UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays in Applied Microeconomics

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Economics

by

Arman Khachiyan

Committee in charge:

Professor Gordon Dahl, Chair
Professor Judson Boomhower
Professor James Fowler
Professor Alexander Gelber
Professor Gordon Hanson

2022

The dissertation of Arman Khachiyan is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The work in this dissertation was guided by a fantastic committee of advisors and supported by my devoted family and friends. Gordon Dahl has played a large role in shaping my identity as an economist by helping me develop my intuition for identifying causal relationships in observational data. He has also been a steadfast and enthusiastic supporter of my efforts to implement these methods in applications on diverse economic topics and unconventional data sources. His efforts to understand and help advance my goals, and his appreciation of my need for independence, have made him an ideal committee chair for me.

Gordon Hanson has served as an early and constant advisor and colleague throughout my PhD. He granted me early exposure to economic work in remote sensing and patiently supported my development of the skills and knowledge needed to work with these specialized data. Through joint research ventures over several years, he has thoughtfully empowered and consistently prioritized my growth and progress as a student. Finally, his depth of expertise in the topics of urban and labor economics have been invaluable in effectively framing and communicating my work.

Judd Boomhower has been an outstanding resource for connecting my work on fracking to the field of environmental economics. He has the special ability to give comments which identify gaps in the core of an argument while building up constructive ideas for filling these gaps. These suggestions were immensely helpful in the preparation and communication of my job market paper. James Fowler has been a consistent and enthusiastic champion of my interests in connecting topics and methods across disciplines. This is exemplified by his efforts to advance the Omni Methods Group at UCSD, which has served as an important network and presentation venue for me. Finally Alex Gelber has helped me draw strong ties between my research and active areas of policy. His depth of knowledge and experience in this arena has been critical in grounding my work and motivating it to a broad audience.

My family has been a foundation for me through the many difficult phases of my PhD. Most directly, my wife has been immensely helpful in giving me perspective on daily challenges. Her ability to appreciate the nuances of my tasks while reminding me of my broad goals and values

is uniquely comforting. My parents have always encouraged me to do what makes me happy, and provided me with the resources and work ethic to set me up for success. My mother's sense of humor, and her daily efforts to help out in any way she can has been immensely helpful for maintaining a positive and productive mindset. I thank my brother for sharing countless trips, activities, and common interests with me. His enthusiasm for seeking out adventure and maximizing every opportunity is abundant and infectious.

I am grateful for the outstanding network of friends and colleagues I have been surrounded with at UCSD. Among the many fantastic co-authors I had the privilege of working with on the first chapter of my dissertation, Anthony Thomas has been an impeccable friend and collaborator over the years. His deep understanding of the complementarities between computer science and economics, and of the many species of waterfowl in North America, have made him an irreplaceable component of my growth over the last decade. I am also deeply appreciative of Huye Zhou, whose expertise and innovations in implementing neural networks at scale made our joint project possible. Many other students, postdocs, and faculty at UCSD have played an important role in my development as an academic and individual, notably Luke Sanford, Richard Bluhm, Gordon McCord, Ran Goldblatt, Jake Orchard, Zack Goodman, Jackson Somers, Camila Navajas-Ahumada, Alyssa Brown, Julie Cullen, Marc Meundler, Mark Jacobsen, David Arnold, Julian Betts, Molly Roberts, Julie Cullen, Kate Antonovics, Sally Sadoff, Jeff Clemens, Eli Berman, and Joel Sobel.

Chapter 1 has been accepted for publication at the American Economic Review: Insights. Khachiyan, A.; Thomas, A.; Zhou, H.; Hanson, G.; Cloninger, A.; Rosing, T.; Khandelwal, A. The dissertation author was the primary investigator and author of this material. This project was funded through the generous support of the Russell Sage Foundation program on Computational Social Science.

Chapter 2 is currently being prepared for submission for publication of the material. Khachiyan, A. The dissertation author was the primary investigator and author of this material.

Chapter 3 is currently being prepared for submission for publication of the material. Khachiyan, A. The dissertation author was the primary investigator and author of this material.

VITA

2013	Bachelor of Arts in Economics, University of California Berkeley
2022	Doctor of Philosophy in Economics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Essays in Applied Microeconomics

by

Arman Khachiyan

Doctor of Philosophy in Economics

University of California San Diego, 2022

Professor Gordon Dahl, Chair

This dissertation contains three essays studying topics in applied microeconomics. The first chapter is a co-authored paper in which we use daytime satellite imagery and convolutional neural networks to model economic growth at the neighborhood level. In the second chapter, I use this model to examine the spatial distribution of residential impacts from fracking. The third chapter investigates methods of measuring skill distance between occupations and proposes a new method which matches patterns of observed occupational transition. Each chapter uses unconventional data sources and machine learning techniques to contribute to central questions in labor economics

research and policy.

In the first chapter we apply deep learning to daytime satellite imagery to predict changes in income and population at high spatial resolution in US data. Our model predictions achieve R^2 values of and 0.32 to 0.46 in decadal changes, which have no counterpart in the literature and are 3-4 times larger than for commonly used nighttime lights. Our network has wide application for analyzing localized economic shocks.

One such application is my second chapter, which studies changes in total neighborhood income and population in areas near fracking extraction and shale reserves. My microspatial approach identifies that fracking exposure as far as 20 miles away leads to a 2 percent decline in neighborhood income. The spatial gradient and associated mechanisms of this effect indicate that it is driven by local industrialization rather than direct environmental externalities. Examination reveals margins of policy and labor conditions which attenuate the observed impacts.

In the third chapter I show that a regression framework generates a novel, empirical occupational skill distance norm which is disciplined by observed occupation switching patterns. This approach relieves key limitations of existing measures such as linearity and symmetry. It also allows for an analysis of which skill dimensions relate to the portability of human capital, and which do not. Implications for existing results on skill portability are discussed, along with immediate policy applications on employee adjustment costs.

Chapter 1

Using Neural Networks to Predict Micro-Spatial Economic Growth

Arman Khachiyan, Anthony Thomas, Huye Zhou, Gordon Hanson, Alex Cloninger,
Tajana Rosing, Amit K Khandelwal

1.1 Abstract

We apply deep learning to daytime satellite imagery to predict changes in income and population at high spatial resolution in US data. For grid cells with lateral dimensions of 1.2km and 2.4km (where the average US county has dimension of 51.9km), our model predictions achieve R^2 values of 0.85 to 0.91 in levels, which far exceed the accuracy of existing models, and 0.32 to 0.46 in decadal changes, which have no counterpart in the literature and are 3-4 times larger than for commonly used nighttime lights. Our network has wide application for analyzing localized shocks.

1.2 Introduction

Spatial economic analysis evaluates how localized shocks–e.g., infrastructure projects (Redding and Turner, 2015), factory openings (Greenstone et al., 2010), and natural disasters (Boustan et al., 2020)–affect the geographic distribution of economic activity. Standard approaches match administrative or survey data to the geospatial structure of these shocks. Because data tend to be released infrequently (e.g., decadally for Censuses) and for relatively coarse spatial units (e.g., counties or metro areas), this method is suitable for assessing long-run economic impacts at a broad spatial scale (e.g., Faber, 2014; Baum-Snow et al., 2017). By contrast, assessing the impact of shocks at the neighborhood level across all cities nationally would be infeasible with conventional data in most countries.

