Fixed Effects Estimation via Spatially Fused PPML

Reigner Kane and Linghui Wu
ECON 31703

May 23, 2022

1 Introduction

Estimating location-specific fixed effects and detecting local clusterings for counts data is an important task in research that employs spatial data. This project is situated at the intersection of a handful of such literature. It is thoroughly motivated and informed by the theory and practice of modern quantitative spatial models, as employed in trade, urban, and other subfields of economics (Redding and Rossi-Hansberg, 2016; Allen et al., 2020). We attempt to provide a better answer to a specific question raised in this literature, corresponding with the estimation of a particular type of fixed effect. Additionally, however, we draw substantial inspiration from the literature across statistics and related disciplines. From prior work in these fields, we obtain not only the theory necessary to inspire our marginal innovation, but also the computational techniques used to implement them.

We now return to the key question we seek to answer: given a particular model of workers' choice of locations (that is, where they locate their residence and their workplace) and observations of commuting flows (which encode these location-pair choices), what can we infer regarding the expected economic fundamentals which characterize the possible residences and workplaces? This question is interesting not only in its own right, but because it lays the necessary groundwork for further counterfactual analysis, as an answer allows the researcher to fully characterize the discrete choice model under consideration Dingel and Tintelnot (2020). However, prior work in this and related settings has principally sought to perform parameter estimation without explicitly accounting for the spatial realities at play. In particular, we believe that a certain domain-specific intuition (namely, that economic fundamentals are generally continuous or smooth across space) can inform estimation procedures which more accurately recover the parameters of interest, and provide additional information regarding the city structure in the process.

Our results show promise for introducing spatially-fused penalization into the established estimation procedures. Applying to both simulated and real-world data, it appears that at least one of our proposed method can match or exceed the baseline procedure. This is particularly true if we believe that underlying economic fundamentals are smooth in space; our estimation approach harnesses this assumption to inform its results.

This paper is organized as follows. Section 2 discusses the empirical and computational strategies for the fused PPML estimation. Section 3 resorts to Monte Carlo simulations and presents the estimated results obtained from the generated datasets. In section 4, we take the fused lasso and the fused ridge models to the 2010 tract-level New York City commuting

data and compare the estimation results to the baseline fixed effects from the PPML gravity estimation. Section 5 concludes.

2 Empirical and Computational Strategy

2.1 PPML

In the paper which provides the motivation and underlying theoretical model, Dingel and Tintelnot (2020), the authors develop a model for individual's commuting choice. Workers are presumed to make a decision of where to live (indexed by $k \in \{1, \ldots, K\}$) and where to work (indexed by $n \in \{1, \ldots, N\}$) according to their shared expectations regarding the rents and wages (r_k and w_n, respectively) they would face in each location pair, along with the commuting costs for getting from k to n and back (δ_{kn}). By placing this decision in a multinomial logit framework and assuming idiosyncratic type 1 extreme value preference shocks, we obtain the following conditional choice probabilities:

$$\mathbb{P}\left(\ell_{kn} \mid \{r_k\}, \{w_n\}, \epsilon, \alpha\right) = \frac{w_n \left(r_k^{\alpha} \delta_{kn}\right)^{-\epsilon}}{\sum_{k',n'} w_{n'} \left(r_{k'}^{\alpha} \delta_{k'n'}\right)^{-\epsilon}}, \tag{1}$$

where L is the observed total number of individuals in the economy and ℓ_{kn} is the number of workers observed that commute between the tract pair (k, n).

It is crucial that, ex-ante, individuals' beliefs in wages and rents, as well as the commuting elasticity are latent. Instead, we observe the commuting flows ℓ_{kn} and the commuting costs $\bar{\delta}_{kn}$, aiming to produce estimates for the ex-ante price beliefs and the commuting elasticity. We impose an assumption on the unobserved commuting costs which allows us to exploit this observable variable for the purposes of estimation, namely

$$\mathbb{E}[\delta_{kn}^{-\epsilon} \mid \alpha, \epsilon, \delta] = \bar{\delta}_{kn}^{-\epsilon}.$$

We can now take advantage of the econometric literature on the estimation of gravity models via pseudo-Poisson maximum likelihood (PPML), popularized by Silva and Tenreyro Silva and Tenreyro (2006) and theoretically developed in Guimarães et al. (2003). Specifically, we reframe our problem as one of maximizing the following likelihood:

$$\log\big(L(\alpha_k^o, \alpha_n^d, \epsilon; \ell_{kn}, \bar{\delta}_{kn})\big) = \sum_{k,n} \ell_{kn} \left[\alpha_k^o + \alpha_n^d - \epsilon \log\big(\bar{\delta}_{kn}\big) - \log\left(\sum_{k'n'} \exp\big(\alpha_k^o + \alpha_n^d\big) \bar{\delta}_{k'n'}^{-\epsilon})\right)\right],$$

which produces identical results to minimizing the Poisson regression's negative log-likelihood with covariates α_k^o, α_n^d, and $\bar{\delta}_{kn}$.

2.2 Why Regularization?

Aided by the reinterpretation of solving the problem of estimating the parameters as Poisson regression, it is simple and computationally feasible to recover estimates for the α^o and α^d, as well as the elasticity ϵ. However, particularly in the estimation of the fixed effects, the

resulting estimation preserves none of the underlying spatial structure inherent in this model. When considering the spatial fixed effects for two neighboring tracts k and k', the baseline PPML procedure estimates them independently. Though, it aligns with the intuition that the wages and rents in adjacent tracts are not unrelated, especially when space is subdivided sufficiently finely (as in the granular settings Dingel and Tintelnot (2020) are interested in analyzing), we might expect that economic primitives associated with neighboring tracts are likely to be close to one another. This can be loosely formalized by imagining that these location fixed effects are realisations of some continuous function from longitude and latitude (that is, a compact subset of $\mathbb{R}^2$) to $\mathbb{R}$. Provided some additional regularity condition (such as Lipschitz continuity), we would necessarily have a bound on $|\alpha_k - \alpha_{k'}|$. The question, now, is how to implement this goal.

