Recalculating ... : How Uncertainty in Local Labor Market Definitions Affects Empirical Findings*

Andrew Foote[1], Mark J. Kutzbach[2], and Lars Vilhuber[1,3]

[1]Center for Economic Studies, U.S. Census Bureau
[2]Federal Deposit Insurance Corporation
[3]Labor Dynamics Institute, Cornell University

June 21, 2017

[Preliminary Draft - Do not cite without authors' permission]

Abstract

This paper evaluates the use of commuting zones as a local labor market definition. We revisit Tolbert and Sizer (1996) and demonstrate the sensitivity of definitions to two features of the methodology. We show how these features impact empirical estimates using a well-known application of commuting zones. We conclude with advice to researchers using commuting zones on how to demonstrate the robustness of empirical findings to uncertainty in definitions.

1 Introduction

Local labor markets are an important unit of analysis in labor economics. Theoretical papers emphasize characteristics of a local labor market including common wage and rent levels (Roback, 1982; Moretti, 2011) as well as job-finding and unemployment rates (Head and Lloyd-Ellis, 2012; Schmutz and Sidibé, 2014) and often assume fixed or variable costs for

*The analysis, conclusions, and opinions expressed herein are those of the author(s) alone and do not necessarily represent the views of the U.S. Census Bureau or the Federal Deposit Insurance Corporation. All results have been reviewed to ensure that no confidential information is disclosed, and no confidential data was used in this paper. This document is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Much of the work developing this paper occurred while Mark Kutzbach was an employee of the U.S. Census Bureau.

transferring jobs or workers between labor markets. In empirical labor economics, researchers interested in estimating the effect of some local, exogenous shock on labor market outcomes must decide how to define the set of affected jobs or workers. Researchers examining labor markets in the United States often turn to one of several standard geographic definitions that are widely known and compatible with publicly available economic data, including: states (Blanchard and Katz, 1992; Wozniak, 2010; Kennan and Walker, 2011), metropolitan areas (Bound and Holzer, 2000; Card, 2001; Notowidigdo, 2011; Diamond, forthcoming), and counties (Monte, Redding and Rossi-Hansberg, 2015; Foote, Grosz and Stevens, 2015).

Another labor market definition with advantages for some research topics over the above definitions is commuting zones, which are composed of counties and were originally defined by Tolbert and Sizer (1996) (henceforth, TS). Commuting zones are similar to metropolitan areas in that they are meant to capture areas of economic integration that do not necessarily conform to regional political boundaries (such as states) (Office of Management and Budget, 2000, 2010). Unlike metropolitan areas, commuting zones have no urbanized area size requirements and span the entire United States, allowing researchers to measure effects for the entire country rather than just the set of metropolitan areas (or the complements of metropolitan areas within a state).

Given these features, commuting zones have been used in a number of influential papers in the labor economics literature, including Autor, Dorn and Hanson (2013), as well as Chetty et al. (2014), Amior and Manning (2015), Restrepo (2015), and Yagan (2016). Despite their widespread use, to the best of our knowledge, the methodology underlying commuting zone definitions and its impact on empirical estimates has not received much scrutiny and many researchers do not consider how findings may be sensitive to design issues.

Our paper makes two contributions for empirical analysis using commuting zones. First, we describe two methodological issues that researchers should be aware of when they use the commuting zone definitions. Second, we show how these methodological issues impact empirical estimates using Autor, Dorn and Hanson (2013) as an example.

Our findings suggest that researchers should consider evaluating the sensitivity of their results, and we propose two ways that researchers can test if their results are robust to the uncertainty inherent in this definition of local labor markets. These findings are informative to the use of commuting zones for defining local labor markets specifically, but also suggest caution for researchers in general when measuring treatment effects in geographically distinct areas where treatment may not be as discretely related to geography as is implied by the unit of measure.

The remainder of the paper proceeds as follows. We describe the data we use and the commuting zone definitions and methodology in detail in Section 2. In Section 3 we outline the extent to which commuting zone definitions are sensitive to data inputs and design decisions, and in Section 4 we discuss how those issues affect empirical estimates and provides guidance for researchers in light of our results. Section 5 concludes and discusses next steps.

2 Commuting Zone Data and Methodology

The Economic Research Service (ERS), an agency under the U.S. Department of Agriculture for which commuting zones were originally developed, distributes definitions on its website.[1] Commuting Zones are especially relevant for the economic analysis of rural areas, a focus of ERS, because they include all counties, not just urban counties.

As an alternative to metropolitan statistical areas, or Core Based Statistical Areas (Office of Management and Budget, 2013), many researchers examining local labor markets use commuting zones, which also combine sets of counties based on commuting flows, because they cover the entire country. However, few researchers are familiar with the methodology used to develop these zones. To that end, this section describes the methodology used by Tolbert and Sizer (1996) in developing the zones.[2]

[1]ERS has released commuting zone definitions based on 1980, 1990, and 2000 commuting data. All three definitions are available at http://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.aspx.

[2]The methodology was originally used in Tolbert and Killian (1987), but the 1996 paper is much more

We describe two especially important design components: the dissimilarity matrix, which measures how "far" nodes are from one another, and the clustering method, which decides how nodes are assigned to groups.

