Empirical Bayes with Side Information: Incorporating Geographic Relevance into Neighborhood Impact Estimation

Mohammad H. Seyedsalehi*

Octobor 2026

Abstract

This paper introduces a novel methodology for estimating neighborhood impacts on intergenerational mobility. The approach employs a nonparametric empirical Bayes framework that addresses geographic heterogeneity in the underlying data-generating process. First, I propose a theoretical framework to address the long-standing yet unresolved question of how to make better use of highly relevant samples—in this context, data from geographically closer regions—when designing a shrinkage estimator for a target area. I demonstrate, both theoretically and empirically, how neglecting this factor can lead to systematically distorted estimates. Second, using publicly available data, I demonstrate that the relevance-based empirical Bayes method reduces Mean Squared Error (MSE) by 26% to 34% relative to benchmark estimates.

1 Introduction

The question of how neighborhoods influence children's future income opportunities has become a key focus in applied research (Chetty and Hendren 2018a, 2018b; Deutscher et al., 2020; Kenedi and Sirugue, 2023; Mogstad et al., 2024). Using geographically detailed datasets, researchers have developed methods, often based on fixed effects, to estimate the impact of regions on intergenerational mobility and economic outcomes. However, when applied to smaller areas, these estimators often produce noisier results, complicating reliable inference.

*m.seyedsalehi@mail.utoronto.ca
Economics Department, University of Toronto, Toronto, Ontario, Canada.

The challenge of reducing measurement error in neighborhood effect estimators has led to the widespread adoption of empirical Bayes methods. These methods improve unreliable estimates by borrowing strength from the full sample, using assumptions about the data-generating process for the unobservable parameter—commonly modeled as a prior distribution. Introduced by James and Stein (1961), the parametric empirical Bayes (PEB) estimator has traditionally been widely applied in research for its simplicity and optimality under normal prior assumptions. More recently, nonparametric maximum likelihood estimation (NPMLE) has gained traction in empirical Bayes studies, as advancements in algorithms have enabled faster and more precise computations (Koenker and Mizera, 2014).

A key overlooked question in the empirical Bayes literature is how to select relevant data samples for "borrowing strength." The common practice of relying on the entire dataset implicitly assumes that all data points are equally relevant, which may not hold true. If the priors generating parameters differ across neighborhoods, incorporating irrelevant data can lead to less reliable estimates.

Our core contribution is the development of a theoretical framework for estimating neighborhood impacts that incorporates geographic relevance. I illustrate its importance using an example from Chetty and Hendren (2018b). Building on a conventional framework where fixed effect estimators of neighborhood impacts reflect the sum of true underlying effects and independent noise, I demonstrate that failing to account for the possibility that geographically smaller areas may have different priors can result in empirical Bayes estimates that are systematically misaligned with local patterns. In Section 3, I formalize how empirical Bayes methods can be adapted to prioritize relevant data and improve the estimation of neighborhood impacts.

I apply this method to estimate intergenerational income mobility across various commuting zones in the United States, using publicly available data from Chetty and Hendren (2018b). While empirical Bayes frameworks usually cannot directly evaluate performance due to the unobservability of true parameters, I use alternative estimates provided by Chetty and Hendren (2018b) as a benchmark. These benchmark estimates were constructed by refining raw causal effect estimates through shrinkage toward average outcomes based on data from permanent residents. Although not perfect, these adjusted estimates are supported by Chetty and Hendren (2018b) as being substantially more reliable than the raw estimates. I use them as an empirical reference to assess the performance of the relevance-based shrinkage method, which achieves a 26% to 34% reduction in Mean Squared Error compared to the original estimates.

The importance of incorporating relevance in empirical Bayes methods has been extensively emphasized by Bradley Efron over the past decades (Efron, 2008; Efron, 2011; Efron and Hastie, 2018; Efron, 2019a; Efron, 2019b; Efron, 2021), with Efron (2011) providing the main concrete attempt to formalize relevance concerns in empirical Bayes shrinkage. However, the application in empirical research has remained limited, perhaps reflecting a gap between the existing theoretical framework and the practical considerations or disciplinary contexts that typically arise in applied settings. In Section 2.2, I detail Efron's approach and demonstrate how the new theoretical framework relaxes its assumptions and structure, offering a more flexible and practical solution.

The methodology proposed in this paper for addressing relevance is closely related to the growing body of literature that leverages covariate information to enhance empirical Bayes estimation. In the parametric empirical Bayes literature, models such as those in Fay and Herriot (1979), Green and Strawderman (1991), Kou and Yang (2017), and more recently Ignatiadis and Wager (2022), specify structures in which covariates are used to predict the mean of the prior distribution. Similarly, Cohen, Greenshtein, and Ritov (2013) pursue generalizations in the nonparametric empirical Bayes setting by considering covariates as a parallel proxy for the unobservable parameter. These methods typically constrain the influence of covariates to a single channel, usually through the prior mean. This contrasts with the broader intuition underlying the Tweedie's formula (Equation 3), which suggests that covariates can shape the prior distribution in a more complex and non-linear way. Luo et al. (2023) is more closely aligned in spirit, proposing what they term the integrative Tweedie approach. However, their formulation takes a different route. In contrast, my method directly builds on Efron's (2011) perspective by combining Bayes' rule with the Tweedie's formula, as formalized in Equation 8. A similar limitation applies to studies in the literature on spatial or cross-unit correlation that reduce the problem to modeling linear dependence among unobservable parameters based on spatial locations and rely on stronger structural assumptions. (e.g., Wang et al., 2025; Yasui et al., 2000).

Overall, the study demonstrates how empirical Bayes shrinkage can be applied to obtain accurate estimates in geographically detailed contexts, avoiding the need for normality assumptions while remaining computationally feasible. Although the focus is on intergenerational mobility, this method is adaptable to a wide range of spatial applications.

The rest of the article is structured as follows: Section 2 introduces the empirical Bayes shrinkage technique, with a detailed discussion of the relevance concern in geographical contexts in Section 2.1. Sections 3 presents the framework for addressing relevance issues

in empirical Bayes and describe the proposed estimator. Section 4 outlines the data and discusses the estimation results.

2 Empirical Bayes Shrinkage and the Problem of Relevance

Empirical Bayes shrinkage techniques have recently become popular for improving parameter estimates by borrowing strength from other observations. Consider the case in which the unobservable parameters μ_i are independently sampled from an unknown prior distribution $g(\cdot)$, and the observed variables Y_i are conditionally distributed from Gaussian distributions with mean μ_i and known variances of σ_i^2. The goal is to estimate each individual μ_i. The posterior mean of μ_i for a known $g(\cdot)$, is an estimator that minimizes the quadratic Bayes risk and can be expressed as follows (Gilraine et al., 2020):

$$\hat{\mu}_i = E(\mu_i \mid Y_i) = \frac{\int \mu \phi(Y_i - \mu; \sigma_i^2) g(\mu)\, d\mu}{\int \phi(Y_i - \mu; \sigma_i^2) g(\mu)\, d\mu}. \tag{1}$$

Here, $\phi(\cdot; \sigma_i^2)$ denotes the Gaussian density function. While the above estimator is infeasible because $g(\mu)$ is unknown, a feasible estimator can be obtained by substituting $g(\mu)$ with its nonparametric maximum likelihood estimator (NPMLE). Formally, the NPMLE is defined as:

$$\hat{g} \equiv \underset{g \in \mathcal{G}}{\operatorname{argmax}} \left\{ \sum_{i=1}^{N} \log \int g(\mu) \phi(Y_i - \mu; \sigma_i^2)\, d\mu \right\}, \tag{2}$$

where $\mathcal{G}$ is the space of all probability distributions over $\mathbb{R}$. Solving the above problem—originally formulated by Kiefer and Wolfowitz (1956)—has only recently become practical through alternative methods, such as Koenker and Mizera (2014). Their method employs a finite-dimensional convex approximation to address the issue of slow convergence effectively.

