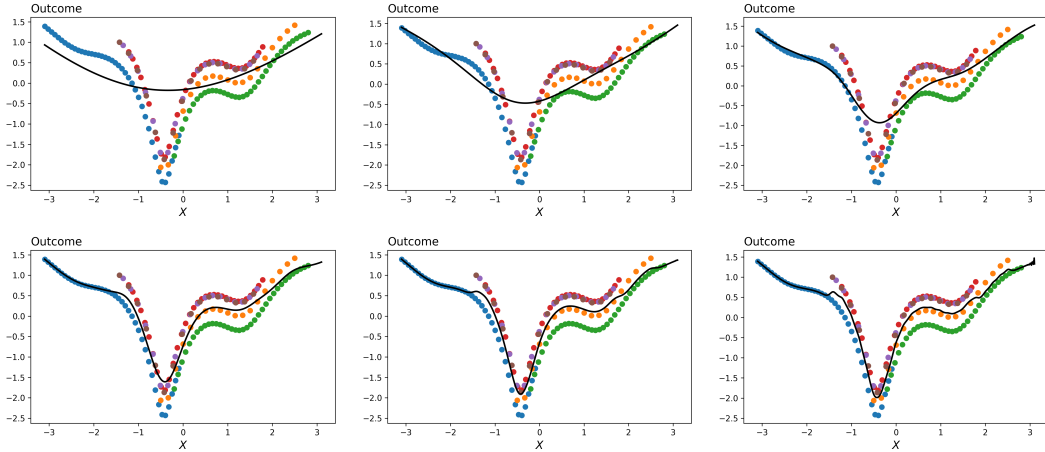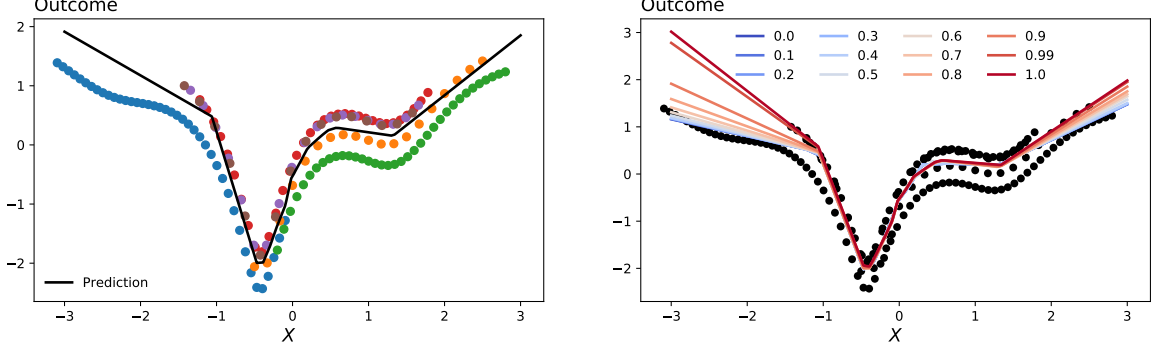

Figure 4: The general pattern that these figures try to highlight is that in order to fit the the 'v'-shaped valley in the function, the model overfits the tails of the function – Reproduced Here

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

3.3   Implicit Cluster Map

TL;DR In this section, we define our implicit cluster map as a regularized version of the popular gradient based meta-learning MAML.[28] While follow-up work such as Raghu et al.

---

[28]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]

Figure 5: 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

[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 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.

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

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 13. 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