2.1 Dissimilarity Matrix

The dissimilarity matrix, D, is a representation of the relative distance between all pairs of counties. TS calculate D (or conversely, the similarity matrix P), where an entry D_{ij} is the dissimilarity of county i from county j, as below:[3]

$$D_{ij} = 1 - \frac{f_{ij} + f_{ji}}{\min(rlf_i, rlf_j)} \tag{1}$$

In the above equation, f_{ij} is the flow of commuters who live in county i and work in county j and f_{ji} is the opposite flow. The resident labor force in county i is $rlf_i = \sum_j f_{ij}$ (including f_{ii}), with a similar calculation for j. Normalizing flows with the minimum resident labor force of a pair upweights the association of outlying areas with metropolitan cores. Note that disimilarity is symmetric, so $D_{ij} = D_{ji}$.

TS1996 compute D using flows from the 1990 Journey-to-Work data, which tabulates the commuting information from the 1990 Census Long-form (U.S. Census Bureau, 1992).[4] The Census Bureau estimates county-to-county flows among 3,141 county equivalents for persons age 16 and older who reported being employed in the week prior to April 1, 1990.[5] The Economic Research Service maintains the 1990 commuting zones.

widely cited, and the zones from that paper are the ones most commonly used. For more background on the development of commuting zones, see Fowler, Rhubart and Jensen (2016).

[3]The clustering method used requires a dissimilarity matrix; one is just the complement of the other, by element, so $D_{ij} = (1 - P_{ij})$.

[4]Journey to Work data on county to county commuting flows are available for the 1990 Census, the 2000 Census, and 5-year samples of the American Community Survey at https://www.census.gov/hhes/commuting/data/commutingflows.html.

[5]Employment status is based on responses to the question "Did this person work at any time LAST WEEK?" Place of work is geocoded using the response to "At what location did this person work LAST WEEK?" Residence location is compiled from the mailing frame of the Census.

2.2 Clustering Method

After constructing this dissimilarity matrix, TS use it as an input into their clustering method. In general, data scientists use clustering methods to assign interrelated items, or items with similar features, into groups. In their application, TS use the average-linkage hierarchical clustering algorithm (PROC CLUSTER in SAS software).[6] The hierarchical clustering method uses the dissimilarity matrix in the following way. To begin, every county is its own cluster. Then, it finds the lowest value D_{ij} in the dissimilarity matrix, and combines those two counties together. It then recalculates the dissimilarity values between the new cluster and all other clusters. For each pair of clusters C_K and C_L, the dissimilarity, D_{KL}, is calculated as:

$$D_{KL} = \frac{1}{N_K N_L} \sum_{i \in C_k} \sum_{j \in C_L} D_{ij}, \tag{2}$$

where and D_{ij} is calculated as in Equation 1 and N_K and N_L are the count of nodes, or counties in this case, in each cluster. The process continues until all nodes are clustered, though the designer may stop the process by choosing a maximum "cutoff" height, H, such that if $D_{KL} > H$, then K and L do not merge. TS1990 uses clusters with distances up to 0.98.

We illustrate this process in Appendix Figure A1, which shows the hierarchical progression of how counties are clustered together for California. In the top left-hand corner, only a few counties have joined at a height of $H = 0.80$. As we increase the height to 0.88 and to 0.96, more counties are joined together. Finally, at a height of 1, almost all the counties have merged together, forming one large cluster and a few much smaller clusters.

[6]The hierarchical clustering for this paper using PROC CLUSTER was generated using SAS software, Version 9.2 of the SAS System for Unix. Copyright ©2009 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

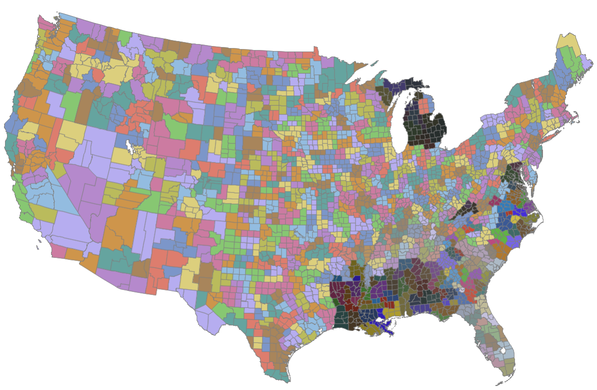
2.3 Our Replication

In order to replicate the clustering result in TS, which we refer to as TS1990, we use the 1990 Census JTW data and the methodology described above, with one important exception. Because of computing power constraints in 1996, TS divided the country into six overlapping regions and performed the clustering algorithm on each region separately, and then manually resolved conflicts in overlapping regions. This decision has two consequences for users: first, the height cutoffs across regions are not the same, because there is a normalization step before the algorithm merges observations. Second, it induces some subjectivity, since conflicts in cluster assignment for counties in the overlapping regions are inevitable.

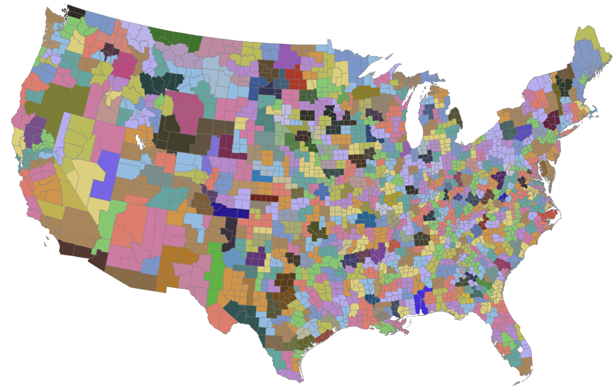
Rather than follow their methodology of dividing the country into regions, which required a subjective expert review, we run the hierarchical clustering algorithm on the entire country. We choose a height cutoff, 0.9365 (compared to 0.98), that most closely replicates their original zones in terms of size distribution, though it results in 810 zones (compared to 741). We find that the clustering algorithm, when attempting to produce the same cluster count as TS1990 with a national run, retains a residual cluster that spreads across many states. Only with the lower cluster height, and more clusters, does the residual cluster break up. We leave evaluation of alternate clustering methods for future work, with the emphasis in the present analysis on the robustness of zone definitions.