Similar questions have been addressed previously across other quantitatively-minded fields. For example, a similar intuition to ours is mentioned in Tibshirani and Taylor (2011), who apply the generalized lasso (with a linear model) to state-level H1N1 infections; Sass et al. (2021) apply this same notion of spatial lasso penalization to a problem in climate statistics; and finally Li and Sang (2019) once again apply a generalized lasso approach to the study of oceanic trends.

Returning to the econometric issue at hand, we present two new or expanded approaches (PPML with fused lasso and fused ridge), and address other alternatives we believe are less well-suited to this particular task.

2.3 Fused Lasso

Our application of fused lasso is positioned within the theoretical framework developed in Ali and Tibshirani (2019). The simplest, and original, form of fused lasso, was first developed in Tibshirani et al. (2005). Initially, it was intended for use in situations where there existed a) a natural ordering of covariates and therefore model parameters, and b) the researcher intended to impose sparsity on the *differences between subsequent parameter estimates*. However, in Ali and Tibshirani (2019), this idea was greatly expanded to include a much wider range of models. Specifically, rather than being restricted to penalizing only subsequent estimates in settings where a natural ordering existed, this expanded framework allows for a penalization term of the form

$$\lambda \|D\beta\| \tag{2}$$

where D is a $m \times p$ matrix, resulting in penalization for m arbitrary linear combinations of the components in β. Additionally, superseding the original application to linear models, this generalized form was shown to have desirable properties (namely unique solution) when such this type of penalization is applied to any generalized linear model. Relevant to our desired applications, this includes Poisson regression.

Moving beyond theoretical results to the question of computation, the construction of our estimation procedure begins with Tibshirani and Taylor (2011). The algorithm outlined there, and made accessible in the `genlasso` `R` package, is flexible to penalization as in 2. However, it only allows for an underlying linear model, and so is unsuitable for direct application here. In Choi et al. (2018), the authors expand upon the `genlasso` algorithm

to the Poisson case with fixed effects; however, their procedure only permits unit-specific fixed effects. The baseline that they develop, provided in the `pglasso` `R` package, is what we modify to develop the setting-specific algorithm needed.

Their approach is a hybrid and iterative one, exploiting the efficiency and accessibility of `genlasso` in conjunction with Majorize-Minimization (MM). At each step, this algorithm approximates the regularized likelihood minimization problem as a regularized linear least-squares problem, allowing its solution to be computed using `genlasso`. Specifically, given intial values for each of the α_k^o, α_n^d, ϵ, and the Poisson regression intercept β_0, the algorithm proceeds as follows at step m (initially described on page 545 on Choi et al. (2018), modified to this setting):

1. Compute the next value of $\beta = (\beta_0, -\epsilon)$ while holding fixed $\alpha = (\alpha^o, \alpha^d)$, using MM and the resulting closed-form solution, as documented on page 544 of Choi et al. (2018).
2. Compute the next value $\alpha^{(m)}$, by
   (a) Compute the second-order Taylor expansion of the negative log-likelihood, holding fixed $\beta^{(m)}$
   (b) Reformulate the resulting minimization problem as a generalized fused lasso problem with a linear least-squares objective function, and call `pglasso`.

The changes necessary are all contained in step 2, and are detailed below. First, recall the negative log-likelihood we are intending to minimize, subject to regularization:

$$-\log\big(L(\alpha, \beta; \ell_{kn}, \bar{\delta}_{kn})\big) = \sum_{n,k} \left[-\ell_{kn}\left(\alpha_k^o + \alpha_n^d - \epsilon \log\big(\bar{\delta}_{kn} + \beta_0\big)\right) + \exp\big(\alpha_k^o + \alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon} + \beta_0\right]. \tag{3}$$

Now, the MM procedure attempts to bound and approximate Equation **??** with a function which is more easily optimized. In the case of MM estimation applied to maximum likelihood estimation, we begin with the second-order Taylor expansion of the negative log-likelihood, appropriately penalized. Focusing on the un-penalized component of Equation **??**, we obtain the following partial derivatives:

$$\begin{aligned}
-\frac{\partial \log\big(L(\alpha; \beta, \ell_{kn}, \bar{\delta}_{kn})\big)}{\partial \alpha_{k^o}} &= \sum_n -\ell_{kn} + \exp\big(\alpha_k^o + \alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon}, \\
-\frac{\partial \log\big(L(\alpha; \beta, \ell_{kn}, \bar{\delta}_{kn})\big)}{\partial \alpha_{n^d}} &= \sum_k -\ell_{kn} + \exp\big(\alpha_k^o + \alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon},
\end{aligned}$$

which together characterize the gradient of interest. Furthermore, for any $k, k' \in \{1, \ldots, K\}$, we have

$$-\frac{\partial^2 \log\big(L(\alpha; \beta, \ell_{kn}, \bar{\delta}_{kn})\big)}{\partial \alpha_{k^o} \partial \alpha_{k'^o}} = \sum_n \left(\exp\big(\alpha_k^o + \alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon}\right) \mathbb{1}(k = k').$$

The analogous result holds for n and n'; define $\mu_k = \sum_n \exp\big(\alpha_k^o + \alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon}$, $\mu_n = \sum_n \exp\big(\alpha_k^o + \alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon}$, and $\mu_{kn} = \exp\big(\alpha_k^o + \alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon}$. This allows us to characterize the Hessian rather simply, as a block matrix of the form

$$H_\alpha(\alpha; \beta, \ell_{kn}, \bar{\delta}_{kn}) = \begin{pmatrix} A_\alpha & B_\alpha \\ B_\alpha^T & C_\alpha \end{pmatrix},$$

where A_α and C_α are a diagonal matrix with entries μ_k and μ_n, respectively, and B_α is a matrix with entries $(B_\alpha)_{ij} = \mu_{ij}$.