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

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

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

We illustrate the relative importance of our design choices in figure 6. In figure 6a we see that supverized learning overfits to the clusters in the tails of their distributions. In figure 6b, we see that MAML meta initialization is in the wrong space. And finally in figure 6c, we see that a regularized version of MAML appears to learn something reasonable.
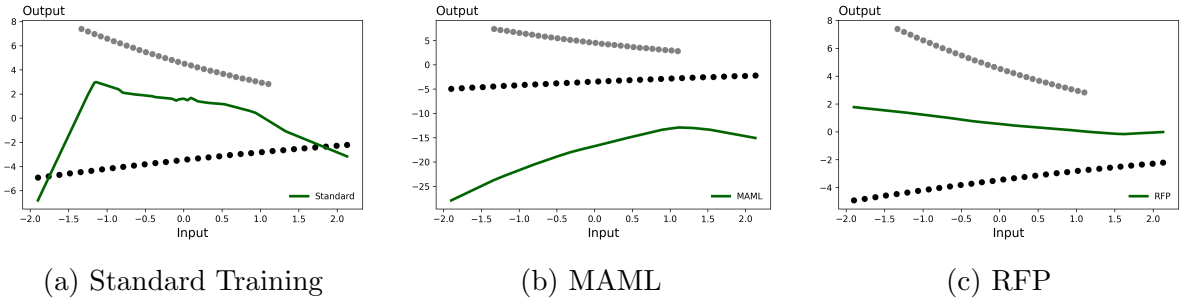

(a) Standard Training	(b) MAML	(c) RFP

Figure 6: 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
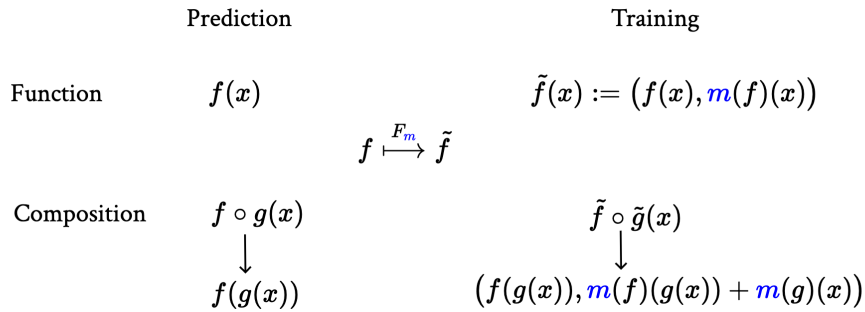

Figure 7: High Level Summary

4 Toy Model

4.1 Overview

In order to clarify our estimand of interest, we write down the following model. The key detail is that we parameterize the constraint function with the Right to Counsel Policy. Everything else is rather standard in that we allow the parameters of the constraint and utility functions to differ across individuals, and define our estimand as the expected value take 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.

4.2 Model

We define an element of the choice set as a bundle of housing related features. Most relevant to our data set is the housing search length which we include as the first component. It's important to acknowledge that individuals may be substituting across a wide range of housing aspects including the quality of the house and the price among others features.[29]

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

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 landlords actions/policies towards individual tenants.

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

The uttility function then captures some (present discounted) value of each housing option.

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

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.

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

---

[29]To introduce uncertainty into the model, we could augment each housing bundle with a stochastic process that captures the probability of remaining in the house across each time period. Note, this stochastic process would likely vary across individuals.

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.

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

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

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

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. As we illustrate in figure 8, the percentage point effect of the Right to Counsel (the y-axis) is relatively consistent 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 8b, we see a relatively consistent negative effect on the probability of moving into housing by a certain date.
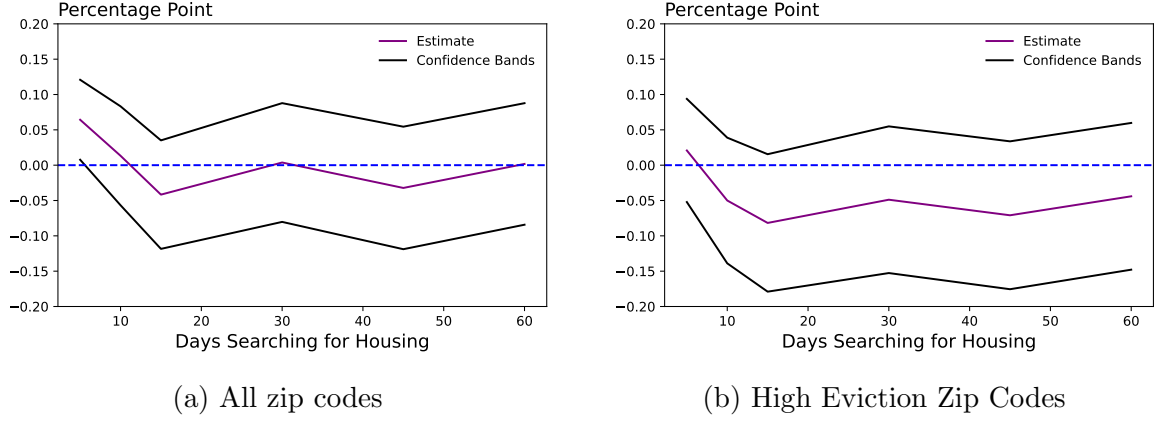
(a) All zip codes
(b) High Eviction Zip Codes