In the special case of homogeneous variances ($\sigma_i = \sigma$), the posterior mean of μ_i can be expressed as the sum of Y_i and a function of the marginal distribution of observables, f, commonly referred to in the literature as the Tweedie's formula (Efron, 2011):

$$\hat{\mu}_i = E(\mu_i \mid Y_i) = Y_i + \sigma^2 \frac{d}{dy} \log f(y) \Big|_{y=Y_i}, \tag{3}$$

where the marginal density $f(y)$ is the convolution of the prior $g(\mu)$ and the Gaussian

density function ϕ, given by:

$$f(y) = \int g(\mu)\phi(y-\mu;\sigma^2)\,d\mu. \tag{4}$$

The term $\sigma^2 \frac{d}{dy} \log f(y)|_{y=Y_i}$ represents the shrinkage applied to the observable Y_i, incorporating information from other observations through the derivative of the marginal distribution of Y_i. To obtain a feasible estimator, this term can be estimated using Poisson regression, as described in Efron (2011) and Efron (2008). This approach is commonly referred to in the literature as f-modeling, in contrast to methods that estimate the prior g, known as g-modeling (Efron, 2011; Efron, 2016).

2.1 An Example of the Relevance Problem

A key overlooked question in the empirical Bayes literature is how to select relevant data samples from which to borrow strength for various applications, such as shrinkage estimation, inference, or hypothesis testing. The common practice of using all available Y_i as a default solution can be problematic. When auxiliary information suggests that certain parameters are less relevant than others, incorporating such irrelevant data samples into empirical Bayes techniques may result in suboptimal shrinkage outcomes. An example from Chetty and Hendren's (2018b) estimation of U.S. commuting zones' effects on intergenerational mobility can illustrate this issue.

Their study estimates the causal impacts of spending an additional year in U.S. geographic units on children's adult incomes using national tax records. They present various estimates under different conditions, segmented by gender, age groups, parental income levels, and geographic divisions—such as commuting zones and counties. Here, I focus on applying an empirical Bayes shrinkage method to the estimates of the causal impact of commuting zones, derived solely from movers whose parents were at the 25th percentile of the income distribution. I refer to these data as the **original estimates** of CZ effects, and they are presented in Figure 1a.

Since the original estimates are highly noisy, it is natural to consider a shrinkage estimator based on the assumption that the CZ impacts, μ_i, within a given state $X_i = x_0$, are drawn from a common prior distribution g_{x_0}, given that commuting zones within the same state are more likely to exhibit similar socioeconomic characteristics. However, the small number of commuting zones within each state makes it impractical to estimate g_{x_0} and therefore apply Equation 1 to derive the posterior mean estimator.

As an alternative approach, one could assume a common prior distribution $g(\cdot)$ for all CZ impacts across the U.S. to derive posterior mean estimators via Equation 1. I refer to these estimates as **pooled estimates** for CZ effects and denote them as $\hat{\mu}_i^{\text{Pooled}}$. As shown in Figures 1a and 1b, while shrinkage substantially reduces the range of the estimated effects compared to the original estimates, both figures reveal clear geographic variation in the range of observed effects across different regions of the U.S. This underscores the concern that pooled shrinkage may incorporate information from distant commuting zones that are not relevant to a specific state's context, as they may provide little meaningful information about the parameters of that state's commuting zones. As a result, such an assumption could lead to the construction of a less informative prior and, consequently, less accurate estimators that suffer from over- or under-shrinkage.

To clarify further, Figure 2 shows the histogram of the pooled estimates, with the portion corresponding to observations from one of the seven adjacent southeastern states (South Carolina, Georgia, North Carolina, Florida, Mississippi, Alabama, and Tennessee) highlighted in blue. The graph indicates a noticeable potential difference between the posterior distributions of estimates from these southeastern states and those from other regions, as they are concentrated mainly on the left side of the larger histogram. This difference could result from one or both of the following factors: (1) differences in the prior distributions from which μ_i are drawn in the southeastern states compared to other areas, or (2) differing levels of observational noise in these regions, which influence the posterior distribution. Here, I focus on addressing the first possibility.

To handle this concern, I propose using a framework called statistical relevance (Efron, 2011), which I reinterpret and simplify by introducing the concept of a relevance-based prior to preserve its powerful capabilities while ensuring it is suitable for more general contexts. Instead of using a common prior $g(\cdot)$ for all unobservable parameters, I approximate $g_{x_0}(\cdot)$ by defining the relevance-based prior $g^r_{x_0}(\cdot)$ as a mixture of prior distributions from states surrounding x_0, weighted by their geographic relevance. This approximation is then used in Equation 1. The relevance weight $\lambda_{x_0}(x_k)$ assigns higher values to states x_k that are geographically closer to x_0, ensuring that $g^r_{x_0}(\cdot)$ borrows more strength from relevant states. I discuss Efron's (2011) method in detail in Section 2.2 and introduced my own approach in Section 3.

2.2 Relevance Framework in the Literature

In the case of homogeneous variances ($\sigma_i = \sigma$), Efron (2011) proposes a relevance framework for applying empirical Bayes shrinkage that explicitly models the structure of priors around a reference location x_0. Assume I observe $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, where the unobservable parameter μ_i is drawn from an unknown prior $g_{x_0}(\cdot)$, representing the data-generating prior at location $X_i = x_0$. The direct estimation of $g_{x_0}(\cdot)$ is not possible due to insufficient data, as only a small number of other μ_k values are drawn from $g_{x_0}(\cdot)$. Its approach assumes that the prior at any other location x is either equal to $g_{x_0}(\cdot)$ with probability $\rho(x_0, x)$ or an irrelevant distribution $g_x^{\text{irr}}(\cdot)$ otherwise, where I^r is a binary indicator of relevance.

$$g_x(\cdot) = I^r \cdot g_{x_0}(\cdot) + (1 - I^r) \cdot g_x^{\text{irr}}(\cdot). \tag{5}$$

The function $\rho(x_0, x)$ quantifies relevance via a kernel that assigns higher probability of sharing the prior to nearby locations. Imposing this relevance structure on the data-generating process yields a lemma that expresses the marginal density $f_{x_0}(y)$, which is the density of the random variable Y_i at location $X_i = x_0$, evaluated at a realized value y. This is written in terms of the overall marginal density $f(y)$, scaled by the expected relevance weight at that value (see Efron, 2011, for proof).

$$f_{x_0}(y) = f(y)\frac{\mathbb{E}[\rho(x_0, x) \mid y]}{\Pr(I^r = 1)}. \tag{6}$$

Recall that $f(y)$ is defined as in Equation 4, and $f_{x_0}(y)$ represents the convolution of the local prior $g_{x_0}(\cdot)$ with the Gaussian density $\phi(\cdot; \sigma)$.

Incorporating Equation 6 into Tweedie's formula adds a corrective term, derived from the first derivative of $\log \mathbb{E}[\rho(x_0, x) \mid y]$ with respect to y, to adjust the posterior mean.

$$E(\mu_i \mid Y_i, X_i = x_0) = Y_i + \sigma^2 \left(\frac{d}{dy} \log f(y)\Big|_{y=Y_i} + \frac{d}{dy} \log \mathbb{E}[\rho(x_0, x) \mid y]\Big|_{y=Y_i} \right). \tag{7}$$

Although the proposed extension seems promising, its application in the literature, particularly in the context of spatial data, remains absent to the authors' knowledge. This gap could be attributed to structural assumptions, such as modeling priors with a probabilistic structure (Equation 5) and centering them around a single point x_0. These assumptions make the method difficult to apply across datasets, since they cannot hold simultaneously

for all locations. Building on Efron's idea, this article introduces an alternative yet compatible relevance framework. Our design achieves the same relevance-based correction in Tweedie's formula, and is flexible to be applied across different locations without requiring mutually inconsistent assumptions.

3 Relevance-Based Empirical Bayes Shrinkage

I begin by modeling a setting similar to that introduced in Section 2, now incorporating an observable covariate X_i, which represents the location associated with each observation and has marginal density denoted by $\omega(\cdot)$ defined over the support χ. In this framework, the unobservable parameter μ_i is assumed to be drawn from the prior distribution $g(\cdot \mid X_i = x_0)$ which becomes equal to the location-specific prior $g_{x_0}(\cdot)$ upon conditioning, reflecting how the distribution of parameters may vary across different covariate values. Each observed outcome Y_i is then generated independently from a Gaussian distribution with mean μ_i and known variance σ^2. Combining Tweedie's formula with Bayes' rule, the posterior expectation of μ_i given (Y_i, X_i) is presented in Theorem 1.