While this allows us to characterize the second-order Taylor expansion we need, it is alone insufficient. Instead, we are interested in further bounding by a simpler function; to do so, we need to establish bounds on the magnitude of the eigenvalues of H_α. Thankfully, due to the simple structure described above, we can do so by way of the Gershgorin Circle Theorem (GCT). In particular, GCT states that all eigenvalues of a square matrix lie within at least one of the following disks,

$$D_i = \{x \in \mathbb{C} \,|\, \|x - h_{ii}\| \leq R_i\},$$

where h_{ii} are the diagonal entries of the matrix, and

$$R_i = \sum_{j \neq i} |h_{ij}|.$$

This alone guarantees that our Hessian is positive semi-definite, because we have that $R_i^{(m)} = h_{ii}^{(m)}$, from the earlier derivations of $A_\alpha^{(m)}$ and $B_\alpha^{(m)}$. Additionally, we now have an explicit upper bound on the eigenvalues of $H_\alpha^{(m)}$, namely

$$\sigma^{(m)} := \max_i 2h_{ii}.$$

This bound allows us to return to the procedure outline by Choi et al. (2018). We now define

$$z^{(m)} := \alpha^{(m)} + \frac{\big(\mu^{o(m)}, \mu^{d(m)}\big)^T}{\sigma^{(m)}},$$

so that we have

$$\begin{aligned}
\frac{\sigma^{(m)}}{2}\left\|z^{(m)} - \alpha\right\|_2^2 &= \frac{\sigma^{(m)}}{2}\left(z^{(m)} - \alpha\right)^T\left(z^{(m)} - \alpha\right) \\
&= \frac{\sigma^{(m)}}{2}\left(\alpha^{(m)} - \alpha + \frac{\big(\mu^{o(m)}, \mu^{d(m)}\big)^T}{\sigma^{(m)}}\right)^T\left(\alpha^{(m)} - \alpha + \frac{\big(\mu^{o(m)}, \mu^{d(m)}\big)^T}{\sigma^{(m)}}\right) \\
&= \frac{\sigma^{(m)}}{2}\left(\left\|\alpha^{(m)} - \alpha\right\|_2^2 + 2\left(\alpha^{(m)} - \alpha\right)^T\left(\frac{\big(\mu^{o(m)}, \mu^{d(m)}\big)^T}{\sigma^{(m)}}\right) + \left\|\frac{\big(\mu^{o(m)}, \mu^{d(m)}\big)^T}{\sigma^{(m)}}\right\|_2^2\right) \\
&\geq \frac{\sigma^{(m)}}{2}\left[\left\|\alpha^{(m)} - \alpha\right\|_2^2 + 2\left(\alpha^{(m)} - \alpha\right)^T\left(\frac{\big(\mu^{o(m)}, \mu^{d(m)}\big)^T}{\sigma^{(m)}}\right)\right] \\
&\geq \frac{1}{2}\left(\alpha^{(m)} - \alpha\right)^T H_\alpha^{(m)}\left(\alpha^{(m)} - \alpha\right) + \left(\alpha^{(m)} - \alpha\right)^T\left(\big(\mu^{o(m)}, \mu^{d(m)}\big)^T\right),
\end{aligned}$$

where the final line is simply the second-order Taylor approximation of the negative log-liklihood, with the constant term dropped. The inequality there is guaranteed by the definition of $\sigma^{(m)}$. It immediately follows, then, that (up to higher-order approximation error)

$$\frac{\sigma^{(m)}}{2}\left\|z^{(m)}-\alpha\right\|_2^2+\lambda\|D\alpha\|_1\geq-\log\big(L(\alpha,\beta;\ell_{kn},\bar{\delta}_{kn})\big)+\lambda\|D\alpha\|_1.$$

Critically, the left side of the above inequalitiy conforms with the requirements of `genlasso`, and so the remainder of this step of the algorithm is left to the `fusedlasso` function within that package.

2.4 Fused Ridge

In addition to the fused lasso, we propose the use of fused ridge; that is, applying the ℓ_2 norm to arrive at a penalization term of the form $\lambda||D\alpha||_2^2$. In contrast with the fused lasso where the estimated fixed effects are required to be constant within a given cluster, fused ridge smooths the estimates across spaces.

Specifically, we minimize, with respect to the commuting elasticity ϵ, the origin and destination fixed effects $\{\alpha_k^o\}_{k=1}^K$ and $\{\alpha_n^d\}_{n=1}^N$,

$$\min_{\epsilon,\{\alpha_k^o\}_{k=1}^K,\{\alpha_n^d\}_{n=1}^N}\left\{l^{\text{ridge}}=\sum_{n,k}\left[-\ell_{kn}\left(\alpha_k^o+\alpha_n^d-\epsilon\log\big(\bar{\delta}_{kn}\big)\right)+\exp\big(\alpha_k^o+\alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon}\right]+\lambda\|D\alpha\|_2^2\right\}. \tag{4}$$

Denote $\mathcal{N}^o(k)$ and $\mathcal{N}^d(n)$ the neighbors of residence tract i and workplace tract j, and the ridge penalization term can be rewritten as

$$\sum_{(k,n)\in\mathcal{N}^o(k)}(\alpha_k^o-\alpha_n^o)^2+\sum_{(k,n)\in\mathcal{N}^d(n)}(\alpha_k^d-\alpha_n^d)^2. \tag{5}$$

Note that the objective function of the fused ridge is differentiable, unlike in the fused lasso case. This allows us to utilize standard gradient-based methods for solving the optimization problem above and recovering the fused ridge PPML estimates for a given penalty λ. We compute the partial derivatives for the fused ridge as follows.