TS1996's zones and our replication are in Figure 1, and our summary statistics comparing the zones are in Table 1. While our replication does not perfectly match their zones and the cutoff height differs, this is not surprising, because we did not split the country into overlapping regions. Our replication has more evenly sized clusters with fewer clusters made up of a single county. We refer to our replication as FKV1990.

Figure 1: Replication of Commuting Zones from TS1996: County Mapping



Commuting Zones - TS1990



Replication of Commuting Zones - FKV1990

Table 1: Replication of TS1990 Commuting Zones: Summary Statistics

	TS1990	FKV1990
Mean Cluster Size	4.24	3.88
Median Cluster Size	4	4
Number of Clusters	741	810
Number of Singletons	62	16

Notes: Both TS1990 and FKV1990 are based on JTW tabulations from the 1990 Census. Summary statistics for TS1990 are from Table 8 of TS.

3 Design Sensitivity

While commuting zones are used by researchers as a convenient measure of local labor markets, they have a number of shortcomings for empirical research that are not regularly discussed in the literature. In this section, we evaluate the sensitivity of commuting zone definitions, focusing on two aspects of the TS1990 methodology. First, we show that if there is uncertainty in the input data, the resulting commuting zone definitions can vary substantially. Second, the resulting clusters are sensitive to the decision of when to stop merging clusters, which implies that small changes in the chosen cutoff height can affect the number and size of the clusters. Overall, this uncertainty and subjectivity in the commuting zone definitions contributes to conventional standard errors understating the true level of uncertainty in the estimates, which we show when we return to this issue with our empirical replication in the next section.

3.1 Sensitivity of Clustering Results to Underlying Error

Given the reliance of TS1990 on the commuting flows data, we want to analyze the extent to which the outputs of the TS methodology are sensitive to errors in this data. First, recall Equation 1 for the entries of the dissimilarity matrix. If f_{ij} is measured without error, then the distance between counties i and j is also measured without error. However, if the flows are measured with error, ϵ_{ij}, then we actually have an estimate of D_{ij}, $\hat{D}_{ij}$, which can be expressed as below (assuming without loss of generality that $rlf_i < rlf_j$):

$$\begin{aligned}\hat{D}_{ij} &= 1 - \frac{f_{ij} + \epsilon_{ij} + f_{ji} + \epsilon_{ji}}{r\hat{l}f_i} \\ &= 1 - \frac{f_{ij} + f_{ji}}{rlf_{ij} + \sum_j \epsilon_{ij}} + \frac{\epsilon_{ij} + \epsilon_{ji}}{rlf_{ij} + \sum_j \epsilon_{ij}}\end{aligned}$$

Even if $E[\epsilon_{ij}] = 0$, that does not imply that $E[\hat{D}_{ij}] = D_{ij}$. Furthermore, we cannot rely on the limit properties of the error distribution, because we only have one realization of the commuting flow, which is calculated from survey responses. Additionally, we know that $\frac{\epsilon_{ij}}{f_{ij}}$ is larger for small flows. This will increase D_{ij} for some small counties and decrease it for others. Because of the hierarchical nature of the clustering method, this error will affect the formation of all other clusters in the data.[7]

To demonstrate how this measurement error affects the outcome of the clustering procedure, we use the published margins of error (MOE) from the 2009-2013 ACS Journey to Work data to calculate the ratio MOE_{ij}/f_{ij} for different sized flows.[8] We project these ratios onto the flow bins from the 1990 Journey to Work data (which does not publish margins of error), to have an estimated MOE.[9] Using these estimated MOEs, we then obtain different realizations

[7] Additionally, because heights are normalized in the procedure, it also affects where the effective cutoff is, even for counties unaffected by errors in flows.

[8] These flow size bins are the following percentile bins: 0-50; 50-90; 90-95; 95-99; and 99+.

[9] There are other possible projections of the margins of error from one dataset to another. The Cen-

of the commuting zones in the following way:

1. For each origin-destination pair (i, j), we draw ϵ_{ij} from a normal distribution with mean 0 and standard deviation $MOE_{ij}/(1.64)$, since the MOE is scaled to be the 90% confidence interval.[10]