Figure 8: Reproduced Here: Confidence Bands formed via stratified bootstrapped sampling without replacement (75%)

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 9b.

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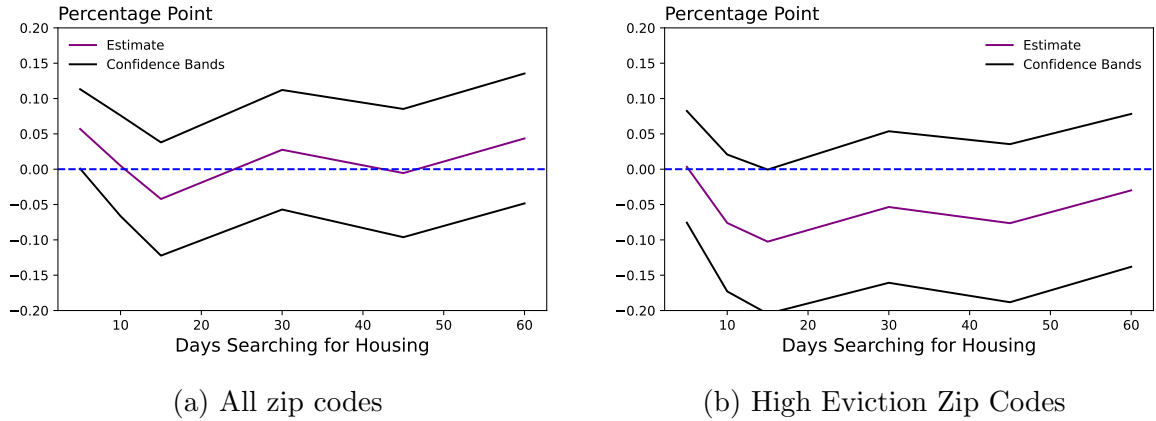$$+ \beta_3 X_i + \beta_4 Z_i + \varepsilon_i$$

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

Figure 9: 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 10 can be problematic in certain contexts.

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

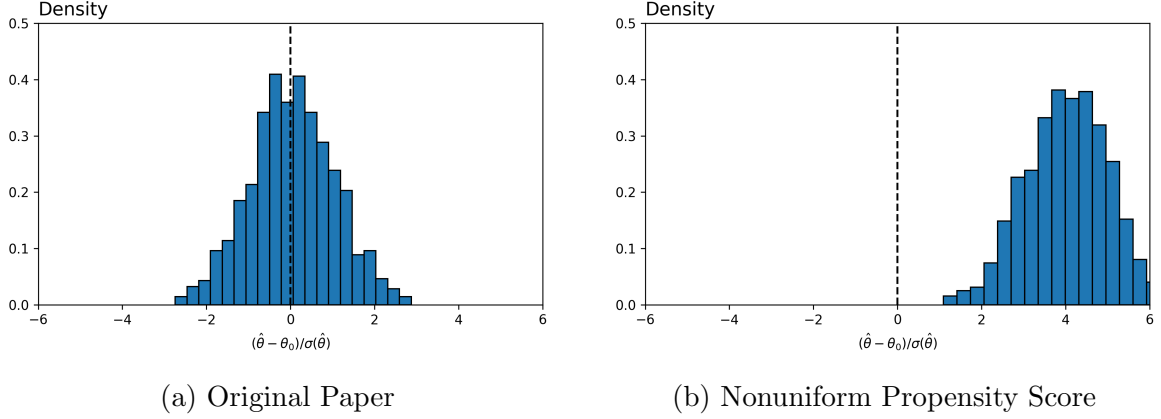(a) Original Paper   (b) Nonuniform Propensity Score

Figure 10: Reproduced Here: The sampling distribution of a Double Machine Learning Based Neural Network Estimator under selection on observables. In (a), we replicate the results of the original paper where the authors assume a constant treatment effect. Under a non-constant treatment effect as captured in (b), a failure to correct for the propensity score can introduce bias.