$$\begin{aligned}
\frac{\partial l^{\text{ridge}}}{\partial\epsilon}&=\sum_{kn}\log\big(\bar{\delta}_{kn}\big)[\ell_{kn}-\exp\big(\alpha_k^o+\alpha_n^d\big)\bar{\delta}_{kn}^{-\epsilon}],\\
\frac{\partial l^{\text{ridge}}}{\partial\alpha_{k'}^o}&=\sum_n\left[-\ell_{k'n}+\exp\big(\alpha_{k'}^o+\alpha_n^d\big)\bar{\delta}_{k'n}^{-\epsilon}\right]+2\lambda\sum_{\tilde{k}\in\mathcal{N}(k')}(\alpha_{k'}^o-\alpha_{\tilde{k}}^o),\\
\frac{\partial l^{\text{ridge}}}{\partial\alpha_{n'}^d}&=\sum_k\left[-\ell_{kn'}+\exp\big(\alpha_k^o+\alpha_{n'}^d\big)\bar{\delta}_{kn'}^{-\epsilon}\right]+2\lambda\sum_{\tilde{n}\in\mathcal{N}(n')}(\alpha_{n'}^d-\alpha_{\tilde{n}}^d).
\end{aligned} \tag{6}$$

These, together, characterize the gradient which we will use to actually compute the fused ridge estimates.

2.5 Selection of Penalization Parameter

To search for the optimal value of the tuning parameter, we can apply cross-validation where the model is estimated on the training set and the prediction error is evaluated on the testing set. This search strategy can be done with a single training-validation split, or through multi-fold cross validation (Tibshirani et al., 2005). It is also applicable to use the information criterion to assess the model's performance. For example, Choi et al. (2018) employs the Bayesian information criterion to select the optimal tuning parameters over a grid of possible options.

However, due to practical considerations which limit the viability of either approach, the penalty terms here are determined by minimizing the prediction mean squared error (MSE). The challenges lie in a) the simulated and real data sets are cross-sectional rather than panel and different sample splittings would potentially result in different spatial dependency in the training and testing sets, and (b) the penalized model selection criteria require the number of estimated spatial clustering for calculation, while the fused ridged does not force the fixed effects estimates to be constant within a unit. Thus, determining the groups (or, rather in the case of ridge, the degrees of freedom) involves the use of additional methods beyond the scope of the paper.

2.6 Other Approaches

We would be remiss if we neglected to mention prior work which is clearly related to the goals laid out here. Given our interest in the spatial continuity of our quantities of interest (fixed effects), our setting may seem like a natural application of two-dimensional series regression. For example, we could understand α^o and α^d as evaluations of continuous functions f^o and f^d, both mapping a compact subset of $\mathbb{R}^2$ to $\mathbb{R}$. Through regularized series regression, we could recover estimates $\hat{f^o}$ and $\hat{f^d}$; however, this framework imposes some strong assumptions we are unwilling to accept in our setting. First and foremost, this approach requires that we assume global continuity. We are aware that in the real world, there are occasionally sharp boundaries (be they purely geographic, political, or social) which represent a true break point of a location relative to its neighbors. Such a feature of economic-geographic reality would both violate the usual assumptions of series regression and potentially cause practical problems with estimation.

Our setup also closely echoes the method of kriging in spatial and geostatistics, a method for optimally predicting values given a set of observed values Cressie (1990). The substantial literature now devoted to expanding and applying this technique, however, aim to predict *after observing some number of spatially distributed realizations*. However, in the context we are interested in, we are not attempting simply to optimally predict the commuting flows to and from any point in space. Rather, we are intending to work within the theoretical framework that requires us to recover estimates for spatial fixed effects (discretized over Census tracts) for the ultimate aims of counterfactual analysis. As such, kriging is ultimately not applicable to solving our specific problem.

3 Simulation Results

To better understand the properties and behavior of our proposed solutions, we apply them to simple simulated datasets. In particular, we use both fused lasso and fused ridge to estimate the spatial fixed effects and commuting elasticity on data simulated according to the following process. In all cases, we consider a 20×20 grid of locations, where we apply a fusion penalty along rook adjacency, and the observed commuting cost $\bar{\delta}_{kn}$ is determined by the taxicab distance (or, ℓ_1 distance) between cells k and n. Given the fixed effects which are established according to the rules below, we infer probabilities according to formula established in Equation 1. Finally, we make draws from the multinomial distribution characterized by those probabilities, with total population size $L = 500,000$. This value for the total labor supply was selected to result in approximately the same ratio between L and the number of locations as in the true New York City dataset, which has $L \approx 2.5$ million and $\sim 2,000$ locations.

DGP_1: **The "Neighborhood-constant" DGP:** Numbering the grid cells in lexicographic order, we enforce rectangular groupings on the origin fixed effects, and group all locations together for the destination fixed effects. Then, values for the fixed effects are assigned to each of these groups.

In the context of the true theoretical model, this would correspond to uniform wage expectations across our grid city, and rents in a cell are entirely characterized by the neighborhood which contains the cell.