Satellite imagery offer a path forward. Recent work leverages nighttime light intensity to study regional economies where conventional data are sparse (see, e.g., Donaldson and Storeygard, 2016). Although nightlights can detect changes in economic activity across cities, states, and countries, they are problematic at smaller spatial scales. High luminosity in city centers may saturate satellite sensors, leading to top coding, while surface reflectance may cause light to bleed across space, making urban footprints appear artificially large. Aggregating imagery addresses these problems, but dampens spatial variation. To increase granularity, recent work in remote sensing and computer science uses convolutional neural networks (CNNs) to predict outcomes from multispectral daytime satellite imagery at high spatial resolutions. This research detects cross-sectional variation in spending and wealth for villages in Africa (Jean et al., 2016) and poverty rates across a diverse sample of cities (Babenko et al., 2017; Piaggesi et al., 2019). In related work on $1km$ grid cells in the US, Rolf et al. (2021) develop a "task-agnostic" learning approach to predict a broad set of localized outcomes.

This paper makes two advances over the existing literature. First, we implement a CNN to predict changes in local economic activity from changes in high-resolution daytime satellite imagery. We achieve high predictive accuracy in the *cross section*, as others have done, and in

predicting localized outcomes in the *time series*, which has not been the focus of previous work. Second, we demonstrate that our approach far outperforms nighttime lights at predicting changes at fine spatial scales.[1]

For inputs in model training, we use multi-spectral imagery from Landsat; for labels, we use household income and population for Census Blocks in the US Census and American Communities Surveys (ACS). Working in the data-rich US setting, we are able to train a CNN from scratch using hundreds of thousands of images and training labels. Matching Census data with Landsat to construct square images with side lengths of $1.2km$ or $2.4km$, we predict levels and changes in income and population.[2] In the test set, model predictions achieve R^2 values of greater than 0.85 in levels and 0.32 in time differences, which compare to R^2 values for predictions in levels of 0.42 for income and 0.75 for population in Rolf et al. (2021). There are no estimates in the literature to benchmark our predictions of changes in local income and population.

Methodologically, we advance the scale and specificity at which machine learning is used to predict local changes in economic activity. Rather than beginning with image features generated by existing models for prediction–which is the standard practice of transfer learning–we train and tune CNN models for all urbanized pixels in the contiguous US from the ground up. This computationally demanding approach allows us to detect the low-level image features (i.e., shapes, shades, edges, clusters) that are informative for predicting income and population, beyond those that have proven useful in other image tasks (Rosenstein et al., 2005).

Our approach complements Rolf et al. (2021), who aim for generality rather than specificity in predicting outcomes from satellite imagery. They use a layer of randomly initialized filters–based on sampling a small patch from the imagery–to extract features from the raw images. These features are then used to predict outcomes of interest. Their process requires little training, is undemanding computationally, and is suitable to predicting many outcomes, but may not

[1] Given their wide use in spatial analysis, nightlights are a natural benchmark for comparison. See, e.g., Chen and Nordhaus (2011), Henderson et al. (2012), Gennaioli et al. (2013), Michalopoulos and Papaioannou (2014), Storeygard (2016), Bruederle and Hodler (2018), Henderson et al. (2018), Hjort and Poulsen (2019), and Jedwab and Storeygard (2021). In the policy domain, the World Bank has produced a quarterly data set, Light Every Night, which records localized nighttime light intensity from 1992 to 2020.

[2] For comparison, in 2010 US Census Blocks had an average size of $0.9km$ x $0.9km$.

be well tuned to specific prediction tasks. Our approach, while highly intensive in training and computation, is bespoke for predicting local changes in income and population.

Our model and code can be used to impute high-frequency outcomes in between the periodic data drawn from large-scale surveys, to train models with imagery where Census data exist but are sparse, and to predict levels and changes in income and population for spatially disaggregated units where Census data are unavailable entirely.[3] We conclude with a discussion of potential applications.

1.3 Data & Methods

1.3.1 Imagery and Label Data

For satellite imagery, we use daytime surface reflectance detected by the USGS Landsat 7 satellite, which has 7 spectral bands (3 visible, 2 near-infrared, 1 thermal, 1 mid-infrared), covers the Earth's surface biweekly, and has a spatial resolution of 30m (USGS, 2020). Using Google Earth Engine (Gorelick et al., 2017), we construct annual composites of surface reflectance for the May-August median of cloud-free images each year.[4]

To avoid populating the data with a large number of images covering uninhabited areas, we limit the sample to Landsat pixels corresponding to urbanized US Census Block Groups.[5] We first rank Block Groups according to population density in 2000 and identify those in descending rank order that collectively comprised 85% of the continental US population in that year. We then draw a 1-mile buffer around these Block Groups and include all images within the buffer in our sample. Following this procedure, our data cover 93% of the continental US population in 2000. We construct individual images from Landsat imagery as squares. We test two image sizes, one

[3] Our code, model, and output are available at https://github.com/thomas9t/spatial-econ-cnn.git. This repository includes scripts and computed weights which can be used to augment or extend our modeling approach. It also includes data and instructions for direct applications using our generated income and population measures.

[4] Using summer months averts irregularities due to persistent clouds or snow.

[5] Census Blocks (600 to 3,000 residents) are the smallest geographic unit in the Census; Block Groups are the next smallest unit. In 2000, there were 211,267 Block Groups, with a mean of 39 Blocks per Group. We exclude Census Blocks in which more than 10% of the population was living in group quarters in 2000.

with 2.4*km* sides and one with 1.2*km* sides (see Figure 1.1).[6] Smaller images, which increase the spatial resolution of the ultimate predictions, may be more useful in some applications, but may also be more challenging to model as they have fewer pixels, and therefore less information available, per image.

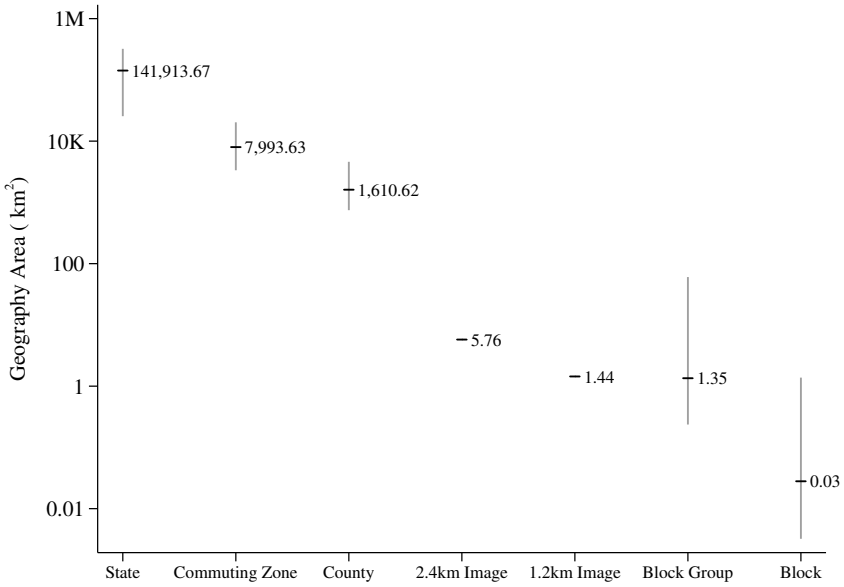


Figure 1.1: Geographic Area of Census and Image Units
Note: This figure shows the geographic area covered by various Census geographic units alongside our constructed images. Horizontal black dashes display the median area for each geographic unit; grey vertical lines show the range from the 10th percentile of area to the 90th percentile of area for each geography. Note that the y-axis is a log-scale of area.