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 11. 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

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

Repeated Cross-Sectional Setting

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

$$\hat{\theta} := \int \hat{\theta}(X)d\mathbb{P}_X, \qquad \text{Average Treatment Effect on the Treated}$$
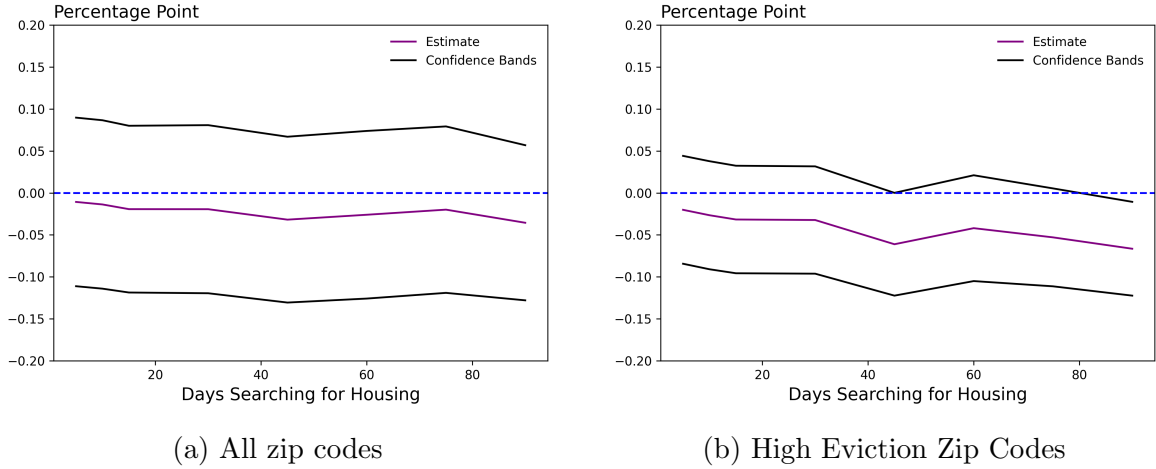

(a) All zip codes

(b) High Eviction Zip Codes

Figure 11: 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.[30]

---

[30]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

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

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

The recent events of Covid-19 and rising inflation have magnified the importance and fragility of housing for low-income individuals. 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.[31] 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,[32] 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 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.

---

costliest cases matters significantly for the mean costs reported above."

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

[32]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.

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.

Russell Engler. Connecting self-representation to civil gideon: What existing data reveal about when counsel is most needed. *Fordham Urb. LJ*, 37:37, 2010.

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.

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.

Ashley Gromis, Ian Fellows, James R Hendrickson, Lavar Edmonds, Lillian Leung, Adam Porton, and Matthew Desmond. Estimating eviction prevalence across the united states. *Proceedings of the National Academy of Sciences*, 119(21):e2116169119, 2022.

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.

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.

Bartosz Milewski. *Category theory for programmers*. Bartosz Milewski, 2019.

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.

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.
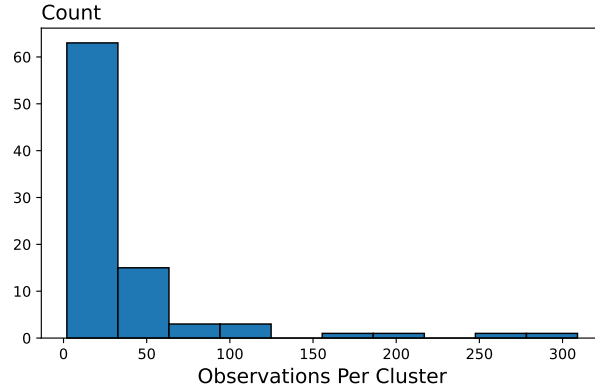
7 Appendix