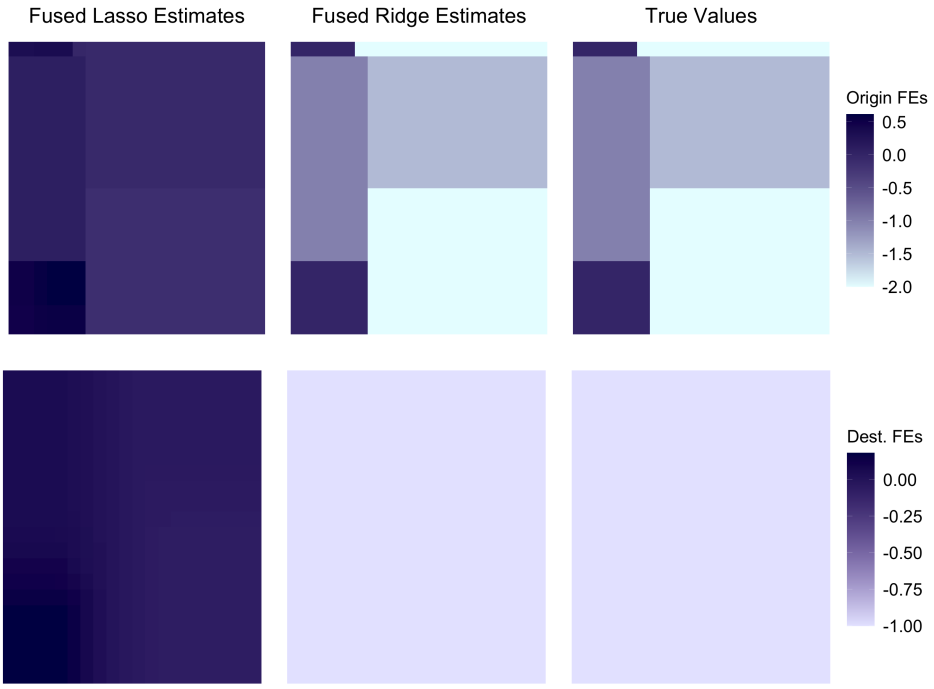
DGP_2: **The "Independently Drawn" DGP:** Rather than establishing groups on the grid, we allow for every fixed effect to be instead drawn independently from a normal distribution. This results in no structural relationship between the cells, and should be understood as the opposite extreme as DGP_1.

Our simulation results seem to suggest that the fused ridge approach is highly capable of recovering the true parameter values, while the fused lasso approach struggles. The results from DGP_1 are visualised in Figure 1. which is essentially tailored to the strengths of the fused lasso method, both the fused lasso and fused ridge procedures were able to recover the apparent boundaries of the neighborhoods. (Of course, the fused ridge does so only up to some threshold, whereas the fused lasso explicitly reports the groups).

We also notice another strong tendency in the fused lasso results: they seem to be severely biased towards zero. This is easily seen even in the absence of any direct penalization on the magnitude of the estimates themselves. We believe this may be indicative of an unknown error in the estimation algorithm itself, or perhaps due to numerical issues when using `genlasso` with too high-dimensional models. Regardless, this characteristic is made exceptionally clear in the estimation of destination fixed effects, which are constant at -1 in truth, but are estimated very close to 0 by fused lasso.

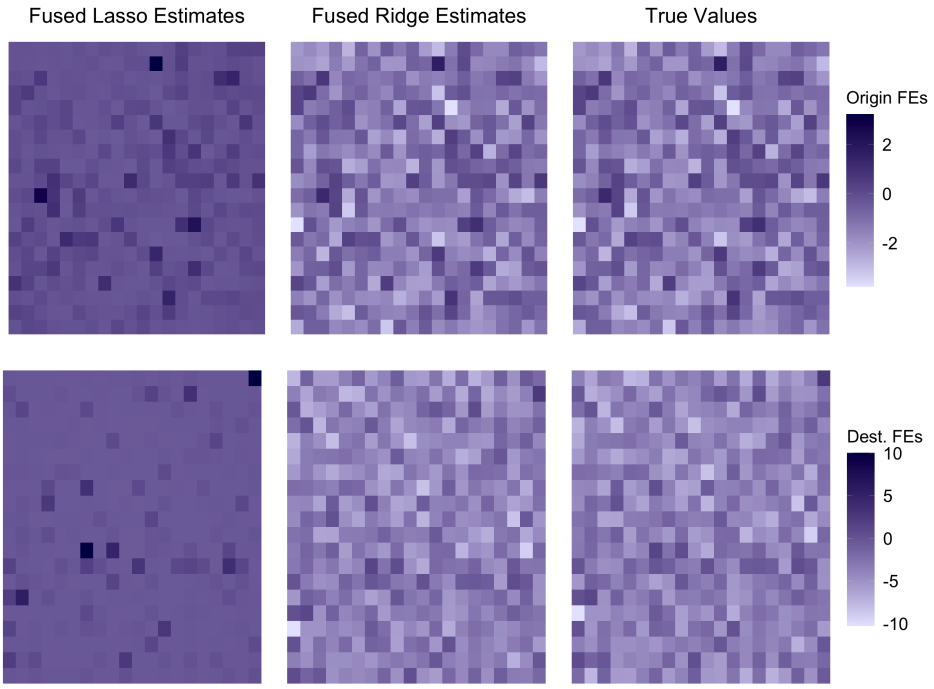
We now turn our attention to the alternative case, DGP_2. We might expect, before seeing results, that 1) the fused lasso procedure would not behave especially well (as this essentially corresponds to applying standard lasso to a model that does not truly possess sparsity), and 2) that fused lasso would behave better, likely with a small optimal penalization value. This

Figure 1: Estimates and True Values for Spatial Fixed Effects, DGP_1



intuition is encapsulated in the results displayed in Figure 2. Indeed, we see that fused ridge does a much better job of capturing the fine-grained variation in the underlying fixed effects. The fused lasso, while it is able to pick out a few prominent features, is far too aggressive in its imposing of local homogeneity, resulting in estimated values which are quite far from the true values. From these simple examples, then, it seems that the fused ridge approach is substantially more robust to failure of our local smoothness assumption.

Figure 2: Estimates and True Values for Spatial Fixed Effects, DGP_2



4 New York City Results

We now examine the fixed effects estimates from the gravity model and the fused PPML models on real word data. The 2010 tract-level New York City commuting data come from the Longitudinal Employment-Household Dynamics, Origin-Destination Employment Statistics (LODES). We obtain the neighboring structures of New York City census tracts from American Communities Project to construct the origin and the destination adjacency matrices, where we consider rook-adjacency. Following Dingel and Tintelnot (2020), we compute the observable commuting costs as

$$\bar{\delta}_{kn} = \frac{H}{H - t_{kn} - t_{nk}}, \tag{7}$$

where $H = 9$ hours and t_{kn} is the public transit time from origin tract k to destination tract n according to Google Maps collected by Davis et al. (2019).