2. Calculate the new flow value, $\hat{f_{ij}} = f_{ij} + \epsilon_{ij}$, with negative values set to zero.

3. Re-calculate each dissimilarity matrix entry $\hat{D}_{ij}$.

4. Re-run the hierarchical clustering procedure, using the same cutoff as the replication.

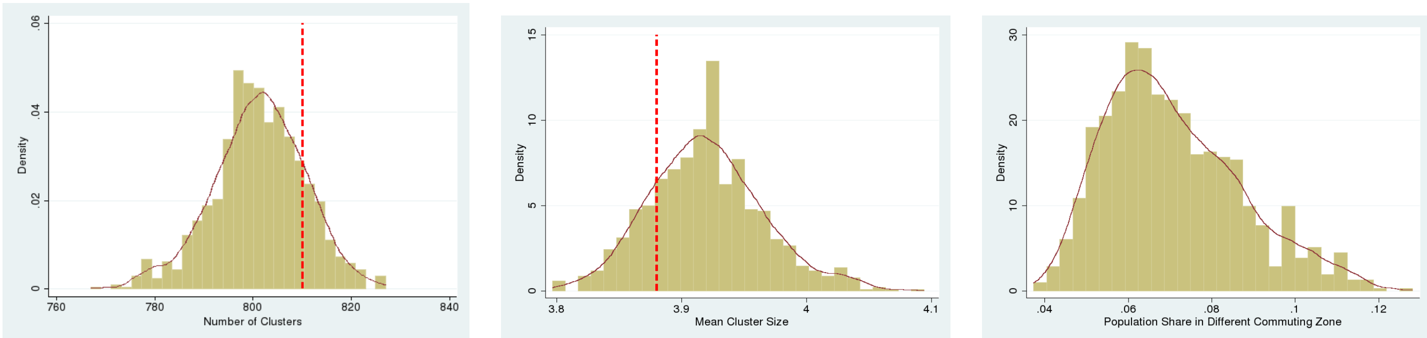
5. Store the new clusters, and calculate the following statistics: average number of counties in a cluster; number of clusters; and total number of counties in a different cluster than the one they were originally assigned.

We iterate over this procedure 1000 times in order to obtain distributions for these statistics. These graphs are shown in Figure 2, where the red vertical dashed lines are the values that would be obtained using only the published figures using our replication height. The figures show that the average cluster size varies considerably from the result the published figures would yield. Additionally, the share of the population that is mismatched is larger than 5% of the US population on average, which implies that using commuting zones to assign treatment mis-measures treatment for over 5% of the population. Additionally, this

sus Long form is designed to be a one-in-six sample for one year, while the ACS 5-year summary is designed to 5 years with a one-in-fifty sample each year. The smaller sample size typically results in higher margins of error in the ACS for comparable statistics. The uncertainty implied by our implementation likely overstates the underlying MOEs in the 1990 JTW. For more information on the construction of the ACS MOEs, refer to https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/MultiyearACSAccuracyofData2013.pdf, pages 10-12.

[10] In doing this, we assume that $\epsilon_{ij} \perp \epsilon_{ik} \forall k$, for simplicity. In reality, it is likely that $corr(\epsilon_{ij}, \epsilon_{ik}) < 0$, which means in our setting that we are understating the error by treating them as independent. In the JTW data, there are likely some origin-destination pairs that are not reported due to the sample design. In our current resampling approach, we only resample from non-zero flows in the data. A more complete approach would be to model the likelihood that a zero reported is actually a positive flow, and resample accordingly. This modeling is beyond the scope of this paper. For more detail on the 1990 Decennial Census sample design, consult https://www2.census.gov/prod2/decennial/documents/D1-D90-PUMS-14-TECH-01.pdf.

Figure 2: Results from Re-sampling Commuting Flows



(a) Number of clusters (b) Average Number of Counties (c) Mismatched Counties

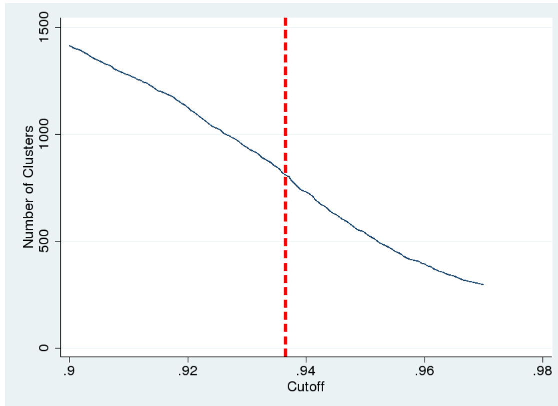
uncertainty in the cluster definitions is exacerbated by the sharp cutoff imposed in cluster analysis.

One complication with this sampling method is that observed flows f_{ij} are bounded by zero, though the published MOE may extend below zero. Many small flows, 65 percent, are not distinguishable from zero and are at risk to be censored, but because these tend to be small flows, they account for only 1.7 percent of jobs. Even so, our procedure will leave larger remaining flows, which decreases the distance between counties. This censoring implies that for a given cutoff, C, and two flows matrices, F and $\hat{F}$, where $\hat{F}$ is a resampled version of F, the flows matrix F would expect to have more clusters than $\hat{F}$, because more clusters merge at heights below C. This feature of the resampling is why the distribution in Figure 2 is not centered on the replicated estimate.

3.2 Choosing Cluster Height

Another sensitive feature of the methodology used by TS1996 is choosing the cutoff value above which no clusters can form, which determines the number of clusters. Tolbert and Killian (1987) describe the algorithm for choosing a cutoff value as follows (see page 15): "As a rule of thumb, a normalized average distance of 0.98 was considered sufficient distance between sets of counties to treat them as separate [Labor Market Areas]." The article does not provide an analysis of the sensitivity to changing the cutoff marginally up or down. Tolbert

Figure 3: Effect of Cluster Height on Number of Clusters



Note: Authors' calculations using methodology outlined in Section 2.

and Sizer (1996), in an effort to minimize methodological differences between commuting zones for 1980 and 1990, use the same cutoff with no further evaluation for the 1990 data. In this subsection, we investigate how sensitive the resulting clusters are to the choice of the cutoff value.