7.1 Context Plots

Figure 12: Reproduced Here

7.2 High-Level Overview of Model

Test

$$l \circ f \circ c$$

Training

$$l >=> f >=> c$$

7.3 Category Theory Explanation

$$\text{class Functor } f \text{ where}$$
$$\text{fmap } :: \ (a \to b) \to (f\ a \to f\ b)$$

$$\text{instance Functor RFP where}$$
$$\text{fmap } f\ (x, r) = (f\ x, g(f)\ x + r)$$

Type Class: It's not enough to simply restrict the number of iterations of the inner loop. You need to fmap! That is, fmap is regularizing the forward pass: a statistic you compute during the forward pass. To be exact, we're defining a specific functor which is captured in haskell via the implementation of fmap.

If we make the simplification that we have two types, Data and Params, then our neural networks models can be thought of as functions between these sets: the clusterMap maps from Params to Params and the featureMap from Params to Data. With this, we have a category.

As explained in section 3, during training, we would like our neural network models to return a regularization value in addition to the predictions. We can do so with the help of a functor[33] which maps our category into a new category. For the types, it augments them the the regularization value. And for our neural network models, it embelleshes them so that they return both desired outputs. Just from this, it is then evident that when we are training the model, we are working in one category while during inference we're working in another.

Type Constructor

$$\text{data Reg a} = \text{Float \& } a$$

fmap

$$\text{fmap} :: (a \rightarrow b) \rightarrow (\text{Reg a} \rightarrow \text{Reg a})$$

This structure highlights that the key design choice is the fmap (how do you explicitly penalize the neural network).
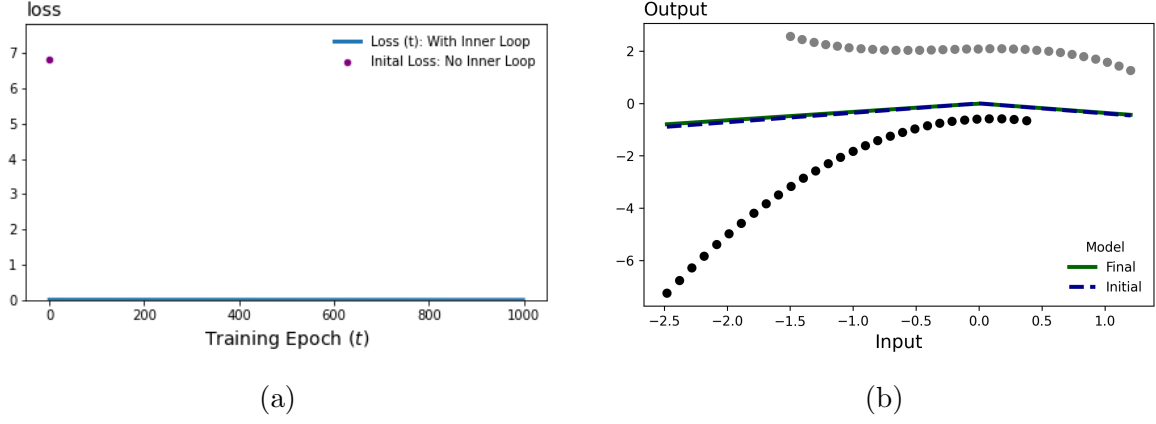
Figure 13: Without regularizing the implicit cluster maps, we risk losing the signal of the gradient – Reproduced Here

7.4 Framework Motivation

7.5 Loss Function

$$
\begin{aligned}
\hat{\theta} &= \operatorname*{argmin}_{\theta} \ \mathcal{L}(\theta) \\
&= \operatorname*{argmin}_{\theta} \ \sum_{c} \mathcal{L}_c(\theta) \\
&= \operatorname*{argmin}_{\theta} \ \sum_{c} F(\theta, \theta_c^*(\theta)) \\
&= \operatorname*{argmin}_{\theta} \ \sum_{c} F(\theta, \operatorname*{argmin}_{\theta_c} F(\theta, \theta_c))
\end{aligned}
$$