Theorem 1 (Empirical Bayes Posterior Mean with Side Information)**.** *Let $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$, where $\mu_i \sim g(\cdot \mid X_i)$, and $X_i = x_0$ is observed. Then, the posterior mean of μ_i given Y_i and X_i is:*

$$E(\mu_i \mid Y_i, X_i = x_0) = Y_i + \sigma^2 \left(\frac{d}{dy} \log f(y) \Big|_{y=Y_i} + \frac{d}{dy} \log \omega(x_0 \mid y) \Big|_{y=Y_i} \right), \tag{8}$$

where $f(y)$ is the unconditional marginal density of Y_i, and $\omega(x_0 \mid y)$ is the density of $X_i = x_0$ given $Y_i = y$.

The additional term $\frac{d}{dy} \log \omega(x_0 \mid y)|_{y=Y_i}$ serves as a corrective adjustment to the standard Tweedie's formula, accounting for the presence of covariate information. Notably, Theorem 1 does not rely on any structural assumptions like those in Equation 5 regarding the data-generating process. In what follows, I introduce a method for estimating this correction term by developing the notion of relevance-based priors.

A relevance-based prior is defined as a convex combination of location-specific prior distributions, with weights typically structured to decay with distance from a reference location such as x_0. This distribution, denoted $g^r_{x_0}(\cdot)$, is constructed as follows:

Definition 1 (Relevance-Based Prior and Marginal Distribution)**.** *Let $\lambda_{x_0}(x)$ be a pre-specified relevance function on $x \in \chi$, relative to a fixed reference point x_0. The function λ_{x_0} is assumed to satisfy the following convexity conditions:*

$$\lambda_{x_0}(x) \geq 0 \quad \textit{for all } x \in \chi, \quad \textit{and} \quad \sum_{x \in \chi} \lambda_{x_0}(x) = 1.$$

The weights $\lambda_{x_0}(x)$ are typically structured to decay with distance from the reference location x_0. Given these weights, I define:

1. *The relevance-based prior:*

$$g^r_{x_0}(\mu) = \sum_{x \in \chi} \lambda_{x_0}(x) g_x(\mu). \tag{9}$$

2. *The relevance-based marginal distribution:*

$$f^r_{x_0}(y) = \sum_{x \in \chi} \lambda_{x_0}(x) f_x(y). \tag{10}$$

For example, in the context of Chetty and Hendren (2018b), x_0 may represent a particular state, and a possible formulation for λ_{x_0}, where d represents the spatial distance between states, could be:

$$\lambda_{x_0}(x) = \frac{d_{\max} - d(x, x_0)}{\sum_{x' \in \chi} (d_{\max} - d(x', x_0))}, \tag{11}$$

where $d_{\max}$ is the maximum spatial distance between any two states in the dataset.

The constructed distribution $g^r_{x_0}(\cdot)$ serves as an alternative model, although it may be a misspecified one, for constructing the empirical Bayes estimator of μ_0. It implicitly approximates the true conditional prior $g(\cdot \mid X_i = x_0)$ used in that estimation. In contrast to the approach that assumes a universal prior $g(\cdot)$ across all units, this specification introduces flexibility by partially accounting for spatial heterogeneity, thereby providing a closer approximation to $g(\cdot \mid X_i = x_0)$. This is motivated by the relevance assumption, which posits that the priors at x and x_0 are more alike when x is geographically close to x_0.

Consequently, it helps construct an empirical Bayes estimator that includes a term approximating the corrective term in Equation 8, rather than omitting it entirely as in

the conventional pooled shrinkage approach. The statistical consequences of this model approximation are analytically tractable, and the resulting estimation error can be formally bounded, as shown later in Theorem 2.

By applying Bayes' rule to Equation 10, the relevance-based marginal distribution can be expressed as a multiple of $f(y)$, a relationship I refer to as the Relevance Lemma. This serves as an alternative to Equation 6 proposed by Efron (2011).

Lemma 1 (Relevance Lemma)**.** *The relevance-based marginal distribution $f^r_{x_0}(y)$ can be expressed as:*

$$f^r_{x_0}(y) = \sum_{x \in \chi} \frac{\lambda_{x_0}(x)\omega(x \mid y)}{\omega(x)} f(y) = E\left(\frac{\lambda_{x_0}(x)}{\omega(x)} \mid y\right) f(y). \tag{12}$$

The Relevance Lemma offers a useful interpretation of the relationship between the relevance-based marginal distribution $f^r_{x_0}(\cdot)$ and the true location-specific distribution $f_{x_0}(\cdot)$. Specifically, $f_{x_0}(\cdot)$ can be expressed as

$$f_{x_0}(y) = \frac{\omega(x_0 \mid y)}{\omega(x_0)} f(y) = E\left(\frac{I_{x_0}(x)}{\omega(x)} \mid y\right) f(y). \tag{13}$$

Here, $I_{x_0}(x)$ denotes the indicator function for $\{x = x_0\}$. Comparing Equation 13 to the Relevance Lemma reveals that the relevance-based marginal distribution approximates $f_{x_0}(\cdot)$ by replacing the sharp indicator $I_{x_0}(x)$ with a smoother relevance function $\lambda_{x_0}(x)$, thereby allowing information from nearby locations to contribute in a graded fashion. This approximation also clarifies the connection between Equation 8 and Equation 14, as discussed further below.

Finally, to construct a posterior mean estimator for μ_i, under the working approximation that it is drawn from a relevance-based prior $g^r_{x_0}$, I apply the Relevance Lemma to Tweedie's formula. This yields an enhanced empirical Bayes shrinkage estimator that incorporates covariate information, expressed as follows:

Definition 2 (Relevance-based Empirical Bayes Estimator)**.** *The relevance-based empirical*

Bayes estimator for μ_i is given by:

$$\begin{aligned}\hat{\mu}_i^r = \hat{E}(\mu_i \mid Y_i; \mu_i \sim g_{x_0}^r) &= Y_i + \sigma^2 \frac{d}{dy} \widehat{\log f_{x_0}^r(y)}\Big|_{y=Y_i} \\ &= Y_i + \sigma^2 \left(\frac{d}{dy} \widehat{\log f(y)}\Big|_{y=Y_i} + \frac{d}{dy} \log E\left(\widehat{\frac{\lambda_{x_0}(x)}{\omega(x)}} \mid y \right) \Big|_{y=Y_i} \right).\end{aligned} \tag{14}$$

The term $\frac{d}{dy}\log f(y)|_{y=Y_i}$ can be estimated using the NPMLE approach proposed by Koenker and Mizera (2014), and $\frac{d}{dy}\log E\left(\frac{\lambda_{x_0}(x)}{\omega(x)} \mid y\right)|_{y=Y_i}$ can be estimated through smooth regression techniques, employing a similar approach as Efron (2011). One computational advantage of Equation 14 is its efficiency when applied simultaneously across multiple observations Y_j. The prior $g(\cdot)$ only needs to be estimated once and then used to construct the term $\frac{d}{dy}\widehat{\log f(y)}\big|_{y=Y_j}$ across all observations, which significantly reduces computational burden.

A central concern is how the choice of the relevance function influences the estimator's bias, arising from the discrepancy between $g_{x_0}^r(\cdot)$ and the true prior $g(\cdot \mid X_i = x_0)$. A narrower relevance function, which places greater weight on observations near the target point x_0, tends to reduce bias by more closely capturing the local structure of the prior. Conversely, a broader function pools information from a wider region, potentially introducing bias by incorporating less relevant observations.