Figure 3 shows the number of clusters that form at various height cutoffs using the national 1990 JTW data, with the vertical line indicating the cutoff value we chose to replicate TS1990 (0.9418). The key takeaway from this figure is that it is theoretically ambiguous where a researcher should choose to stop merging clusters.[11] Additionally, increasing or decreasing the cutoff has implications for the number of resulting clusters. Increasing it to 0.9428 decreases the number of clusters by 19, while using a cutoff of 0.9408 cause the number of clusters to increase by 17.

As we described above, the measurement error in commuting flows causes some uncertainty in terms of the true dissimilarity matrix, and hence the true cluster heights. Because of the presence of a strict cutoff, some clusters that would have formed if D_{ij} were measured without error do not form, and vis-versa. More broadly, TS provide no empirical guidance for choosing the optimal cutoff and cluster size other than referring to expert knowledge.

[11]Decisions on clustering methods, clustering counts, and validation criteria depend on the application and are inherently somewhat subjective. Because clustering is an unsupervised method, there may be no indication of the ideal number of clusters (Halkidi, Batistakis and Vazirgiannis, 2001).

While outside the scope of the current paper, future work may explore data-driven methods to determine whether there is an optimal number of clusters for certain uses.

4 Empirical Sensitivity

In the previous section, we showed that there are a number of margins on which clustering methodologies are sensitive: uncertainty in the input data and the choice of the number of clusters. However, these issues are only important for empirical labor economists to the extent that these sensitivities impact empirical estimates in a meaningful way. To that end, in this section, we demonstrate the impact these issues have on the empirical findings of a well-known paper that uses commuting zones.

Autor, Dorn and Hanson (2013) (hereafter, ADH) estimate the impact that increased trade competition from China had on manufacturing employment in the United States. The magnitude of the main finding has been widely discussed and debated in economics and in the popular press.[12] To estimate this effect empirically, they use variation in the initial distribution of manufacturing employment at the commuting zone level, and national increases in imports from China by manufacturing subsector. Because ADH use commuting zones as their definition of local labor markets, their empirical analysis may be impacted by the issues outlined above.[13]

Their main estimating equation in the paper is as follows:

$$\Delta L^m_{it} = \gamma_t + \beta_1 \Delta IPW_{uit} + \beta_2 X_{it} + e_{it} \tag{3}$$

Where ΔL^m_{it} is the decadal change in manufacturing employment in Commuting Zone i following year t, ΔIPW_{uit} is the import exposure measure for the United States, and X_{it} are

[12]For example, see The Economist, March 11, 2017 "Economists argue about the impact of Chinese imports on America" http://www.economist.com/news/finance-and-economics/21718513-china-shock-has-not-been-debunked-it-worth-understanding.

[13]We want to acknowledge that the authors have been incredibly helpful in the process of replicating their paper, both in providing data and helping to troubleshoot, and were receptive to this exercise.

control variables. All regressions are weighted by population share of the commuting zone.

4.1 Replication of ADH

Since we use slightly different methods of aggregating data, we compare the main estimates from ADH (Table 1 in their paper) to our replication, which we show in Table 2.[14] Each cell in the table is a coefficient from a different regression, and for simplicity we just display estimates for the time period 1990-2000 (other results available upon request). The first column shows the estimates from their paper, while the second column changes the import exposure measures to our replicated measure. In the third column, we use our estimate of the change in manufacturing employment and weights. The fourth column clusters on commuting zone, and the final column shows the estimate using our replication of commuting zones. The estimate using our replicated commuting zones is somewhat smaller than the original estimates, but still statistically significant and of a similar size and magnitude.

Overall, the estimates are considerably stable, giving us confidence that we are properly replicating their main finding. We now turn to demonstrating how these estimates are affected by the concerns with the commuting zone definitions themselves.

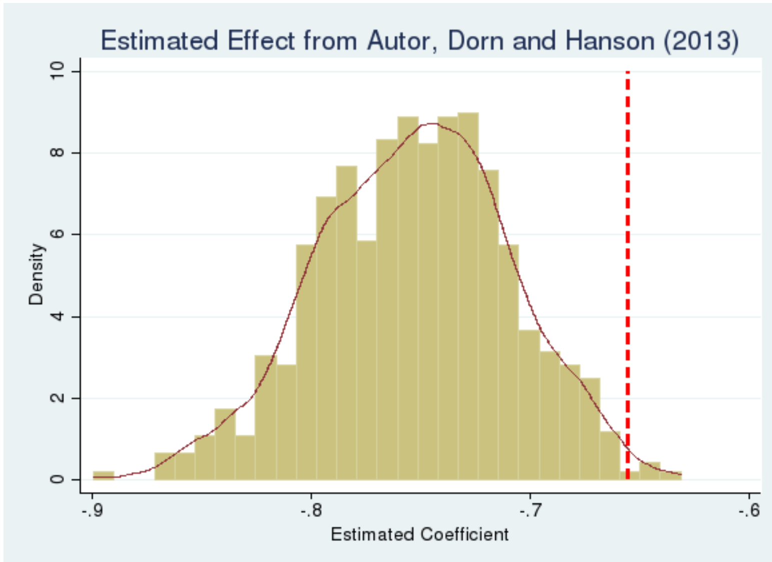
[14] ADH use individual level IPUMS data, which as a PUMA geography, and assign those observations to commuting zones based on population weights (more detail at http://www.ddorn.net/data.htm). We just use county-level tabulations, which aggregate to the commuting zone level.