Labels for the analysis are constructed from the US Census for 2000, 2010, and 2020, and the ACS five-year samples for 2005-2009, 2008-2012, and 2015-2019, all extracted from Manson et al. (2020). From each sample, we use population by Census Block and total personal income, for residents ages 15 years and older, by Census Block Group.[7] Because income data are only published at the Block Group level, we interpolate income from Block Groups to Blocks according to the population distribution across Blocks within Groups.[8] We further interpolate income and

[6]The 2.4*km* and 1.2*km* images have pixel dimensions of 80x80 (6,400 pixels) and 40x40 (1,600 pixels).

[7]Personal income includes wages and salaries, tips and bonuses, proprietor's income, government cash transfers, interest and rental income, and retirement benefits. In-kind government transfers, capital gains, and revenue from property sales are not included (Manson et al., 2020). All values are in 2012 dollars.

[8]Because Block population is unavailable in the ACS data, we use the 2010 population to interpolate 2007 income

population from Census Blocks to images based on the geographic overlap between the two.

1.3.2 Convolutional Neural Networks for Spatial Economic Analysis

Although images are an information-rich medium, their unstructured and high-dimensional nature make them difficult to use with conventional learning algorithms, such as LASSO regression. The ability of CNNs to learn structure from data has revolutionized image processing (LeCun et al., 2015). A CNN consists of a sequence of layers, each of which implements a parameterized nonlinear transformation of its inputs. The inputs to the first layer are raw images, in our case 7-dimensional images from Landsat. The output of the first layer is used as input by the second layer and so on. The transformation implemented by each layer is typically either a convolution or pooling operation (Goodfellow et al., 2016), which can be visualized by sliding a rectangular window (e.g., $3 \times 3 \times 7$) over the input image. At each position, an inner product is performed, which aggregates the pixel values in the window into a single number. The output of either a convolution or a pooling operation is another image in which the pixels are these aggregated values.[9] After a sequence of convolutional and pooling layers, the transformed image passes through a fully-connected layer, which is a nonlinear regression that maps the image features extracted by the convolutional and pooling layers to a predicted outcome. The parameters of the model are fit using a gradient-based optimization algorithm known as stochastic gradient descent, which minimizes the MSE over labeled training examples.

In our context, a CNN extracts economic information that is latent in spectral data. Asphalt, cement, gravel, soil, water, vegetation, and other materials vary in their reflectance intensity across the light spectrum (e.g., De Fries et al., 1998). The presence of these materials varies enormously

from Block Groups to Blocks, and similarly use 2020 population to interpolate 2017 income.

[9] In a convolutional layer, the window contains coefficients used to compute a weighted sum of the pixel values within each window via convolutional filtering. The CNN learns these weights to identify a feature of the image. By applying a sequence of transformations that learn features at increasingly coarse spatial scale, CNNs are able to represent complex spatial relationships between pixels in an image. In a pooling layer, we condense all pixel values within the window to a single number—typically the maximum pixel value within the window. Pooling differs from convolution primarily in that it does not require any learned weights. Pooling serves to reduce the size of the image, which lowers the computational burden of subsequent layers, and helps make the features detected by convolutions robust to small spatial transformations.

within an urban area: more vegetation and loose soil in green spaces; more asphalt and cement around motorways; more steel and wood, together with concrete, in houses and buildings (Zha et al., 2003). The shapes of these materials exhibit similarly wide variation: irregular edges in green spaces, intermittent grids of grass and roofing material in suburbs, larger rectangular clusters in apartment complexes and shopping malls, and compact, interconnected grids in urban centers (Ural et al., 2011, Pesaresi et al., 2016). It is this complexity that makes a neural network powerful—the network learns the mapping of materials and shapes to the level of economic activity and changes in materials and shapes to changes in economic activity. As an empirical regularity, the features learned by the network are often organized into a hierarchy of complexity (Zeiler and Fergus, 2014), in which early layers learn to identify simple features, such as edges or basic shapes, and subsequent layers learn to compose these simple features into complex objects, such as office buildings, industrial parks, suburban developments.

The predicted values that our analysis generates will be subject to error. In regression analysis, measurement error in the outcome variable does not generate bias in estimating treatment effects if this error is uncorrelated with the treatment being studied.[10] Because treatments may be correlated with initial levels of economic development, we wish to eliminate any correlation between prediction errors and initial conditions. To do so, we include controls for local economic characteristics in the initial time period (as measured in Census data) in our CNN models.[11] An added virtue of this approach is that it may improve model accuracy, thereby reducing the scope for prediction errors to contaminate analysis that uses our predictions as outcome variables in the first place. Implementing our approach, we find minimal correlations between prediction errors and initial conditions in our data.[12]

[10]For example, if the assigned treatment (a new highway) had a strong positive correlation with the measurement error in the outcome (larger positive deviations between actual and predicted population or income near the highway), this would lead to an overestimate of the true treatment effect.

[11]A full list of variables included can be found in Table 1.3.

[12]The largest correlation coefficient for the income differences model in the test set is 0.057 (for employment in hospitality services), and the median correlation is 0.002. See Table 1.3 for details.

1.3.3 Training, Tuning, and Testing Procedure

CNNs contain a large number of tunable parameters—known as hyperparameters—which control the model architecture and optimization process (e.g., the dimension of convolution filters, number of channels produced by each convolution layer, strength of regularization on weights, and step size used by the optimization algorithm). CNNs are prone to overfitting, in which a model generates accurate predictions on the data used to fit parameters, but fails to generalize on out-of-sample data. To obtain accurate estimates of the model's out-of-sample performance and to determine the best values for hyperparameters, we follow standard practice in empirical machine learning by partitioning our data into three disjoint subsets for training, validation, and testing (Friedman et al., 2001). The training set is used to fit model parameters, and the validation set is used to estimate the out-of-sample error for a given set of hyperparameters. The final model is obtained by selecting the hyperparameters that yield the lowest prediction error in the validation set. The test set is used to obtain an estimate of out-of-sample error for the final model. Ideally, we would repeat this partitioning many times to obtain an estimate of the distribution of out-of-sample error. However, this is infeasible at our data scale.

Models are trained to minimize the MSE of the prediction using the Adam optimizer (Kingma and Ba, 2014). When training models in levels, we pool training data for the years 2000 and 2010, and train a single model to predict outcomes in this combined sample. An alternative approach would be to specialize models in levels to a particular year. However, this method led to greater over-fitting, where training on pooled data resulted in only modest losses in accuracy. We tune hyperparameters for the learning rate (step size and decay rate) and strength of L2-regularization on weights. The training images are randomly augmented to prevent overfitting (cropping, flipping and zooming). We stop the optimization process after 200 epochs or if the R^2 on the validation set fails to increase for 50 epochs. In the latter case we retain the weights which maximize the validation R^2. Further details are in the online appendix.

To obtain reliable estimates of out-of-sample performance, the training, validation, and test

sets must be disjoint. To construct these subsets, we partition the full set of images meeting our inclusion criteria into contiguous urban areas. We randomize selection into training, validation, and test sets at the level of the urban area, rather than the level of the image. Maintaining a disjoint split of the images removes the possibility of data leakage between the training and testing sets (which may result if we allowed images from the two sets to be adjoining). This procedure leads to a total of 4,710 urban regions, which are each randomly assigned to either the train (roughly 50%), validation (roughly 20%), or test (roughly 30%) sets. An image receives the subset designation of the urban region it is contained by, where we discard images located on borders between urban areas (e.g., images on the border between Minneapolis and St. Paul, which are separate urban areas). Figure 1.6 shows the distribution of images into each of these sub-groups.