Theorem 2 shows the order of this bias when the relevance function is viewed as a kernel-based choice with bandwidth h. While a sharper relevance function reduces the model specification bias, it potentially increases the estimation error of $\frac{d}{dy}\log E\left(\widehat{\frac{\lambda_{x_0}(x)}{\omega(x)}} \,\middle|\, y\right)|_{y=Y_i}$. When an estimation method such as a smooth regression is fixed, its variance often depends on the sharpness (e.g. derivative properties) of λ_{x_0} at the point of observation. The optimal MSE, and hence the optimal bandwidth, can be derived based on the chosen estimation procedure, which could be explored in future research.

Theorem 2 (Model Bias for the Relevance-Based Estimator)**.** *Suppose the conditional density $\omega(x \mid y)$ and the marginal density $\omega(x)$ are twice continuously differentiable in x in a neighborhood of x_0. Let $K : \mathbb{R}^d \to \mathbb{R}_+$ be a symmetric, bounded, and compactly supported kernel satisfying $\int K(u)\,du = 1$. For any bandwidth $h > 0$, define*

$$\lambda_{x_0,h}(x) \;=\; h^{-d}\, K\!\left(\frac{\|x - x_0\|}{h}\right),$$

and let $g^r_{x_0,h}(\cdot)$ *denote the relevance-based prior constructed with* $\lambda_{x_0,h}$.

Then, as $h \to 0$*, the bias induced by relevance-based model misspecification satisfies*

$$E\big(\mu_i \mid Y_i;\, \mu_i \sim g^r_{x_0,h}\big) \;-\; E\big(\mu_i \mid Y_i;\, \mu_i \sim g_{x_0}\big) \;=\; O(h^2).$$

Overall, the approach bypasses the structural assumptions underlying Equation 5 and their limitations, which cannot be satisfied simultaneously across multiple locations. In contrast, various relevance-based priors can be defined independently as pointwise estimators at different locations x_j, allowing us to construct relevance-based estimators for all observations Y_k simultaneously. This is a direct consequence of treating the relevance-based estimator as a tool to approximate the combined effect of Bayes' rule and Tweedie's formula—formalized in Equation 8.

4 Empirical Analysis

4.1 Data

In this section I apply my method to the estimation of intergenerational income mobility across different locations in the United States, using publicly available data from Chetty and Hendren (2018a, 2018b).

Chetty and Hendren (2018a) shows how children's expected incomes differ across geographical units such as commuting zones (CZ) and counties, conditional on their parents' incomes. Their primary measure of interest is $\bar{Y}_{cp}$, indicating the expected income rank of children from parents at the pth percentile of the national income distribution, based on the area c in which they were raised. Although the estimates $\bar{Y}_{cp}$ have very low sampling variances, they are potentially biased due to residential sorting. I suppress the index p henceforth for notational simplicity.

The second dataset, provided by Chetty and Hendren (2018b), focuses on the estimation of causal effect of regions on income mobility using a sample of cross-county movers. They study Y^*_c, which measures the effect of spending one more year of childhood in a given region. Y^*_c is mean-centered to reflect each region's influence relative to the national average. In Section 2 I referred to Y^*_c as original estimates of CZ effects.

To improve the accuracy of noisy causal effect estimates derived from movers, Chetty

and Hendren (2018b) implements a shrinkage procedure that minimizes mean squared error (MSE). Specifically, they construct a forecast $\hat{\mu}_c^f$ by optimally combining the raw estimate Y_c^* with the average outcome of permanent residents in the same region, $\bar{Y}_c$. The resulting forecast $\hat{\mu}_c^f$ is designed to be the linear MSE-minimizing predictor of the true CZ-level causal effect μ_c, assigning more weight to the mover-based estimate when it is more precisely estimated (see Chetty and Hendren 2018a; 2018b for further details).

Building on the Chetty and Hendren (2018b) framework, I temporarily ignore the availability of estimates from permanent residents. I apply the relevance method to the raw causal effect estimators of CZs, Y_c^*, within the context of commuting zones at the 25th percentile of parental income. The heatmap of the raw data, before applying shrinkage, is presented in Figure 1a, and its histogram is shown in Figures 2 and 3a. I then use their shrunken forecast toward permanent residents, $\hat{\mu}_c^f$, as a benchmark, to measure the performance of the method. The histogram and heatmap of $\hat{\mu}_c^f$ are shown in Figure 3b and Figure 5a. An interesting observation from comparing the heatmaps is that, although spatial correlation was not explicitly used in estimation of $\bar{Y}_c$, the resulting estimates exhibit strong spatial coherence—nearby regions tend to have similar values. This pattern provides further support for the relevance-based approach.

While I acknowledge that using $\hat{\mu}_c^f$ as a benchmark is not without limitations—since it remains an estimate subject to noise—it nonetheless serves as a valuable empirical reference point for evaluating the performance of the method. Beyond its construction as an MSE-minimizing estimator—which leverages permanent-resident data with very low sampling variability—Chetty and Hendren (2018b) argue that $\hat{\mu}_c^f$ closely approximates the true causal effect. For instance, they show that Y_c^* is highly correlated with various observable area-level characteristics, including the fraction of Black residents, racial segregation as measured by the Theil index, income inequality (Gini index), the prevalence of single motherhood, social capital, and public school expenditures per student. They find that once permanent residents' outcomes $\bar{Y}_c$ are incorporated through shrinkage, these additional characteristics explain little residual variation in Y_c^*. As a result, the shrinkage of Y_c^* toward $\bar{Y}_c$ performs nearly as well as more complex models with richer covariate structures.

4.2 Empirical Strategy

In the presence of heteroskedastic noise, where $Y_c^* \sim N(\mu_c, \sigma_c^2)$, with μ_c denoting the latent causal effect and σ_c either known or consistently estimated (as reported in Chetty

and Hendren 2018b), applying Equation 14 to estimate the posterior mean becomes more nuanced. To account for this variance heterogeneity, Equation 8 can be adapted into the following modified version:

$$E(\mu_c \mid Y_c;\, \mu_c \sim g_{s(c)},\, \sigma_c) = Y_c + \sigma_c^2 \Big(\frac{d}{dy}\log f(y;\sigma_c)\Big|_{y=Y_c} + \frac{d}{dy}\log \omega(s(c) \mid y;\sigma_c)\Big|_{y=Y_c}\Big), \tag{15}$$

where $s(c)$ denotes the state in which commuting zone c is located.

Although variance heterogeneity is present, the term $\frac{d}{dy}\log f(y;\sigma_c)\big|_{y=Y_c}$ can still be estimated using g-modeling techniques, as described by Koenker and Mizera (2014). Approximating the second term, $\frac{d}{dy}\log \omega(s(c) \mid y;\sigma_c)\big|_{y=Y_c}$, in a manner similar to the homogeneous case requires estimating the conditional expectation $\frac{d}{dy}\log E\left(\frac{\lambda_{s(c)}(x)}{\omega(x|\sigma_c)} \mid Y_c;\sigma_c\right)$. For this approximation, the range of σ_c values is divided into eight interquantile intervals, and smooth regressions are performed using only observations whose associated σ_c values fall within the same interval.

4.3 Empirical Results

I begin by employing three distinct relevance functions to assess the robustness and compare the resulting estimates. The first is a linear relevance function, assigning weights that are inversely proportional to the geographic distance between states.

$$\lambda_{x_0}^{Linear}(x) = \frac{d_{\max} - d(x, x_0)}{\sum_{x' \in \chi}\left(d_{\max} - d(x', x_0)\right)}, \tag{16}$$

where $d_{\max}$ is the maximum spatial distance between any two states in the dataset, approximately 4,228 km. The states of Alaska and Hawaii are excluded from the analysis, as they are not geographically contiguous with the rest of the United States.

The second relevance function assigns weights that decay quadratically with distance. Specifically, for all $x \in \mathcal{X}$, it is defined as:

$$\lambda_{x_0}^{Quad}(x) = \frac{\left(d_{\max} - d(x, x_0)\right)^2}{\sum_{x' \in \chi}\left(d_{\max} - d(x', x_0)\right)^2}. \tag{17}$$

The third relevance function uses an indicator function to assign equal weights to states

closer than $0.3 \cdot d_{\max}$, and zero weight otherwise:

$$\lambda_{x_0}^{Binary}(x) = \begin{cases} \frac{1}{N} & \text{if } d(x, x_0) < 0.3 \cdot d_{\max}, \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

where N is the number of states satisfying $d(x, x_0) < 0.3 \cdot d_{\max}$.