Table 2: China Syndrome Replication and Comparison, 1990-2000

	ADH Estimate	Our RHS	Our LHS and Weight	CZ Clustering	Using FVK1990
$\Delta IPW_{cz,t}$	-0.8875	-0.8871	-0.8748	-0.8748	-0.6556
	(0.1812)	(0.1811)	(0.1527)	(0.1243)	(0.1110)

Notes: Table from author's calculations, using data from Autor, Dorn and Hanson (2013) and constructed data, based on equation 3. Column 1 is Table 2, Column 1 from ADH (2013). Column 2 replaces their measure of import exposure to ours. Column 3 replaces their measure of change in manufacturing employment and CZ-specific weights with ours. Column 4 does not cluster on state. Standard errors are in parentheses. All coefficients are significant with p-values less than 0.01.

Figure 4: Distribution of Effect, 1990-2000



Note: Histogram plots estimates of β_1 from equation 3, based on commuting zone realizations as outlined in Section 3. Red vertical line shows estimates using replicated FKV1990 commuting zones.

4.2 Sensitivity to Errors in Flows Data

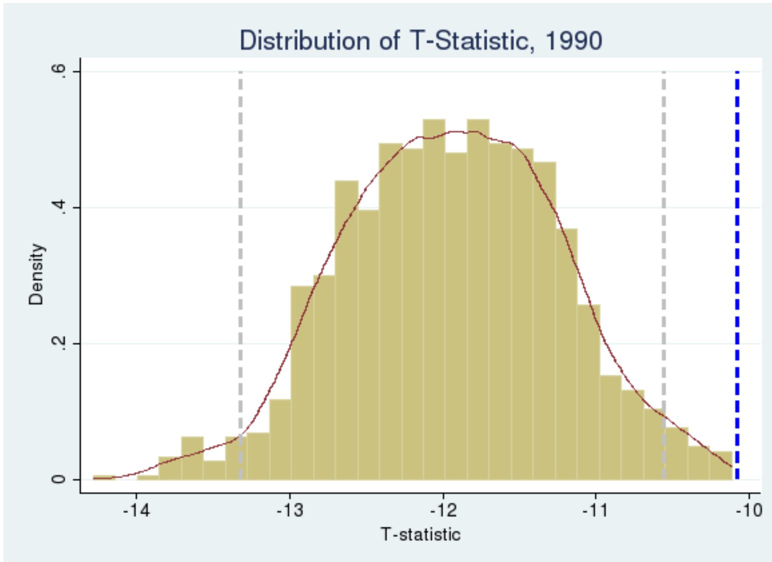
To demonstrate how sensitive the results of Equation 3 are to different commuting definitions, we re-estimate the equation using the 1000 realizations of commuting zones that were generated in the previous section.

The coefficients from this exercise are graphed in Figure 4, which shows the distribution of the estimated effect for the 1990-2000 period. The red vertical line shows the estimate using the published flows data from our national replication of TS1990.[15]

Another way to summarize the results of this exercise is to look at the distribution of t-statistics, which incorporates information about se_{β_1} into the analysis as well, and comparing that distribution to the critical values. To use the distribution of t-statistics in an empirical setting, researchers construct a 95% confidence interval of the t-statistic by using the values at the 97.5th and 2.5th percentiles of the 1000 realizations. If this confidence interval is

[15] One reason the distribution is not centered on the red vertical line is that the realizations of commuting zones have systematically larger clusters than the commuting zones using published figures; this issue is also discussed in the previous section.

Figure 5: Distribution of T-Statistic, 1990-2000



Note: Histogram plots t-statistics derived from estimating equation 3, based on commuting zone realizations as outlined in Section 3. Blue vertical line is t-statistic using FKV1990, and gray vertical lines are the 2.5th and 97.5th percentiles of the t-statistic distribution.

outside the critical value $t_{0.025}$, then the null hypothesis can be rejected at $\alpha = 0.05$.

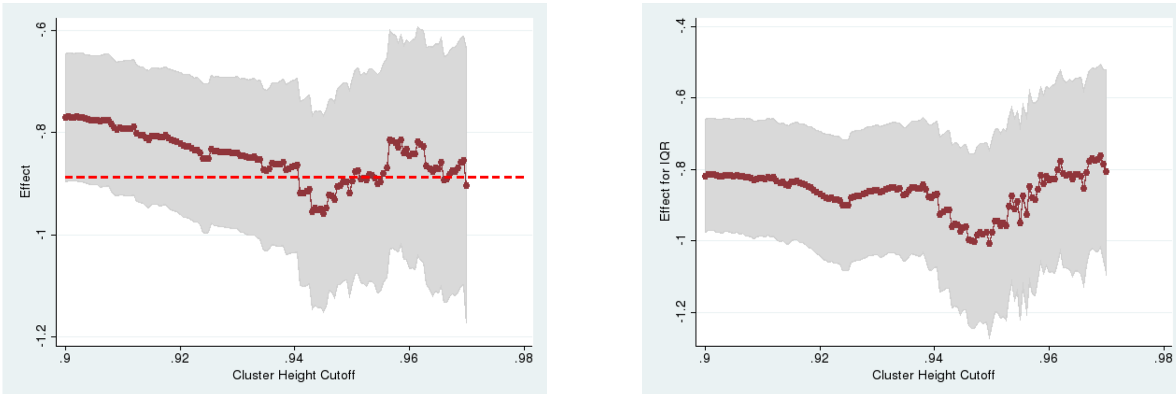
To give an empirical application, Figure 5 shows the distribution of t-statistics obtained from estimating 3. The blue vertical line is the original t-statistic, and the light gray vertical lines are the 2.5th and 97.5th percentiles. Clearly, in this application the result is still significant, because the entire confidence interval of t-statistics is less than the critical value (-1.96). Once again, the t-statistic from the original estimate is one of the smallest in magnitude.