We compare the estimated origin and destination fixed effects from the baseline PPML, the fused lasso and the fused ridge regressions. As discussed before, we expect estimates from the gravity model to be independently distributed where the predicted fixed effects can exhibit arbitrarily large jumps in the values for adjacent geographical units, those from the fused lasso model show clustering structures where the estimates within the unit are identical, and those from the fused ridge model that display both grouping structure and variations within the group.

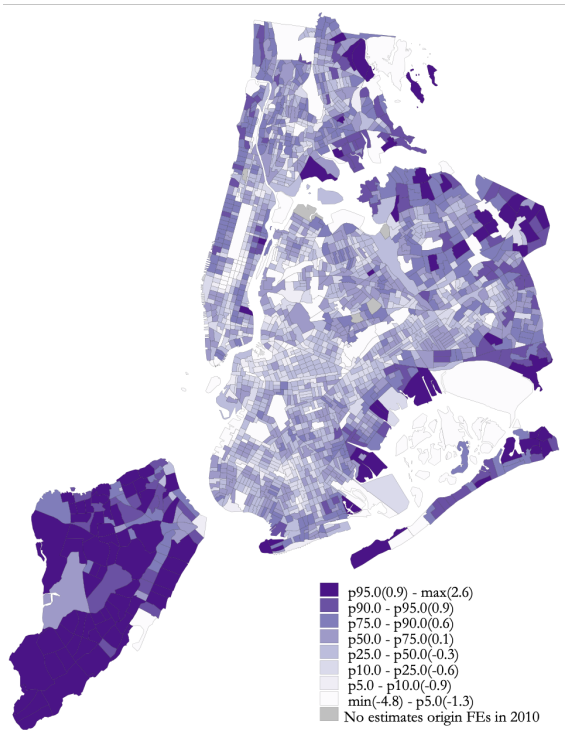
The estimated commuting elasticities from the three models are -7.896 from the unpenalized procedure, -6.818 from the fused lasso, and -7.934 from the fused ridge. Figure 3 hence reports the predicted residence fixed effects. The PPML gravity regression estimates

the fixed effects for residence tracts of the largest ranges, from -1.3 to 2.6, as shown in Figure 3a, and the estimates are distinctive and discrete values. By contrast, the fused lasso and ridge estimates (Figures 3b and 3c) exhibit clustering structures. There exist 289 distinctive groups from the fused lasso estimation, given the MSE-optimal penalization value. Furthermore, the estimates from the models with the regularization are reduced in magnitude, the penalization drives the fixed effects for the neighboring tracts towards each other. Though, it is noteworthy that the origin fixed effect estimates is influenced by its neighboring tracts, and are not unbiased and asymptotically consistent (Rinaldo, 2009). For instance, the estimated residence fixed effects for Central Park is -3.117 in the PPML mode, while are -0.990 and 1.522 from the fused lasso and the fused ridge models, which indicates beliefs in relatively low rents for living in Central Park. (See Figure 5 for a visualization comparing the empirical distributions of the fixed effects produced by each procedure)

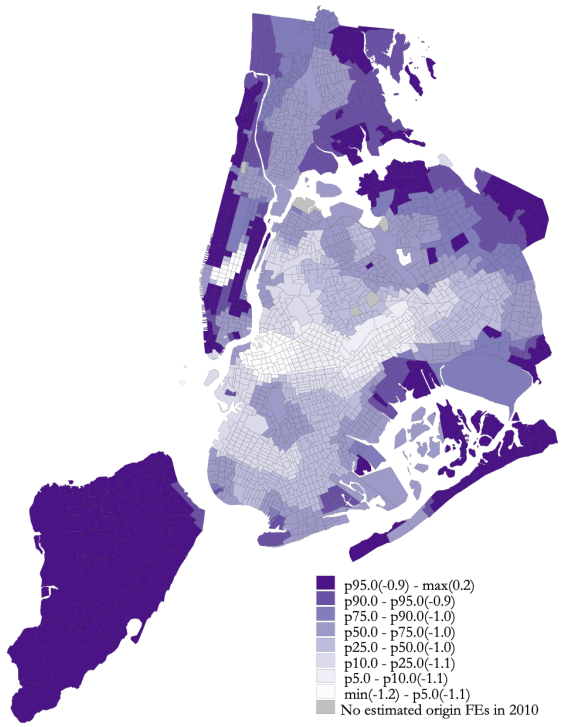
Figure 4 shows the predicted workplace fixed effects. The destination fixed effect estimates from the PPML model are granular and distinctive, with drastic changes appearing for neighboring tracts. On the other hand, the fused lasso groups the estimated workplace fixed effects into 403 different units, with the bottom 10 percentiles associated with Brooklyn and the upper five percentiles linked to Manhattan and Staten Island. The fused ridge model allows for greater variation in the estimates.