I denote the three relevance-based empirical Bayes estimators, derived using these relevance functions as $\hat{\mu}_c^{r,Linear}$, $\hat{\mu}_c^{r,Quad}$, and $\hat{\mu}_c^{r,Binary}$, respectively. Their descriptive statistics are presented in Table 1, with Figure 3 displaying their respective histograms and Figure 5 illustrating their heatmaps.

Table 1: Descriptive Statistics for the Estimation Methods

Method	Variable	M	Mean	SD	Median	Min	Max
Relevance-Based Shrinkage (Linear)	$\hat{\mu}_c^{r,\text{Linear}}$	588	0.062	0.146	0.033	−0.208	1.045
Relevance-Based Shrinkage (Quadratic)	$\hat{\mu}_c^{r,\text{Quad}}$	588	0.054	0.227	0.023	−1.570	1.075
Relevance-Based Shrinkage (Binary)	$\hat{\mu}_c^{r,\text{Binary}}$	588	0.059	0.143	0.030	−0.179	1.297
Pooled Shrinkage	$\hat{\mu}_c^{\text{Pooled}}$	588	0.044	0.121	0.019	−0.164	0.764
Original	Y_c^*	588	0.184	0.615	0.055	−2.416	3.341
Benchmark	$\hat{\mu}_c^f$	588	0.188	0.541	0.152	−0.908	1.993

Table 2 presents the Mean Squared Error (MSE) relative to the benchmark ($\hat{\mu}_c^f$) for $\hat{\mu}_c^{r,Linear}$, $\hat{\mu}_c^{r,Quad}$, and $\hat{\mu}_c^{r,Binary}$. The results show substantial reductions in MSEs, ranging from 26% to 34%, compared to the original estimates of Y_c^*, provided by Chetty and Hendren (2018b). These estimators also perform noticeably better than Pooled estimates. The standard deviation of each MSE is calculated using a jackknife method, where I iteratively drop one observation at a time and recompute the MSE across the remaining data.

Table 2: Mean Squared Error (MSE)

Method	Variable	MSE	SD (MSE)	Improvement (%)
Relevance-Based Shrinkage (Linear)	$\hat{\mu}_c^{r,\text{Linear}}$	0.234	0.016	30.4%
Relevance-Based Shrinkage (Quadratic)	$\hat{\mu}_c^{r,\text{Quad}}$	0.223	0.019	33.6%
Relevance-Based Shrinkage (Binary)	$\hat{\mu}_c^{r,\text{Binary}}$	0.249	0.017	25.9%
Pooled Shrinkage	$\hat{\mu}_c^{\text{Pooled}}$	0.262	0.018	22.0%
Original	Y_c^*	0.336	0.0351	

The table compares the Mean Squared Error (MSE) of various relevance-based and pooled estimation methods relative to the benchmark $\hat{\mu}_c^f$. The Improvement column highlights the extent to which each estimator reduces the MSE compared to the original estimates Y_c^*.

This outcome is significant. It shows that applying the relevance-based empirical Bayes

technique to Y_c^*, by borrowing strength from geographically neighboring states, captures meaningful patterns. These patterns are similar to those identified by Chetty and Hendren (2018b) when they adjusted Y_c^* to align the estimates toward the expected income $\bar{Y}_c$ of permanent residents.

Figure 4 presents scatter plots that provide empirical evidence of bias reduction, albeit at the cost of increased variance. In all panels, the x-axis and the dotted black line represent the benchmark values; black dots represent the pooled estimates, and blue dots represent the relevance-based estimates. The fitted blue line for the relevance-based estimates lies closer to the 45-degree benchmark line compared to the pooled alternative, suggesting that the relevance-based estimator more effectively mimics the benchmark pattern on average. However, the wider dispersion of the blue dots, relative to the black ones, reflects higher variance—an expected consequence of estimating an additional term in the relevance-based formulation. Overall, as reported in Table 2, the total mean squared error (MSE) is reduced, confirming the net performance improvement of the relevance-based approach.

Additionally, the heatmaps in Figure 5—particularly the one based on the quadratic relevance function—visually confirm the effectiveness of relevance-based shrinkage. It recovers the underlying geographical structure present in the benchmark but missing from the original estimates. This improvement is reflected in the smoother and more locally coherent estimates across neighboring states.

One concern is that the data used to construct the relevance-based estimators overlap with the data Chetty and Hendren (2018b) used to construct their benchmark estimates, $\hat{\mu}_c^f$. This overlap can induce correlation between the error terms of the relevance-based estimators and those of the benchmark, relative to the true unobservable parameter of interest, thereby complicating the interpretation of MSE. To provide additional support for the performance of the relevance-based shrinkage—with evidence that it is not directly influenced by this overlap—I report in Table 3 the correlation between the amount of shrinkage induced by the adjustment, $\hat{\mu}_c^r - \hat{\mu}_c^{\text{Pooled}}$, and $\bar{Y}_c$. Our relevance correction is based on causal effect estimates derived from movers data in nearby states, whereas $\bar{Y}_c$ comes from an entirely separate dataset of permanent residents in the target location.

Table 3: Correlation with $\bar{Y}_c$

Estimator	$\mathbf{Corr}(\hat{\mu}_c^{\mathrm{r}} - \hat{\mu}_c^{\mathrm{Pooled}}, \bar{Y}_c)$
Relevance-Based Shrinkage (Linear)	0.401
Relevance-Based Shrinkage (Quadratic)	0.408
Relevance-Based Shrinkage (Binary)	0.18

These notable correlations indicate effective performance in capturing the underlying socioeconomic and characteristic patterns embedded in the permanent resident estimates, $\bar{Y}_c$.

To sum up, results show that relevance-based shrinkage outperforms pooled shrinkage in terms of MSE, bias, and capturing geographical heterogeneity, with the linear and quadratic designs of the relevance function performing better than the binary alternative.

5 Conclusion

In this paper, I introduced a novel theoretical framework to address the question of statistical relevance, emphasizing how information from more relevant samples should influence the analysis of a particular sample. Despite being recognized as a significant issue, this topic has been largely overlooked in the literature. To address this gap, I proposed the concept of a relevance-based prior, maintaining a fully nonparametric approach that accounts for heterogeneity, offering a theoretically grounded method to properly incorporate relevance into empirical Bayes estimations.

Using an example from Chetty and Hendren's (2018b) estimation of neighborhood causal effects on children's income, I demonstrated the difference made by accounting for relevance in the estimation process. I also utilized its improved estimates, which were based on combining information with permanent resident data, as a benchmark. I showed clear evidence of improvements in the estimation results after incorporating relevance adjustments.

Our method presents a promising approach for applications where both parametric and nonparametric estimators are widely applied. With increasingly detailed geographical data becoming available, it has the potential to provide more accurate and precise estimates for individual units or smaller areas. Notable applications include, but are not limited to, the estimation of hospital quality assessments (Chandra et al., 2016; Hull, 2020; Kunz et al., 2024), intergenerational mobility studies (Alesina et al., 2021; Deutscher et al., 2020;

Kenedi & Sirugue, 2023), discrimination and racial bias (Goncalves & Mello, 2021; Schendstok, 2024), and school effects and educational outcomes (Angrist et al., 2017; Bruhn, 2019; Brandt, 2023; Berg, 2024; Crema, 2024; Hanushek et al., 2024).

From a theoretical perspective, the concept of incorporating a relevance-based prior into empirical Bayes analysis extends beyond the current focus on shrinkage estimation. It offers significant opportunities for future research, including ranking and selection (Mogstad et al., 2024; Gu & Koenker, 2023; Klein et al., 2020; Andrews et al., 2024) and hypothesis testing (Efron, 2004; Efron, 2008; Liang, 2019; Bickel, 2023) within the empirical Bayes framework.

References

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2021). Intergenerational mobility in Africa. Econometrica, 89(1), 1-35.