1.4 Results

1.4.1 CNN Model Performance

Baseline Results

Here, we present our main results on the predictive power of CNNs. Table 1.1 Panel A reports R^2 values for model accuracy, again in levels (2000 and 2010) and time differences (2000 to 2010) for $2.4km$ images; Table 1.1 Panel B repeats the results for $1.2km$ images. Our smaller images are close in dimension to the $1km$ images that Piaggesi et al. (2019) and Rolf et al. (2021) use in their machine-learning approaches to model, respectively, poverty levels and levels of average income and population density in US data. We report performance in the training, validation, and test sets, with and without incorporating initial conditions in model training.[13] For models in levels, we report results for a single model trained to predict both years; performance in each year

[13]The complete set of initial conditions, all measured for the year 2000, are at the county level, log population, log personal income, and the shares of employment in business services, non-business services, and industrial production; and at the Census Block level, population shares for individuals who are female, ages 25 to 54, Black, non-Hispanic white, Hispanic, and living in group quarters, and employment shares for two-digit manufacturing industries, business services, and non-business services (Census, 2020).

separately is very similar (see Table 1.7).

Beginning with larger images in Table 1.1 Panel A, we first consider model performance for outcomes in levels. For income and population, and with initial conditions, the R^2 in the test set are 0.90 and 0.91, respectively. Without initial conditions, performance deteriorates moderately, with the R^2 falling by 0.05 to 0.07. Comparing these results to those for smaller image sizes in Table 1.1 Panel B, the R^2 for income and population are 0.85 and 0.86, with initial conditions, and 0.09 to 0.11 lower, without them. The weaker performance of smaller relative to larger images is expected. For smaller images, the network must form predictions based on a smaller number of underlying pixels, which tends to undermine accuracy.

Table 1.1: R^2 Values for Baseline Models of Large and Small Images

	2000 and 2010 Levels			2000 to 2010 Difference		
	Train	Valid	Test	Train	Valid	Test
Panel A: National 2.4*km* Imagery						
Income						
With Initial Conditions	0.9254	0.8934	0.9018	0.4863	0.4126	0.3962
Without Initial Conditions	0.8625	0.8289	0.8374	0.4951	0.3960	0.3702
Population						
With Initial Conditions	0.9611	0.9029	0.9132	0.5410	0.4839	0.4573
Without Initial Conditions	0.9187	0.8636	0.8684	0.7004	0.4496	0.4202
Panel B: National 1.2*km* Imagery						
Income						
With Initial Conditions	0.8957	0.8620	0.8543	0.3819	0.3061	0.3216
Without Initial Conditions	0.7969	0.7597	0.7494	0.2959	0.2609	0.2690
Population						
With Initial Conditions	0.9101	0.8716	0.8600	0.4217	0.3401	0.3559
Without Initial Conditions	0.7841	0.7612	0.7492	0.3924	0.3051	0.3036

Note: The table shows R^2 values computed on each subset of the images with 2.4*km* and 1.2*km* sides. The total sample size of spatially unique images in training, validation and test subsets is 112,932 for larger images and 320,880 for smaller images. Income measures the log of total personal income, while population is the log of total population. 2000 and 2010 levels represent a model predicting levels for images in the two years together, while the differences columns show the result predicting the change from 2000 to 2010. Initial conditions included in the model are gender and racial composition, employment shares and county level population and income, all measured in 2000. The results show high accuracy in predicting both levels and differences in income and population; there is not strong evidence of over-fitting in the training set. Model fit is consistently lower on the sample of smaller images; hence, we prioritize the sample of 2.4*km* imagery as our baseline analysis sample.

Turning to our predictions for changes over 2000-2012, for $2.4km$ images the R^2 for income and population growth rates in the test set are 0.40 and 0.46, respectively, with initial conditions, and 0.37 to 0.42 without them. For $1.2km$ images, model performance is again somewhat weaker. The R^2 is 0.32 to 0.36, with initial conditions, and 0.27 and 0.30, without them.

Comparing our results for $1.2km$ images to those for $1km$ grid cells in Rolf et al. (2021), we achieve higher performance for both population density (our R^2 of 0.86 versus theirs of 0.72) and income (our R^2 of 0.85 versus theirs of 0.42). We note that whereas our model is trained from scratch for the express purpose of predicting income and population, their model is constructed for the general purpose of predicting many possible outcomes and therefore may sacrifice accuracy for any specific quantity. Because we are unaware of any prior work that uses CNNs to predict changes in income or population at spatial resolutions similar to our image sizes, we have no benchmark for comparison in the literature for these results.[14]

To evaluate overfitting, we compare predictive accuracy across training, validation, and test sets. Focusing on the time-difference models and on results in validation versus training sets, the R^2 for income growth in $2.4km$ images falls minimally by 0.02 from the validation to the test set, with initial conditions, and by 0.03, without initial conditions; the change in R^2 is slightly larger for population growth. For $1.2km$ images, the R^2 either rises or changes minimally from the validation to the test set, both for income and population and with or without initial conditions. With cross-validation, overfitting in our model training does not appear to be manifest.

Model Prediction Errors

To evaluate prediction errors in our model, Figure 1.2 shows scatter plots of model-predicted values and actual values for log income and population, in levels and time differences. In the models for levels, the data are tightly packed around the 45-degree line, indicating that the model

[14] In Table 1.4, we report results for log income per capita. In levels for 2000 and 2010 and with initial conditions, we achieve R^2 in the test set of 0.65 for $2.4km$ imagery and 0.61 for $1.2km$ imagery; in changes for 2000-2010 and with initial conditions, we achieve R^2 in the test set of 0.07 for both $2.4km$ and $1.2km$ imagery. Differencing population from income, which removes much of the systematic variation in economic activity from the data, appears to complicate extracting information from satellite imagery.

accurately captures log income and population across the entire distributions of each. The results for growth rates in the second row show that the prediction of differences is more challenging. The model captures much of the variation for images in which values are growing, but tends to over-predict growth in images for which values are flat or declining, especially for income. The asymmetry in errors for positive and negative growth rates—for income, in particular—may be a result of the slow depreciation of physical capital. Whereas in expanding regions income growth may lead directly to new construction, in declining regions income loss may result in the change or removal of structures over longer time horizons.