7.6 Haskell-like Signatures

$$
\text{regMAML} :: \text{Data} \rightarrow \text{Params} \rightarrow \big(\text{Params}, \text{Float}\big)
$$
$$
\text{regMAML data } \theta := \big(\text{Update}_m \circ \text{Update}_{m-1} \cdots \circ \text{Update}_1\big) \theta, \quad \mathcal{L}_c(\text{data}, \theta)
$$
$$
\text{where} \quad \text{Update}_t \ \theta = \theta - \alpha_t \nabla \mathcal{L}_c(\text{data}, \theta)
$$

---

[33]As illustrated via the math, as programmers our functors can be thought of as a type constructor - Milewski [2019]

$$\text{regNeuralODE} :: \text{Data} \rightarrow \text{Params} \rightarrow \big(\text{Data}, \text{Float}\big)$$

$$\text{regNeuralODE} \_ x \_ \theta := x + \int f(t, x(t), \theta)dt, \quad \int \left\| \frac{\partial^k}{dt^k} f(t, x(t), \theta) \right\| dt$$

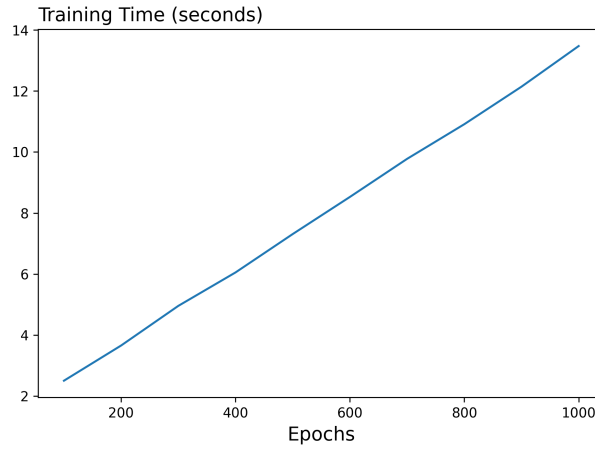$$\text{where} \quad x(0) = x$$

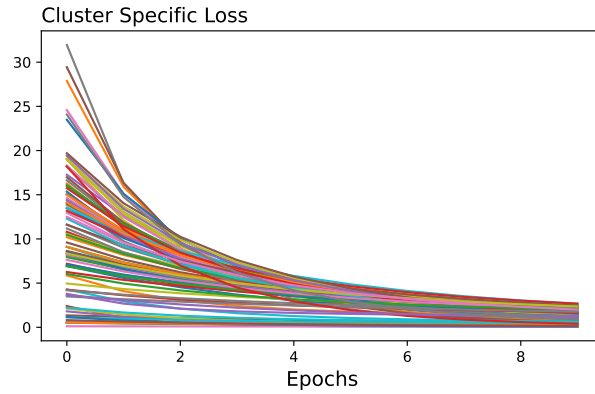## 7.7 Method Implementation

Figure 14: Reproduced Here

Figure 15: Reproduced Here

Using individual level data from the U.S. Department of Housing and Urban Development, and fitting a variety of estimators to the data, from difference-in-means to machine learning based estimators, our preliminary estimates suggest while the policy has no noticeable effects on for the average individuals, it may make it more difficult for African American women to find permanent housing. The estimate, while crude and preliminary, maintains a positive sign above 5 percentage points both across regression specifications as well as

across subsets of zip codes. Eviction number from Gromis et al. [2022]. (Individual Costs): Collinson et al. [2022] writes, "We find that eviction causes significant disruptions that are reflected in increases in residential mobility, homelessness, and hospital use". (Court Costs): As Seron et al. [2001] notes, legal representation actually might decrease housing court costs as the number of appearances and post judgement motions decline. (General Public): As Desmond [2019] writes, "Residential instability often brings about other forms of instability—in families, schools, communities— compromising the life chances of adults and children"

(Multitude of Causes): David Ehrens writes in his letter to the editor of Dartmouth Week of "evictions related to the pandemic, chronic housing supply shortages, inequities in lending, generational poverty, and other harms". (Eviction Proceedings): Missing Reference writes "the vast majority resolved by default or settlement, typically the result of hallway negotiation" (Growing Interest): Engler [2010] writes "a renewed call for a civil right to counsel, or civil Gideon, has gained momentum . . . as well as a surge in membership in the newly-created National Coalition for a Civil Right to Counsel."