Andrews, I., Kitagawa, T., & McCloskey, A. (2024). Inference on winners. The Quarterly Journal of Economics, 139(1), 305-358.

Angrist, J. D., Hull, P. D., Pathak, P. A., & Walters, C. R. (2017). Leveraging lotteries for school value-added: Testing and estimation. The Quarterly Journal of Economics, 132(2), 871-919.

Berg, P., Romero, M., & Singh, A. (2024). Preschool and primary school markets: Evidence from India. Unpublished manuscript.

Bickel, D. R. (2023). Interval estimation, point estimation, and null hypothesis significance testing calibrated by an estimated posterior probability of the null hypothesis. Communications in Statistics-Theory and Methods, 52(3), 763-787.

Brandt, K. (2023). When private beats public: A flexible value-added model with Tanzanian school switchers. Economic Development and Cultural Change, 72(1), 159-206.

Bruhn, J. (2019). The consequences of sorting for understanding school quality. Unpublished manuscript.

Chandra, A., Finkelstein, A., Sacarny, A., & Syverson, C. (2016). Health care exceptionalism? Performance and allocation in the US health care sector. American Economic Review, 106(8), 2110-2144.

Chetty, R., & Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. The Quarterly Journal of Economics, 133(3), 1107-1162.

Chetty, R., & Hendren, N. (2018b). The impacts of neighborhoods on intergenerational

mobility II: County-level estimates. The Quarterly Journal of Economics, 133(3), 1163-1228.

Cohen, N., Greenshtein, E., & Ritov, Y. A. (2013). Empirical Bayes in the presence of explanatory variables. Statistica Sinica, 333-357.

Crema, A. (2024). School competition, classroom formation, and academic quality. Technical report, Yale School of Management.

Deutscher, N., & Mazumder, B. (2020). Intergenerational mobility across Australia and the stability of regional estimates. Labour Economics, 66, 101861.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. Journal of the American Statistical Association, 99(465), 96-104.

Efron, B. (2008). Simultaneous inference: When should hypothesis testing problems be combined? The Annals of Applied Statistics, 2, 197–223.

Efron, B. (2011). Tweedie's formula and selection bias. Journal of the American Statistical Association, 106(496), 1602-1614.

Efron, B. (2016). Empirical Bayes deconvolution estimates. Biometrika, 103(1), 1-20.

Efron, B., & Hastie, T. (2018). Computer age statistical inference, student edition: algorithms, evidence, and data science (Vol. 6). Cambridge University Press.

Efron, B. (2019a). Rejoinder: Bayes, Oracle Bayes, and Empirical Bayes. Statistical Science, 34(2), 234–235.

Efron, B. (2019b). Bayes, oracle Bayes and empirical Bayes. Statistical Science, 34(2), 177-201.

Efron, B. (2021). Empirical bayes: Concepts and methods. In Handbook of Bayesian, Fiducial, and Frequentist Inference (pp. 8-34). Chapman and Hall/CRC.

Fay III, R. E., & Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, 74(366a), 269-277.

Gilraine, M., Gu, J., & McMillan, R. (2020). A new method for estimating teacher value-added (NBER Working Paper No. w27094). National Bureau of Economic Research.

Goncalves, F., & Mello, S. (2021). A few bad apples? Racial bias in policing. American Economic Review, 111(5), 1406-1441.

Green, E. J., & Strawderman, W. E. (1991). A James-Stein type estimator for combining unbiased and possibly biased estimators. Journal of the American Statistical Association, 86(416), 1001-1006.

Gu, J., & Koenker, R. (2023). Invidious comparisons: Ranking and selection as compound decisions. Econometrica, 91(1), 1-41.

Hanushek, E. A., Morgan, A. J., Rivkin, S. G., Schiman, J. C., Shakeel, A., & Sartain, L. (2024). The Lasting Impacts of Middle School Principals (No. w32642). National Bureau of Economic Research.

Ignatiadis, N., & Wager, S. (2022). Confidence intervals for nonparametric empirical Bayes analysis. Journal of the American Statistical Association, 117(539), 1149-1166.

James, W., & Stein, C. (1961, June). Estimation with quadratic loss. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 361–379). University of California Press.

Kiefer, J., & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. The Annals of Mathematical Statistics, 887-906.

Kenedi, G., & Sirugue, L. (2023). Intergenerational income mobility in France: A compar-

ative and geographic analysis. Journal of Public Economics, 226, 104974.

Klein, M., Wright, T., & Wieczorek, J. (2020). A joint confidence region for an overall ranking of populations. Journal of the Royal Statistical Society Series C: Applied Statistics, 69(3), 589-606.

Koenker, R., & Mizera, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. Journal of the American Statistical Association, 109(506), 674-685.

Kou, S. C., & Yang, J. J. (2017). Optimal shrinkage estimation in heteroscedastic hierarchical linear models. In Big and Complex Data Analysis: Methodologies and Applications (pp. 249-284). Cham: Springer International Publishing.

Kunz, J. S., Propper, C., Staub, K. E., & Winkelmann, R. (2024). Assessing the quality of public services: For-profits, chains, and concentration in the hospital market. Health Economics.

Liang, K. (2019). Empirical Bayes analysis of RNA sequencing experiments with auxiliary information. The Annals of Applied Statistics, 13(4), 2452-2482.

Luo, J., Banerjee, T., Mukherjee, G., & Sun, W. (2023). Empirical bayes estimation with side information: A nonparametric integrative tweedie approach. arXiv:2308.05883.

Mogstad, M., Romano, J. P., Shaikh, A. M., & Wilhelm, D. (2024). Inference for ranks with applications to mobility across neighborhoods and academic achievement across countries. Review of Economic Studies, 91(1), 476-518.

Schendstok, M. (2024). Racial Inequality in Labor Market Experiences in the United States. Unpublished manuscript.

Wang, Z., Parker, P. A., & Holan, S. H. (2025). Variational Autoencoded Multivariate Spatial Fay-Herriot Models. arXiv:2503.14710.

Yasui, Y., Liu, H., Benach, J., & Winget, M. (2000). An empirical evaluation of various priors in the empirical Bayes estimation of small area disease risks. Statistics in Medicine, 19(17-18), 2409-2420.

A Appendix

A.1 Proof of Theorem 1 (Empirical Bayes Posterior Mean with Side Information)

Combining Tweedie's formula with Bayes' rule yields the following compact derivation:

$$
\begin{aligned}
E(\mu_i \mid Y_i = y, X_i = x_0) &= y + \sigma^2 \frac{d}{dy} \log f(y \mid x_0) \\
f(y \mid x_0) &= \frac{\omega(x_0 \mid y)}{\omega(x_0)}\, f(y) \\
\log f(y \mid x_0) &= \log f(y) + \log \omega(x_0 \mid y) - \log \omega(x_0) \\
\frac{d}{dy} \log f(y \mid x_0) &= \frac{d}{dy} \log f(y) + \frac{d}{dy} \log \omega(x_0 \mid y) \\
\therefore\ E(\mu_i \mid Y_i = y, X_i = x_0) &= y + \sigma^2 \Big(\tfrac{d}{dy} \log f(y) + \tfrac{d}{dy} \log \omega(x_0 \mid y) \Big).
\end{aligned}
$$

A.2 Proof of Theorem 2 (Model Bias for the Relevance-Based Estimator)

Write

$$
A_h(y) \;=\; \int_{\mathbb{R}^d} \lambda_{x_0,h}(x)\, \frac{\omega(x \mid y)}{\omega(x)}\, dx.
$$

Changing variables $x = x_0 + h\,u$ ($dx = h^d\, du$) gives

$$
A_h(y) = \int_{\mathbb{R}^d} K(\|u\|)\, \frac{\omega(x_0 + h\,u \mid y)}{\omega(x_0 + h\,u)}\, du.
$$

A second–order Taylor expansion of the numerator around x_0 yields

$$
\omega(x_0 + h\,u \mid y) = \omega(x_0 \mid y) + h\,u^\top \nabla_x \omega(x_0 \mid y) + \tfrac{1}{2}\, h^2\, u^\top H_x \omega(x_0 \mid y)\, u + o(h^2).
$$