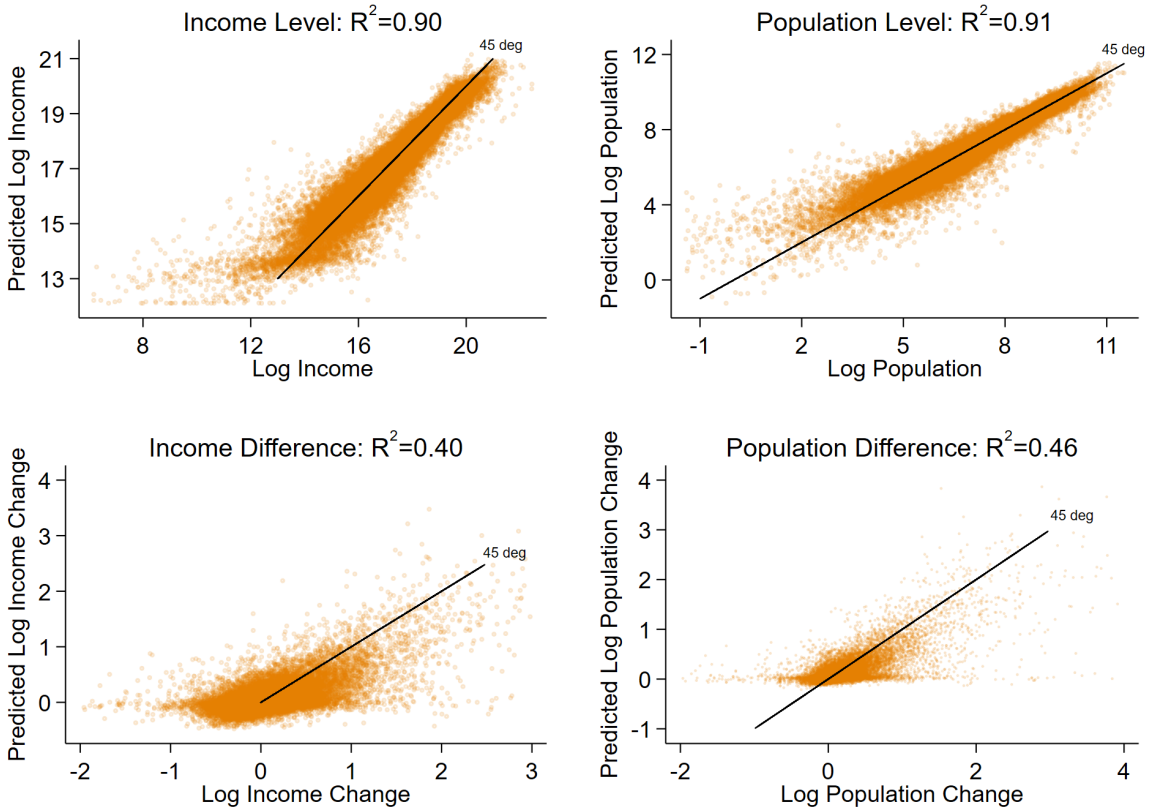


Figure 1.2: Model Predictions against Actual Values

Note: Levels models include data from both 2000 and 2010. Extreme outliers are omitted from this figure to allow visualization of the central tendency in the data.

To see whether our prediction errors are associated with initial economic conditions, we compute the correlation of our prediction errors with initial industry employment shares and de-

mographic characteristics. These correlations are all below 0.1 and mostly well below 0.02, as seen in Table 1.3. Estimating a regression of prediction errors on fixed effects for each urban area in the sample, the fixed effects absorb 11% or less of the variation in the errors, as seen in the last row of Table 1.3. Figures 1.7 and 1.8 further show no systematic variation in prediction accuracy across geographic regions. In all, there appears to be little covariation between prediction errors and initial economic conditions in our sample.[15]

1.4.2 Comparison with Nightlight Intensity

Given the growing use of nightlights to detect GDP, as discussed above, we next compare our CNN performance to how well nightlights predict levels and changes in economic activity. In Figure 1.1, we regress log income or log population on log nightlight intensity (NOAA, 2020), first in levels for the years 2000 and 2010 pooled in a single regression, and then in changes over the 2000 to 2010 time period. The geographies studied range from US states to Census Blocks and include our $1.2km$ and $2.4km$ images. To normalize the size of spatial units, we express all values per km^2.

Figure 3 summarizes the results by presenting the R^2 values for each OLS regression. In the regressions in levels for larger geographies, nightlights are a strong predictor of economic activity, consistent with previous research (Gennaioli et al., 2013; Donaldson and Storeygard, 2016). For income levels in 2000 and 2010, where results for population are very similar, R^2 levels are stable across larger spatial units, at 0.67 for states, 0.66 for commuting zones, and 0.71 for counties. Jumping from counties to our $2.4km$ images, the R^2 drops to 0.57 and drops further to 0.50 for our $1.2km$ images. Even at roughly the neighborhood level—the $1.2km$ images—nightlights are strongly positively correlated with the level of economic activity.

Yet, our CNN trained on daylight imagery substantially outperforms nightlights in cross-section data. Referring to our baseline CNN results in Table 1.1, the CNN trained on daylight satellite imagery with initial conditions yields an R^2 for log income that is 0.33 higher for $2.4km$

[15] In the online appendix, we follow recent literature on interpreting neural network predictions by evaluating saliency maps, which indicate which pixels in an image most influence network prediction.

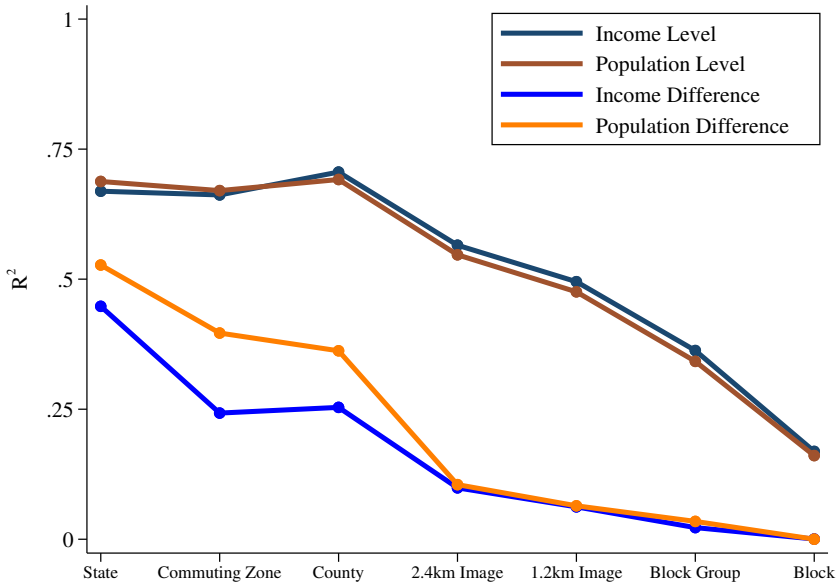


Figure 1.3: Nightlight Predictive Accuracy by Geography

Note: This figure shows the linear fit of log income and log population on log night lights for given geographic units, where measures are in values per km^2. Night light intensity is a spatial sum of DMSP-OLS average visible light in both 2000 and 2010. The regression for each geography is conducted with population weights. Results show that nightlights are a powerful predictor of population and income in large geographies, but their effectiveness in smaller geographies is limited.

images (0.90 versus 0.57) and 0.35 higher for 1.2km images (0.85 versus 0.50); improved accuracy for log population is similar.

The contrast between nightlights and our CNN model is even greater when predicting changes in income or population. For 2000-2010 income changes—where results for population are again similar—R^2 values are 0.10 for nightlights using 2.4km images, compared to 0.40 in our CNN with initial conditions (or 0.37 without them), and 0.06 for nightlights using 1.2km images, compared to 0.32 in our CNN with initial conditions (or 0.27 without them).[16] At the neighborhood dimension of our 1.2km images, changes in nightlights have weak predictive power for changes in economic activity.[17]

1.4.3 Robustness Exercises

We examine the robustness of our results to changes in the satellite imagery and machine-learning methods used in the analysis.

Performance with RGB Only

We consider the effect of limiting the Landsat imagery used for training to the visible spectrum (i.e., the red, green, and blue (RGB) channels). The non-RGB bands in our imagery more than double the size of the data and therefore significantly increase training complexity. It is therefore useful to examine whether the added modelling complexity of using non-RGB data is justified.

Table 1.5 compares test accuracy on models trained with RGB bands alone and those trained with all 7 Landsat bands. For levels models with initial conditions, we find a modest benefit of adding the four non-RGB bands: the R^2 rises by 0.04 for both log income and log population. The gain is larger for difference models: including the additional non-visible Landsat bands

[16] Consistent with previous literature, we find that nightlights have sizable predictive power for long-run income changes in larger geographies, achieving R^2 values of 0.45 for states and 0.25 for counties.