Figure 3: Estimated Residence Fixed Effects

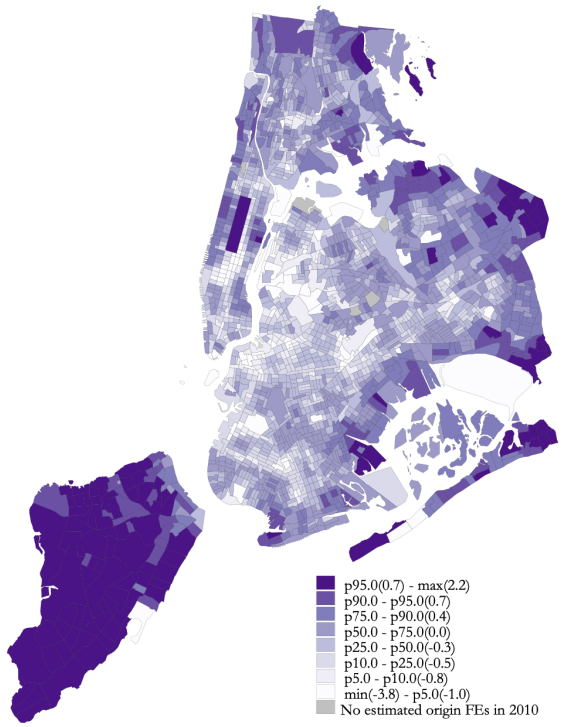
(a) PPML



(b) Fused lasso



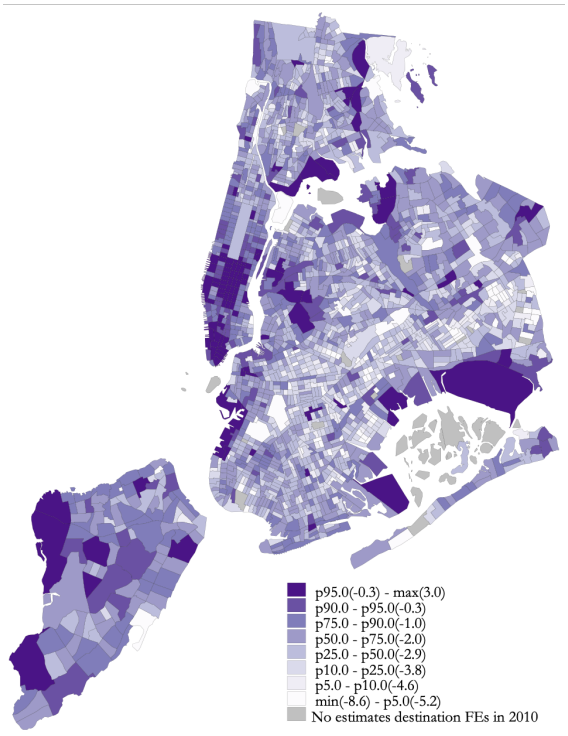
(c) Fused ridge



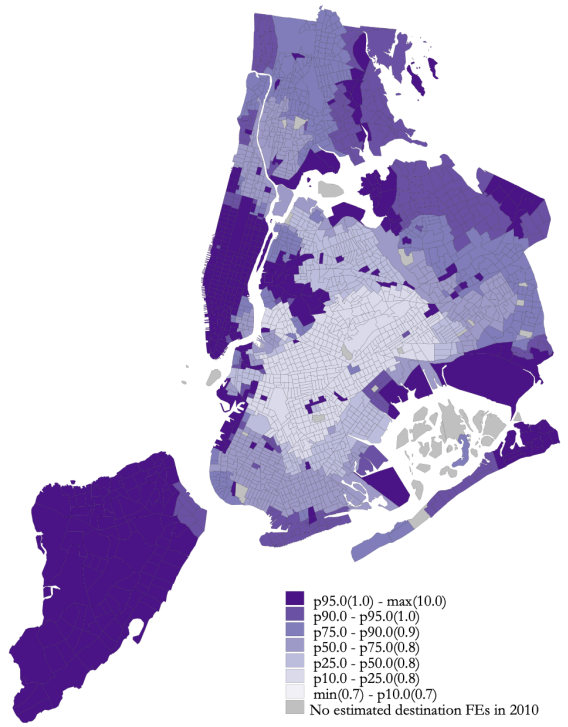
NOTES: Panel 3a plots the estimated origin fixed effects from the benchmark PPML model. Panel 3b plots the estimates from the fused lasso model with the optimal penalty parameter $\lambda = 100$. Panel 3c plots the estimates from the fused ridge model with the optimal penalty parameter $\lambda = 3650$. The legends reports the percentiles corresponding to the fixed effects estimation.