This exercise demonstrates that there is additional uncertainty induced by the construction of the commuting zones that is not addressed in empirical estimates that use these definitions, which may overstate the precision of the results.

4.3 Sensitivity to Chosen Cutoff

In addition to the uncertainty that is induced by underlying errors in the commuting flows, in Section 3.2 we showed that the decision of where to stop the clustering process was rather

Figure 6: Differences in Effect Based on Cluster Cutoff



(a) Effect by Cutoff Height (b) Effect by Cutoff Height, Scaled by IQR

Note: Author's calculations based on replication of Tolbert and Sizer's method. Panel (a) shows estimates of β_1 from equation 3 for different definitions of commuting zones based on height cutoff, while Panel (b) shows estimates of β_1 scaled by the difference in exposure between the 25th and 75th percentile commuting zone. The horizontal line in panel (a) is the main estimate from Autor, Dorn and Hanson (2013)

arbitrary, since there is no clear guidance on what cutoff is most appropriate. To demonstrate how the cutoff choice affects estimates of β_1 from Equation 3, we generate clusters based on cutoffs between 0.9 and 0.97 and estimate the equation using the resulting clusters.

Figure 6 displays the results of this exercise, where panel (a) shows the raw coefficient and panel (b) shows the coefficient scaled by the interquartile range of ΔIPW_{uit}, since the IQR changes based on the composition of the clusters. In panel (a), the red horizontal line is the estimate from ADH.

Again, our results show that there is some variation in the estimate based on the cutoff value. Cutoff values marginally higher or lower can give different results based on how many clusters form at certain points in the cutoff distribution. Given the sensitivity of estimates to the chosen cutoff, best practices for a researcher would be to report estimates for a broad range of possible cutoffs.

4.4 Advice to Researchers

From the results above, it is clear that current commuting zone definitions understate the uncertainty of zone assignment, which has implications for empirical results. Importantly, this uncertainty manifests itself on two different margins: uncertainty about zone assignment due to errors in the flows, and uncertainty about zone assignment due to the chosen cutoff.

Given this uncertainty, we have two pieces of advice for researchers using commuting zones. First, we suggest re-estimating results using multiple realizations of commuting zones, which incorporates the additional uncertainty because of the underlying error in the measurement of flows. Researchers can validate results by examining either the distribution of β or the distribution of the t-statistics, as described in the previous sub-sections. Second, we suggest displaying results for a variety of different cluster counts resulting from a range of cutoff values. This second point is particularly important for researchers applying the methodology from TS1996 to new datasets or for characterizing labor markets outside the United States, given that cluster counts are somewhat subjective and that results can differ considerably based on the count.

To aid researchers in this effort, we provide datasets and code online that include a crosswalk from county to all the realizations of commuting zones used in this paper to characterize uncertainty in inputs as as well as different cutoff values.[16] We also provide our sample code that produced Figures 4 and 6.

5 Conclusion

Numerous influential papers in labor economics have used commuting zones as an alternative definition to local labor markets. However, researchers typically do not evaluate how the methodology used to construct commuting zones may impact their findings, nor have there been any evaluations of the sensitivity of commuting zones to design feature more generally. Our paper contributes to this literature by analyzing this methodology and its implications for empirical applications.

We document that the commuting zone methodology is sensitive to uncertainty in the

[16]Our code is available at https://github.com/larsvilhuber/MobZ/, see also Foote, Kutzbach and Vilhuber (2017).

input data and parameter choices and we demonstrate how these features affect the resulting labor market definitions. Furthermore, we demonstrate that uncertainty in local labor market definitions also affects empirical estimates that use commuting zones as a unit of analysis.

Future work may explore other clustering methods, which are less history-dependent, as they may be better suited for considering a wide range of cluster counts and for evaluating the optimality of cluster counts. Developing metrics to compare zone configurations against one another will facilitate comparisons of the overlap of different clustering outcomes. A more complete characterization of measurement error in the flow measures, reflecting the sparse nature of survey responses, may improve the economic interpretation of flows in rural areas and for long distance flows. Additional metrics of local labor market integration may help to evaluate the overall validity of various definitions.

References

Amior, Michael, and Alan Manning. 2015. "The Persistence of Local Joblessness." Centre for Economic Performance, LSE CEP Discussion Papers dp1357.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6): 2121–68.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*, 23(1): 1–76.

Bound, John, and Harry J Holzer. 2000. "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." *Journal of Labor Economics*, 18(1): 20–54.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics*, 19(1): 22–64.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics*, 129(4): 1553–1623.

Diamond, Rebecca. forthcoming. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." *American Economic Review*.