The denominator $\omega(x_0 + h\,u)$ admits an analogous expansion. Because the kernel K is symmetric, the linear term integrates to zero; hence

$$
A_h(y) = \frac{\omega(x_0 \mid y)}{\omega(x_0)} + \tfrac{1}{2}\, h^2 \int K(\|u\|)\, u^\top H_x \Big[\tfrac{\omega(x_0|y)}{\omega(x_0)} \Big]\, u\, du + o(h^2) = \frac{\omega(x_0 \mid y)}{\omega(x_0)} + O(h^2).
$$

Differentiating with respect to y,

$$\partial_y A_h(y) = \partial_y \Big[\tfrac{\omega(x_0|y)}{\omega(x_0)} \Big] + O(h^2).$$

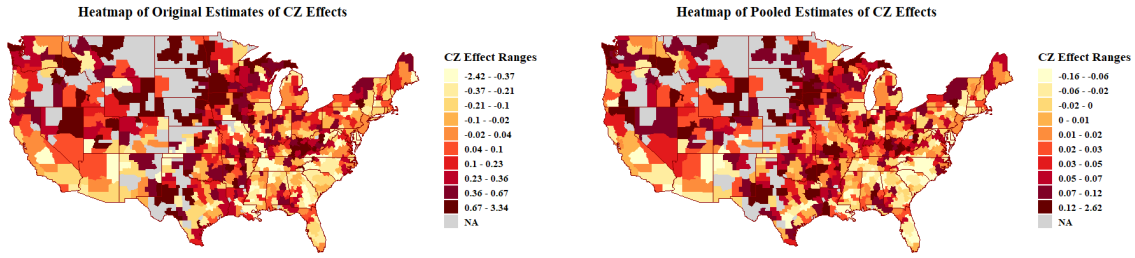
Therefore

$$\frac{\partial_y A_h(y)}{A_h(y)} = \frac{\partial_y \omega(x_0 \mid y)}{\omega(x_0 \mid y)} + O(h^2).$$

Substituting this into the posterior mean difference gives

$$\begin{aligned} & E\big(\mu_i \mid Y_i;\, \mu_i \sim g^r_{x_0,h}\big) \;-\; E\big(\mu_i \mid Y_i;\, \mu_i \sim g_{x_0}\big) \\ &\quad = \; \sigma^2 \Big[\partial_y \log A_h(y) \;-\; \partial_y \log \omega(x_0 \mid y) \Big]_{y=Y_i} \\ &\quad = \; O(h^2)\,. \end{aligned}$$

as required.



(a) **(b)**

Figure 1: Panel (a) depicts the original estimates of the causal effects of commuting zones on children's expected income whose parents are at the 25th percentile of income, as estimated by Chetty and Hendren (2018b). Panel (b) shows the commuting zone effects estimated using pooled shrinkage, $\hat{\mu}_c^{\text{Pooled}}$, assuming all unobservable commuting zone effects are drawn from a shared prior g. Each color range on the right represents a specific percentile of the data, and light grey areas indicate zones with unavailable estimates.

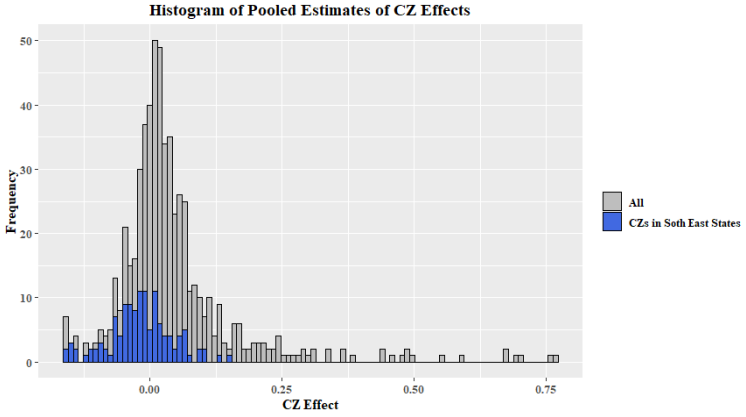


Figure 2: The larger histogram displays $\hat{\mu}_c^{\text{Pooled}}$, the pooled estimates of commuting zones' effects on children's expected income across the United States (US). In contrast, the blue histogram focuses on the same data but is limited to estimates from seven southeastern states: South Carolina, Georgia, North Carolina, Florida, Mississippi, Alabama, and Tennessee.

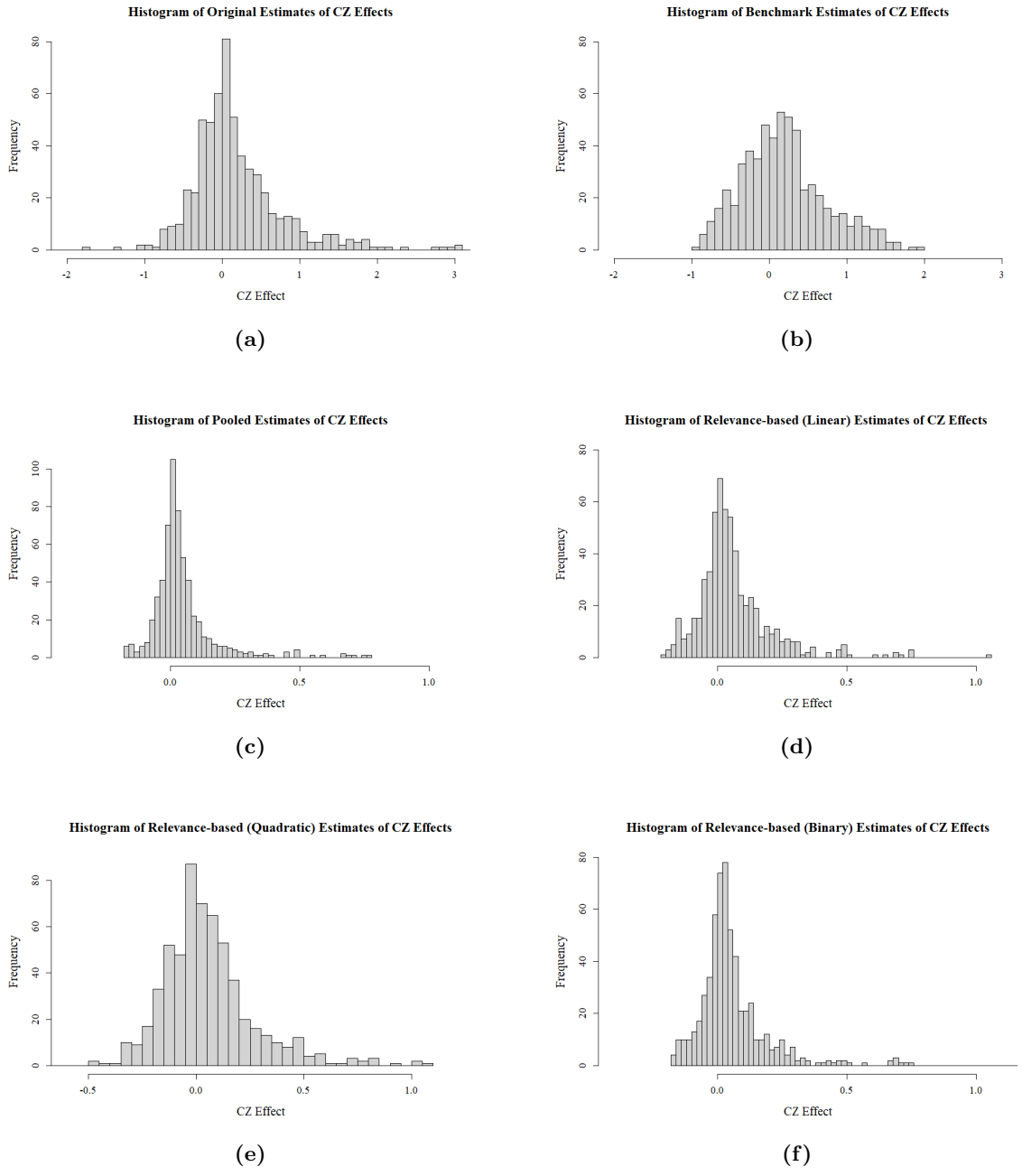
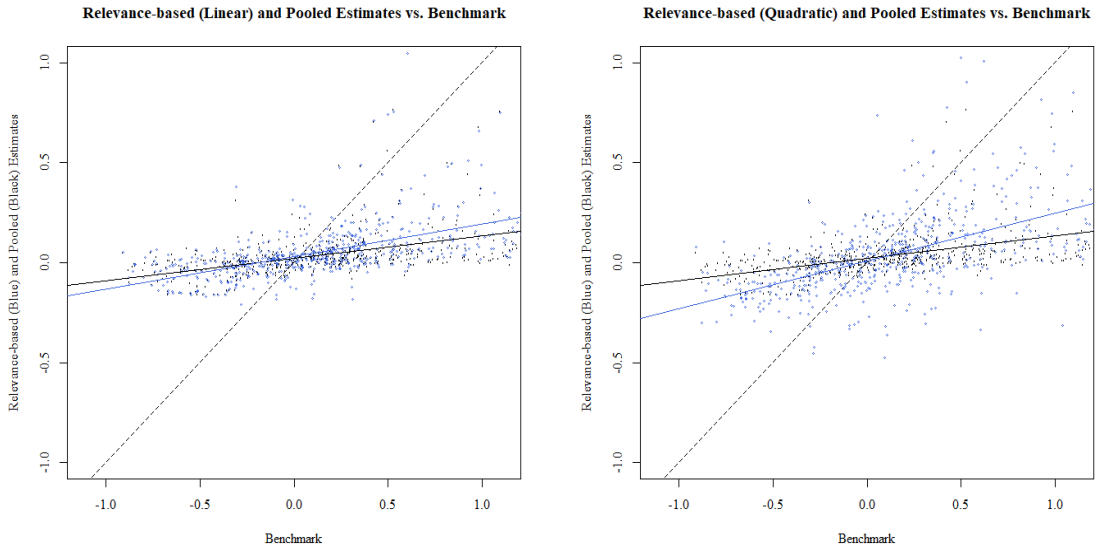
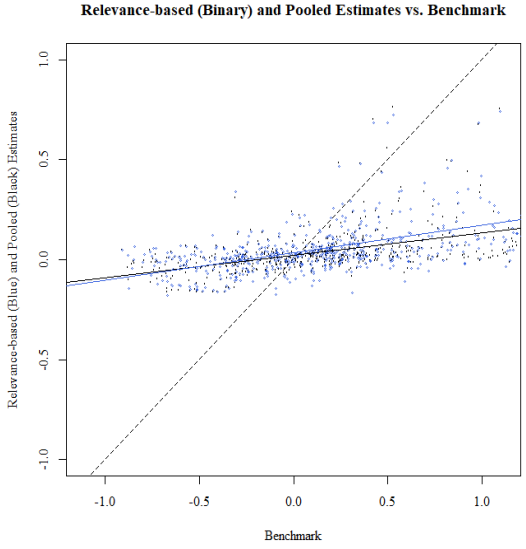