[17] This lack of predictive power for nightlights may be due to the fact that the resolution of 1.2km images is close to that of the 1km pixels for which raw nightlight imagery are available. At the pixel level, perhaps unsurprisingly, changes in nightlights have little information about income or population growth.

raises the R^2 by 0.06 for log income and by 0.11 for log population. For predicting log growth in income and population, having more complete spectral imagery is of substantial value in predictive accuracy.

Performance of 30m (low) vs 15m (high) Resolution Imagery

The resolution of satellite imagery is a key determinant of the information observable in a fixed image region. The USGS Landsat 7 imagery we use has a native $30m$ resolution. Governments and private companies are working to produce more resolute images. DigitalGlobe, for instance, collects and sells satellite imagery with $30cm$ resolution, where a single $30m$ pixel contains 10,000 $30cm$ pixels. Although such high-resolution data promise massive advances in information content, these gains are counter-balanced by similarly massive increases in computational complexity.

To provide a partial evaluation of the gains to prediction from having higher resolution imagery, we compare model performance when doubling the resolution of daytime satellite imagery from $30m$ to $15m$. To perform this comparison, we construct $15m$ Landsat imagery using panchromatic sharpening, as described and used in Jean et al. (2016). This process restricts the Landsat spectral bands to the RGB wavelengths. The results, which appear in Table 1.6, contrast the accuracy of CNN models trained on $1.2km$ images for $30m$ versus $15m$ pan-sharpened RGB bands. To reduce computational complexity, we limit the images used in model training to those in the Mid-Atlantic and Southeast US, as shown in Figure 1.6. Results on test samples indicate that using the higher resolution imagery leads to no meaningful improvement in fit across model specifications. For all models, increases in R^2 are less than 0.005. This finding suggests that modestly higher resolution imagery is unlikely to offer large improvements in a network's ability to learn relevant features for out-of-sample prediction at a fixed geographic scale. However, we cannot speak to the possible model accuracy if substantially higher resolution imagery were coupled with the computational resources to conduct a similar exercise.

1.4.4 Out-of-Sample Predictions

A primary application of our model is to use income and population predictions as outcomes for analyses occurring over periods in which Census data are coarse or unavailable. We offer examples of such analyses in Section 1.4 and guidance on implementing them in the online appendix. To evaluate the accuracy of our predictions in out-of-sample time periods, we train and tune a modified model in which we allocate 70% of our images to training and 30% to validation. In this case, we evaluate model performance in periods outside of 2000 and 2010, rather than in a dedicated set of test images as in our baseline models. To estimate accuracy in periods as far from our sample period as possible, we use 2020 for population and 2017 for income.[18]

Table 1.2: Model R^2 for National 2.4km Imagery in Out-of-sample Periods

	In-Sample Period		Out-of-Sample Period		
Population	**2000, 2010**	**2000-2010**	**2020**	**2010-2020**	**2000-2020**
With Initial Conditions	0.9356	0.5132	0.9193	0.1963	0.4967
Without Initial Conditions	0.8806	0.5030	0.8737	0.1702	0.5106
Income	**2000, 2010**	**2000-2010**	**2017**	**2007-2017**	**2000-2017**
With Initial Conditions	0.9043	0.4910	0.8928	-0.0432	0.4193
Without Initial Conditions	0.8463	0.4331	0.8302	-0.0999	0.3731

Note: The table shows R^2 values computed on all images with 2.4km sides. The sample size of spatially unique images in training and validation subsets is 112,932. Income measures the log of total personal income, while population is the log of total population. The columns delineate fit in the training period and in the out of sample periods, both in terms of levels and differences. Because our imagery panel concludes in 2019, predictions on 2019 imagery are evaluated against actual 2020 population, and 2009 to 2019 change predictions against 2010 to 2020 population change. Initial conditions included in the model are gender and racial composition, residential employment shares and county level population and income, all measured in the initial period (2000 for demographics, 2004 for employment).

Table 2 shows the accuracy of these models when used to predict log population and log income in each period for our larger $2.4km$ images. We find in-period accuracy similar to our baseline model, at 0.90 to 0.94 for levels predictions and 0.49 to 0.51 for time differences (when including initial conditions). This approach also performs well in predicting out-of-sample levels: the R^2 for the levels models including initial conditions is 0.92 for 2020 population and 0.89 for 2017 income. There is little loss in accuracy for predictions in levels when we extend beyond our

[18]Block population for 2020 is from the Census Redistricting File; income for 2017 is from the 2015-2019 ACS and imputed to Blocks using the 2020 population.

sample period.

For the more challenging task of predicting out-of-sample changes, we achieve an R^2 of 0.20 for the change in log population over 2010 to 2020, approximately half of the accuracy seen in our baseline results in the in-sample-period holdout test set. However, the income model is unable to outperform the true mean (i.e., $R^2 = 0$) when forecasting income changes over 2007 to 2017. Performance improves markedly when we instead set our base period to be the in-sample year of 2000 and let the end period extend 7 to 10 years beyond the sample. R^2 values are 0.50 for the 2000-2020 population change and 0.42 for the 2000-2017 income change (with initial conditions), which are similar to results for the 2000-2010 sample period.

Lower performance in predicting changes, particularly for income over 2007-2017, may be related to the sluggish recovery to the Great Recession, which may have dampened changes in the visible properties of economic growth. During this period, falling unemployment drove economic growth, a type of cyclical adjustment for which our CNN may be poorly suited. A second explanation is lower quality label data in the out-of-sample periods, particularly for income. Because block-level population is only available in decennial census years, we use the 2010 and 2020 population distributions to disaggregate 2007 and 2017 income, respectively, from block groups to blocks. The resulting noise may be more problematic over a 10-year period than over the longer periods tested, explaining the difference in accuracy. Because this label quality issue coincides with recessionary years, we are unable to disentangle the two explanations.

We conclude from the results in Table 1.2 that, when evaluated against high quality label data, our approach shows strong potential for producing accurate predictions in out-of-sample periods. The results also indicate that this approach is likely to be most effective when predicting changes over long time horizons, and in periods which do not include large business cycle fluctuations.

1.5 Discussion

Remotely sensed data have the potential to transform spatial economic analysis. Because much of these data are in the public domain, the cost of working at fine geographic scales is now low. We show that applying convolutional neural networks to daytime satellite imagery predicts microspatial changes in income and population at a decadal frequency. An immediate application is to use predictions of income or population at these spatial scales as outcomes in analysis. Our method can also be used to impute income and population between Census years for the US, to extend to other high-income countries where the relationship between multi-spectral imagery and economic activity is likely to be similar, and to initialize layers for training CNNs in other contexts, thereby reducing computational costs. Khachiyan et al. (2021), for example, uses our output to examine the within-county impacts of the US fracking boom.

A related area that would benefit from such data is the study of place-based policies, such as subsidies to firms that invest in designated areas. Justifying these policies hinges on whether new investments have positive spatial spillovers (Kline and Moretti, 2014; Gaubert et al., 2021). Using our model, researchers could evaluate spillovers at much finer spatial scales than is feasible with public data. Estimating the welfare consequences of place-based policies relies further on addressing their non-random location and timing. With our model, researchers could examine pre-existing trends and control for spatial-temporal shocks at much finer resolutions (e.g., county-year levels) than is possible in conventional data (in which the county-year may be the unit of analysis).

Another application is the evaluation of transport infrastructure, which has seen major recent advances (Redding, 2020). Satellite-based measures of income and population would allow researchers to evaluate specific projects, such intra-city bus lanes or subway lines, at the neighborhood level across many cities. Such granularity would permit refined tests of economic theory, such as whether transport links lead to more agglomeration in larger nodes (via home market effects) or less agglomeration in intermediate nodes (due to agglomeration shadows). Although researchers have obtained granular information from smartphone data (e.g., Akbar et al., 2018,

Kreindler and Miyauchi, 2021) and private transport platforms (e.g., Hall et al., 2018), there may be non-random selection of users who supply these data (e.g., taxi riders in New York City may differ from taxi riders in Phoenix). Satellite imagery offers the equivalent of administrative-level data that is consistent across space and time.