Foote, Andrew, Mark Kutzbach, and Lars Vilhuber. 2017. "MobZ - Replication archive for a re-examinination of Local Labor Market Definitions. https://doi.org/10.5281/zenodo.569588."

Foote, Andrew, Michel Grosz, and Ann Stevens. 2015. "Locate Your Nearest Exit: Mass Layoffs and Local Labor Market Response." Center for Economic Studies, U.S. Census Bureau Working Papers 15-25.

Fowler, Christopher S., Danielle C. Rhubart, and Leif Jensen. 2016. "Reassessing and Revising Commuting Zones for 2010: History, Assessment, and Updates for U.S. 'Labor-Sheds' 1990–2010." *Population Research and Policy Review*, 35: 263–286.

Halkidi, Maria, Yannis Batistakis, and Michalis Vazirgiannis. 2001. "On Clustering Validation Techniques." *Journal of Intelligent Information Systems*, 17(2): 107–145.

Head, Allen, and Huw Lloyd-Ellis. 2012. "Housing Liquidity, Mobility, and the Labour Market." *Review of Economic Studies*, 79(4): 1559–1589.

Kennan, John, and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica*, 79(1): 211–251.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2015. "Commuting, Migration and Local Employment Elasticities." National Bureau of Economic Research, Inc NBER Working Papers 21706.

Moretti, Enrico. 2011. “Local Labor Markets.” , ed. O. Ashenfelter and D. Card Vol. 4 of *Handbook of Labor Economics*, Chapter 14, 1237–1313. Elsevier.

Notowidigdo, Matthew J. 2011. “The Incidence of Local Labor Demand Shocks.” National Bureau of Economic Research, Inc NBER Working Papers 17167.

Office of Management and Budget. 2000. “Standards for Dileneating Metropolitan and Statistical Areas; Notice.” Federal Register Notices 249.

Office of Management and Budget. 2010. “2010 Standards for Dileneating Metropolitan and Statistical Areas; Notice.” Federal Register Notices 123.

Office of Management and Budget. 2013. “Revised delineations of metropolitan statistical areas, micropolitan statistical areas, and combined statistical areas, and guidance on uses of the delineations of these areas.” Executive Office of the President OMB Bulletin No. 13-01.

Restrepo, Pascual. 2015. “Skill Mismatch and Structural Unemployment.” Massachusetts Institute of Technology Job Market Paper.

Roback, Jennifer. 1982. “Wages, Rents, and the Quality of Life.” *Journal of Political Economy*, 90(6): 1257–78.

Schmutz, Benoit, and Modibo Sidibé. 2014. “Job Search and Migration in a System of Cities.” Centre de Recherche en Economie et Statistique Working Papers 2014-43.

Tolbert, Charles M, and Molly Sizer. 1996. “US commuting zones and labor market areas: A 1990 update.” Economic Research Service.

Tolbert, Charles M, and Molly Sizer Killian. 1987. “Labor market areas for the United States.” *ERS staff report.*

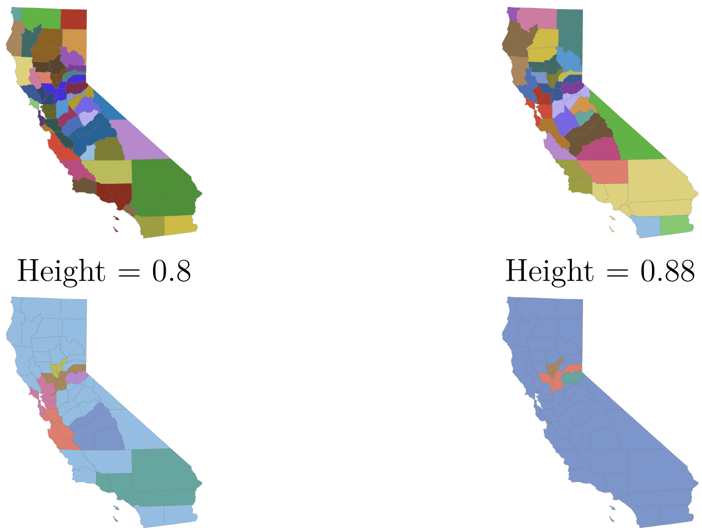
U.S. Census Bureau. 1992. “Census of Population 1990: Number of Workers by County of Residence by County of Work.” Summary Tape File S-5.

Wozniak, Abigail. 2010. “Are College Graduates More Responsive to Distant Labor Market Opportunities?” *Journal of Human Resources*, 45(4): 944–970.

Yagan, Danny. 2016. “The Enduring Employment Impact of Your Great Recession Location.” University of California, Berkeley Working Paper.

Tables and Figures Appendix

Figure A1: Various Height Cutoffs for California



Height = 0.8 | Height = 0.88

Height = 0.96 | Height = 1

Notes: The above graphs are generated using the methodology outlined in Section 2, using 1990 Census JTW data. More detail is in the text.

Table A1: Summary Statistics of Ratio of MOE to Flows

	Mean	25th Pctile	50th Pctile	75th Pctile
All counties	1.236	0.845	1.370	1.600
Flows <100	1.432	1.148	1.500	1.636
Flows 100-1000	0.444	0.301	0.414	0.549
Flows 1000-10000	0.131	0.087	0.124	0.169
Flows 10000+	0.037	0.024	0.036	0.049

Notes: Author's calculation using 2009-2013 ACS Journey-to-Work data.