Figure 3: Histograms of estimated effects of commuting zones on children's expected income. (a) Original estimates Y_c^*. (b) Benchmark forecasts $\hat{\mu}_c^f$ from Chetty and Hendren (2018b). (c) Pooled estimates $\hat{\mu}_c^{\text{Global}}$. (d–f) Relevance-based estimates $\hat{\mu}_c^r$ using three different relevance functions: linear (d), quadratic (e), and binary (f).



(a)

(b)



(c)

Figure 4: Scatter plots comparing various estimators. The x-axis and the dotted line represent the benchmark estimates $\hat{\mu}_c^f$. On the y-axis, black dots represent the pooled estimates $\hat{\mu}_c^{\text{Pooled}}$, and blue dots represent the relevance-based estimates $\hat{\mu}_c^r$.

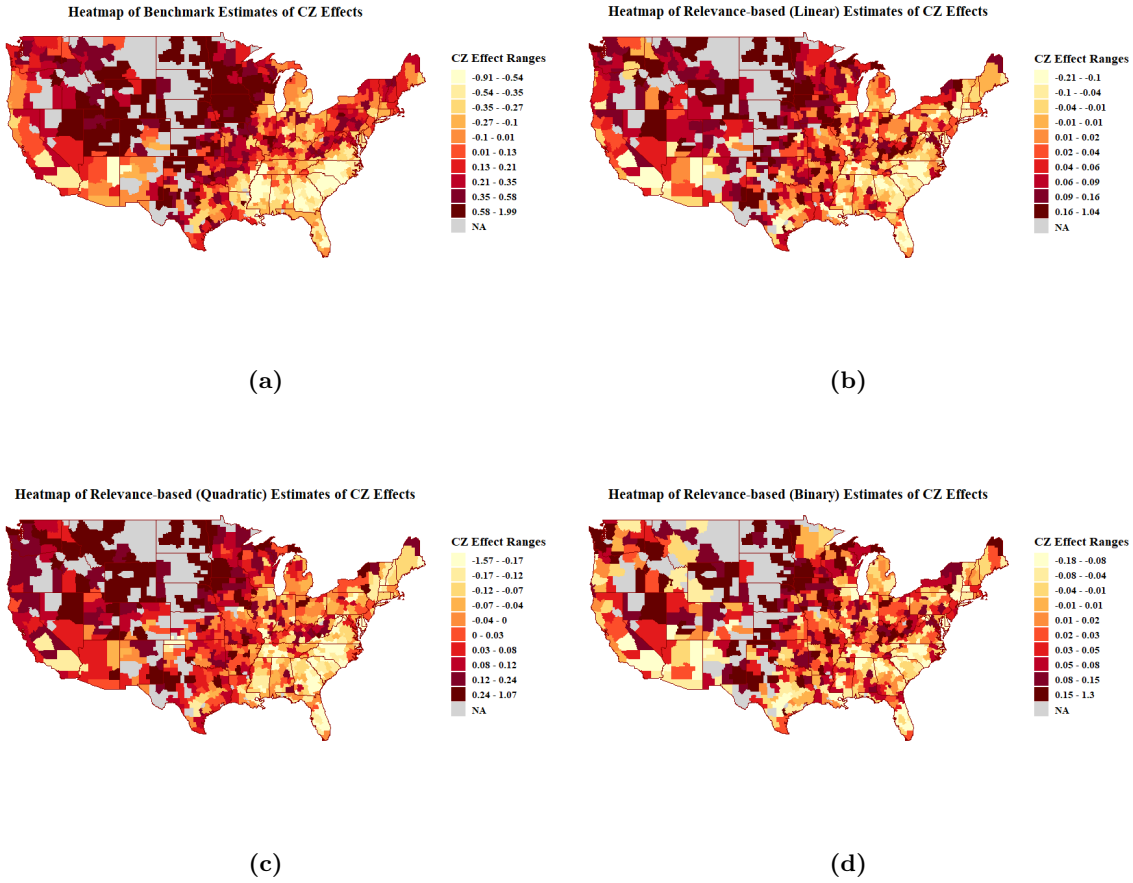


Figure 5: The figures show heatmaps of benchmark and relevance-based estimates of the commuting zone effects on children's expected income. Panel (a) displays the benchmark data $\hat{\mu}_c^f$ provided by Chetty and Hendren (2018b). Panel (b) represents the relevance-based estimates $\mu_c^{r,Linear}$, where shrinkage is applied using a linear relevance function inversely proportional to the spatial distance between states. Panel (c) depicts the relevance-based estimates $\mu_c^{r,Quad}$, where the relevance decreases quadratically with increasing spatial distance. Panel (d) shows the relevance-based estimates $\mu_c^{r,Binary}$, where only observations from regions within a fixed distance threshold contribute to the estimates, treating them as equally relevant. Each color range on the right represents a specific percentile of the data, and the light grey areas, primarily in the center, indicate zones where estimates are unavailable.