A further application is the analysis of natural disasters. Floods, earthquakes, wildfires, and tornadoes tend to have highly localized impacts (Dell et al., 2014). Our model allows analysts to trace the consequences from point of impact to neighboring communities and to broader metro areas. Such disaggregation is important not just for the academic task of evaluating shock transmission across space but for policy makers who, after disasters occur, require tools to assess where need is likely to be acute.

Finally, our results suggest paths for future work developing predictive models from satellite imagery. First, the model does not perform as well in the shorter frequency out-of-sample prediction exercise, although this could be due to business cycles. Addressing this issue could leverage further the ability to use higher-frequency changes in images to predict economic growth. Second, our model is trained on US data, and future work could explore how well model parameters perform in other countries.

1.6 Acknowledgements

Chapter 1 has been accepted for publication at the American Economic Review: Insights. Khachiyan, A.; Thomas, A.; Zhou, H.; Hanson, G.; Cloninger, A.; Rosing, T.; Khandelwal, A. The dissertation/thesis author was the primary investigator and author of this material. This project was funded through the generous support of the Russell Sage Foundation program on Computational Social Science.

Bibliography

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and Xiaoqiang Zheng (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Akbar, P. A., V. Couture, G. Duranton, and A. Storeygard (2018, November). Mobility and congestion in urban india. Working Paper 25218, National Bureau of Economic Research.

Babenko, B., J. Hersh, D. Newhouse, A. Ramakrishnan, and T. Swartz (2017). Poverty mapping using convolutional neural networks trained on high and medium resolution satellite images, with an application in mexico. *arXiv preprint arXiv:1711.06323*.

Baum-Snow, N., L. Brandt, J. V. Henderson, M. A. Turner, and Q. Zhang (2017). Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics 99*(3), 435–448.

Boustan, L. P., M. E. Kahn, P. W. Rhode, and M. L. Yanguas (2020). The effect of natural disasters on economic activity in us counties: A century of data. *Journal of Urban Economics*, 103257.

Bruederle, A. and R. Hodler (2018). Nighttime lights as a proxy for human development at the local level. *PloS one 13*(9), e0202231.

Census (2020). Lehd origin-destination employment statistics data (2004): Version 7 [dataset]. *Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program.*.

Chen, X. and W. D. Nordhaus (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences 108*(21), 8589–8594.

De Fries, R., M. Hansen, J. Townshend, and R. Sohlberg (1998). Global land cover classifications at 8 km spatial resolution: the use of training data derived from landsat imagery in decision tree classifiers. *International Journal of Remote Sensing 19*(16), 3141–3168.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature 52*(3), 740–98.

Donaldson, D. and A. Storeygard (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives 30*(4), 171–98.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies 81*(3), 1046–1070.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*, Volume 1. Springer series in statistics New York.

Gaubert, C., P. M. Kline, and D. Yagan (2021). Place-based redistribution. Working Paper 28337, National Bureau of Economic Research.

Gennaioli, N., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2013). Human capital and regional development. *The Quarterly Journal of Economics 128*(1), 105–164.

Glorot, X. and Y. Bengio (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep learning*. MIT press.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.

Greenstone, M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy 118*(3), 536–598.

Hall, J. D., C. Palsson, and J. Price (2018). Is uber a substitute or complement for public transit? *Journal of Urban Economics 108*, 36–50.

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2018). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics 133*(1), 357–406.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American economic review 102*(2), 994–1028.

Hjort, J. and J. Poulsen (2019, March). The arrival of fast internet and employment in africa. *American Economic Review 109*(3), 1032–79.

Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon (2016). Combining satellite imagery and machine learning to predict poverty. *Science 353*(6301), 790–794.

Jedwab, R. and A. Storeygard (2021, 06). The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 1960–2010. *Journal of the European Economic Association*.

Khachiyan, A., A. Thomas, H. Zhou, G. Hanson, A. Cloninger, T. Rosing, and A. Khandelwal (2021). Using Neural Networks to Predict Micro-Spatial Economic Growth. Working paper.

Kingma, D. P. and J. Ba (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kline, P. and E. Moretti (2014). People, places, and public policy: Some simple welfare economics of local economic development programs. *Annual Review of Economics 6*(1), 629–662.

Kreindler, G. E. and Y. Miyauchi (2021, February). Measuring commuting and economic activity inside cities with cell phone records. Working Paper 28516, National Bureau of Economic Research.

LeCun, Y., Y. Bengio, and G. Hinton (2015). Deep learning. *Nature 521*(7553), 436–444.

Manson, S., J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles (2020). Ipums national historical geographic information system: Version 15.0 [dataset]. *Minneapolis, MN: IPUMS*. `http://doi.org/10.18128/D050.V15.0`.

Michalopoulos, S. and E. Papaioannou (2014). National institutions and subnational development in africa. *The Quarterly Journal of Economics 129*(1), 151–213.

NOAA (2020). Dmsp ols: Nighttime lights time series version 4 [dataset]. *National Oceanic and Atmospheric Administration via Google Earth Engine. NOAA/DMSP-OLS/NIGHTTIME_LIGHTS. https://developers.google.com/earth-engine/datasets/catalog/NOAA_DMSP-OLS_NIGHTTIME_LIGHTS#description.*.

Pesaresi, M., D. Ehrlich, S. Ferri, A. Florczyk, S. Freire, M. Halkia, A. Julea, T. Kemper, P. Soille, and V. Syrris (2016). Operating procedure for the production of the global human settlement layer from landsat data of the epochs 1975, 1990, 2000, and 2014. *Publications Office of the European Union*, 1–62.

Piaggesi, S., L. Gauvin, M. Tizzoni, C. Cattuto, N. Adler, S. Verhulst, A. Young, R. Price, L. Ferres, and A. Panisson (2019). Predicting city poverty using satellite imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 90–96.

Redding, S. J. (2020). Trade and geography. Working Paper 27821, National Bureau of Economic Research.

Redding, S. J. and M. A. Turner (2015). Transportation costs and the spatial organization of economic activity. In *Handbook of Regional and Urban Economics*, Volume 5, pp. 1339–1398. Elsevier.

Rolf, E., J. Proctor, T. Carleton, I. Bolliger, V. Shankar, M. Ishihara, B. Recht, and S. Hsiang (2021). A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*.

Rosenstein, M. T., Z. Marx, L. P. Kaelbling, and T. G. Dietterich (2005). To transfer or not to transfer. In *NIPS 2005 workshop on transfer learning*, Volume 898, pp. 1–4.

Samek, W., A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller (2016). Evaluating the visualization of what a deep neural network has learned. *IEEE transactions on neural networks and learning systems 28*(11), 2660–2673.

Simonyan, K., A. Vedaldi, and A. Zisserman (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.

Simonyan, K. and A. Zisserman (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in sub-saharan africa. *The Review of Economic Studies 83*(3), 1263–1295.

Ural, S., E. Hussain, and J. Shan (2011). Building population mapping with aerial imagery and gis data. *International Journal of Applied Earth Observation and Geoinformation 13*(6), 841–852.

USGS (2020). Landsat 7 surface reflectance tier 1 [dataset]. *United States Geological Survey via Google Earth Engine. LANDSAT/LE07/C01/T1_SR. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LE07_C01_T1_SR#description.*.