Figure 4: Estimated Workplace Fixed Effects

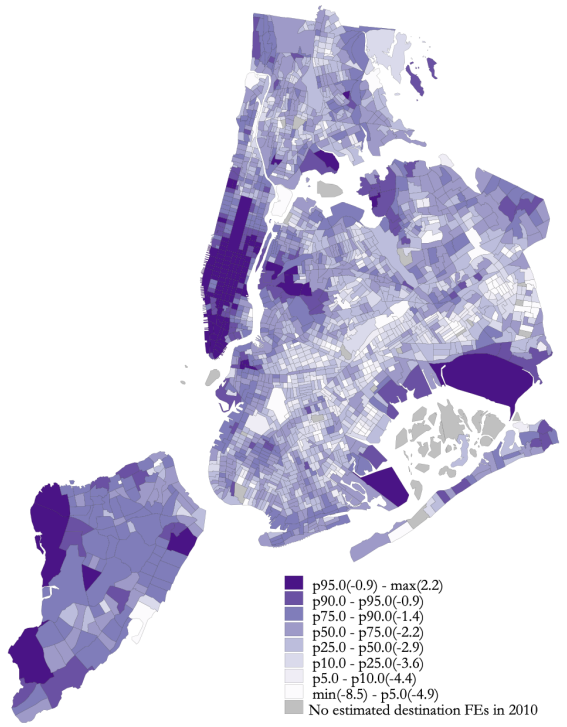
(a) PPML



(b) Fused lasso

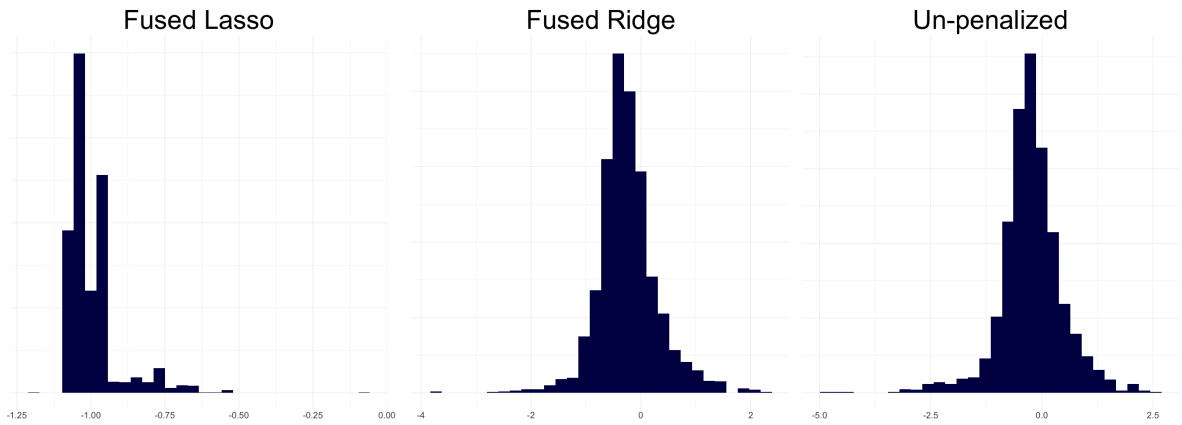


(c) Fused ridge

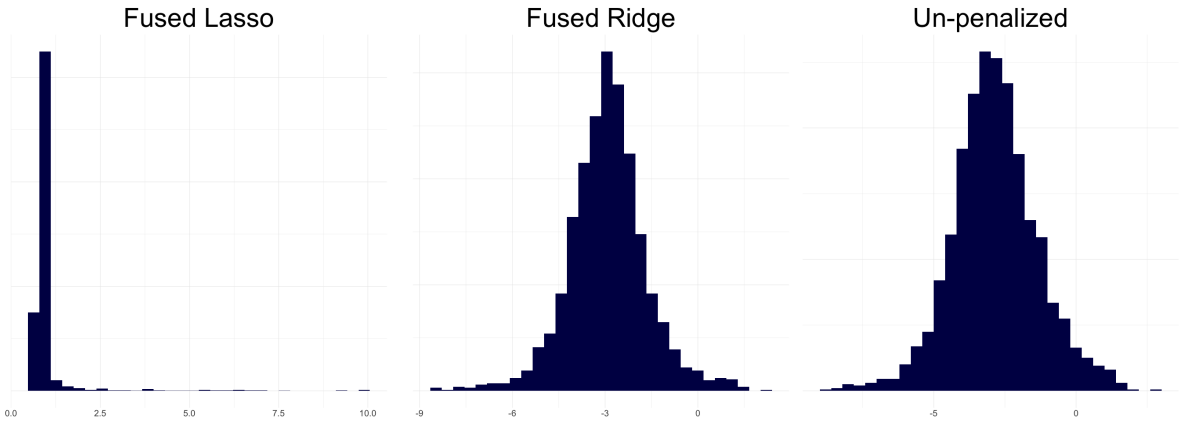


NOTES: Panel 4a plots the estimated origin fixed effects from the benchmark PPML model. Panel 4b plots the estimates from the fused lasso model with the optimal penalty parameter $\lambda = 100$. Panel 4c plots the estimates from the fused ridge model with the optimal penalty parameter $\lambda = 3650$. The legends reports the percentiles corresponding to the fixed effects estimation.

Figure 5: Empirical Distribution of Fixed Effects Estimates



(a) Residence Fixed Effects



(b) Workplace Fixed Effects

5 Conclusion

In this project, we have developed, implemented, and applied two approaches for handling spatial parameters in PPML estimation: PPML with fused ridge and PPML with fused lasso. From our estimates on simple simulated datasets, as well as the apparent behavior on real-world data for which the theoretical model is well-suited, we come to the following two conclusions: a) the fused ridge appears to behave well, smoothing out the fixed effects estimates across space, and b) the fused lasso appears to underperform, producing estimates which are insufficiently dispersed to seem plausible. As such, we believe that fused ridge is a promising tool in the context of estimation procedures for spatial data, specifically PPML estimation with spatial fixed effects. Areas of ripe for further inquiry include alternate means for choosing the optimal penalization parameter, domain-specific optimization procedures to lessen the computational burden of general-purpose gradient-based methods, and further analysis of real-world behavior on other DGPs.

References

Ali, Alnur and Ryan J. Tibshirani, "The generalized lasso problem and uniqueness," *Electronic Journal of Statistics*, 2019, *13* (2).

Allen, Treb, Costas Arkolakis, and Yuta Takahashi, "Universal Gravity," *Journal of Political Economy*, 2020, *128* (2), 393–433.

Choi, Hosik, Eunjung Song, Seung sik Hwang, and Woojoo Lee, "A modified generalized Lasso algorithm to detect local spatial clusters for count data," *AStA Advances in Statistical Analysis*, 2018, *102* (4), 537–563.

Cressie, Noel, "The origins of Kriging," *Mathematical Geology*, 1990, *22* (3), 239–252.

Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales, "How segregated is urban consumption?," *Journal of Political Economy*, 2019, *127* (4), 1684–1738.

Dingel, Jonathan I and Felix Tintelnot, "Spatial economics for granular settings," 2020.

Guimarães, Paulo, Octávio Figueirdo, and Douglas Woodward, "A tractable approach to the firm Location Decision Problem," *Review of Economics and Statistics*, 2003, *85* (1), 201–204.

Li, Furong and Huiyan Sang, "Spatial homogeneity pursuit of regression coefficients for large datasets," *Journal of the American Statistical Association*, 2019, *114* (527), 1050–1062.

Redding, Stephen and Esteban Rossi-Hansberg, "Quantitative spatial economics," 2016.

Rinaldo, Alessandro, "Properties and refinements of the fused lasso," *The Annals of Statistics*, 2009, *37* (5B), 2922–2952.

Sass, Danielle, Bo Li, and Brian J. Reich, "Flexible and fast spatial return level estimation via a spatially fused penalty," *Journal of Computational and Graphical Statistics*, 2021, *30* (4), 1124–1142.

Silva, J. M. and Silvana Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Tibshirani, Robert, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight, "Sparsity and smoothness via the fused Lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2005, *67* (1), 91–108.

Tibshirani, Ryan J. and Jonathan Taylor, "The solution path of the generalized lasso," *The Annals of Statistics*, 2011, *39* (3).