Zeiler, M. D. and R. Fergus (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, pp. 818–833. Springer.

Zha, Y., J. Gao, and S. Ni (2003). Use of normalized difference built-up index in automatically mapping urban areas from tm imagery. *International journal of remote sensing 24*(3), 583–594.

1.7 Appendix

Modelling Appendix

Our modelling approach has two stages. We first train a multi-layer convolutional neural network model, which we use to predict outcomes in levels (income and population in a given year). Our models and training pipelines are implemented in TensorFlow Abadi et al. (2015). Our model architecture is a 7-band version of the VGG16 network model, which is widely used in the computer vision community (Simonyan and Zisserman, 2014) and consists of three convolutional blocks followed by a fully connected block. Each convolutional block consists of three two-dimensional convolution layers followed by a max-pooling layer. The output of the final convolutional block is flattened into a vector, which is used as input to the fully connected block. The fully connected block consists of three hidden layers, each separated by a dropout layer. The weights of each layer in the fully connected block are regularized using an L2 norm penalty. To incorporate initial conditions in the models, we standardize all features to be of zero mean and unit variance and concatenate the resulting feature vector to the flattened representation obtained by the CNN. The resulting augmented image representation is then processed by the fully connected block to form predictions. A detailed description of model architecture, including filter sizes and strides, is in the Appendix and in our code on `GitHub`. Figure 1.4 shows our model architecture.

We use the model trained in levels to construct a model for predicting time differences in the outcome variables over a given time period (e.g., 2000 to 2010). For each year, we first extract the image representation using the convolutional filters learned by training the levels model, as described above. We then concatenate the vectorized representations for each year and use this as input to a new fully connected block, which is used to predict the difference in outcomes between the two years.

More formally, let $I_a, I_b \in \mathbb{R}^{r \times c \times 7}$ be the input images in years a and b, respectively. We first instantiate a copy of the convolutional layers of the levels model described above, which we denote

as a function $f_\phi : \mathbb{R}^{r \times c \times 7} \to \mathbb{R}^d$. The parameters ϕ are initialized to the weights learned by training the levels model. The predicted outcome of interest is then modeled as $\hat{y} = f_\psi(f_\phi(I_a), f_\phi(I_b))$, where f_ψ can be described by concatenating its two arguments and then applying a dense block as described above. The set of parameters ϕ and ψ is then optimized to minimize the mean-squared-error of the prediction. In this process, we use the levels model to "warm-start" the training of the differences model, based on the intuition that features salient for predicting differences are likely related to, but not coincident with, those for predicting levels.

Levels Models. The levels models consist of three convolution blocks, a "flatten" layer which vectorizes the output of the convolution layers, and a dense block, which is used to predict the outcome of interest from the features extracted by the convolution blocks. Weights in all layers are initialized using the Glorot Normal random initialization (Glorot and Bengio, 2010). Each convolution layer block consists of three 2D convolution layers followed by a max pooling layer. The convolution layers use a stride of 1 and a kernel size of 3 with ReLu activations. The convolution kernels are regularized using an L2 norm penalty where the strength of the penalty is chosen using cross-validation as described in the body of the paper. The number of filters is constant within each block and increases by a factor of 2 between each block. In other words, if the first block has n filters, the second block outputs $2n$ filters and the third outputs $4n$. The max-pooling layer pools over a 2×2 window. For models that incorporate nightlight intensity, these are included as another channel in the input image.

The output of the convolution blocks is flattened into a vector which is then passed to the dense block. For models that incorporate baseline features (e.g., county level income or population), these features are concatenated to the vectorized output of the convolution blocks. The dense block consists of three fully connected layers each separated by a dropout layer. The fully connected layers use ReLu activations and are regularized by an L2 norm penalty where the strength of the penalty is again chosen using cross-validation and grid-search. The specific set of parameters considered can be found in our code on GitHub. The dropout probability in dropout layers is fixed at 0.5. The number of hidden units in each fully-connected layer in the dense block is set

based on the number of filters used in the convolution layers. If the first convolution layer outputs n filters, then each fully connected layer uses $l_i \cdot n$ hidden units, where $l_1 = 16, l_2 = 16$ and $l_3 = 8$. The output of the dense block is passed through a final linear layer which produces a scalar value that is the predicted output. This layer is also regularized by an L2 penalty.

Differences Models. The differences model takes a pair of images, of the same spatial region, in different years as input and produces an estimate of the change in the outcome of interest as output. The images for both years are passed through the levels model as described above and the output of the flatten layer is extracted for each year. For models that incorporate auxiliary features, these features are again concatenated to the output of the flatten layer. The image representations extracted for each year are then concatenated and passed to a dense block as described above. The output of the dense block is again passed to a final linear layer which generates the predicted difference in the outcome of interest. The entire architecture, including the convolution filters in the levels models, is then trained end-to-end.

Computing R^2 Values from CNN Predictions

To compute R^2 in our case of highly non-linear CNN models, we use the general formula of $1 - \frac{SSR}{TSS}$. Here SSR is the sum of squared residuals, where each residual is the difference between the predicted and true value for an image. TSS is conversely the Total Sum of Squares, which is the sum across images of squared differences between each true value and the mean true value of the given outcome.

Code and Data Appendix

While highly effective, developing and training CNN models requires significant computational resources and technical expertise. To assist researchers interested in using our predicted outcomes for their own applications, or in adapting our approach to predict other outcomes or generate predictions in periods or countries outside of our analysis, we have made publicly available our entire code pipeline, image labels and predicted values used to generate results in this paper,

and trained CNN models. These resources, along with documentation can be found in our project GitHub at `https://github.com/thomas9t/spatial-econ-cnn.git`.

Specifically, the following resources are available:

- **Code**: The code base used in this paper is available on GitHub. This includes code to (1) extract raw publicly-available imagery from Google Earth Engine and to link imagery with census labels, (2) process image files and convert data to input to the CNN, (3) define and train CNN models, (4) generate predictions using the trained CNN models, and (5) evaluate the accuracy of predictions. Our code base can be directly adapted by researchers to develop new CNN models predicting other outcomes of interest.

- **Model Predictions**: We include CSV files with image-level predictions of income and population levels in each year from 2000 to 2019. We also share predictions of 10-year changes in each outcome for every 10-year period from 2000 to 2019. These predictions are generated for our large images using our out-of-period model (Table 2). These include an image id variable (img_id) and predictions based on models both with and without initial conditions. Each of these files is at the image level, with variables predicting outcomes in a given year and over 10 year changes. Shapefiles of our urban image samples are included for researchers wishing to directly study these geographies. For those interested in aggregating to census geographies, we include a cross-walked version of these predictions for 2010 Census Blocks, which can be further aggregated to containing census geographies (i.e. Counties). Those wishing to study other geographic units will need to construct their own crosswalk between our images and their units of interest. One way to do this would simply be based on the spatial overlap between units of the different geographies.

- **Trained Models**: For researchers interested in generating predictions on geographies or time-periods not described above, we have also made available the trained parameters of our CNN models. Using these pre-trained models, along with the data processing scripts described above, researchers can input their own Landsat data into our models and compute

predicted values. Another use-case would be to use lower levels of our trained CNNs in a transfer learning application in order to reduce the computational cost of training on different economic outcomes.

We also provide documentation and a step-by-step example illustrating how to generate out-of-sample predictions on LANDSAT data not used in our analysis.

Saliency Maps

In addition to validating model performance on a held out test set, it is also useful to assess network performance qualitatively by interpreting the features that appear to be learned by the network. There is a large literature on techniques for interpreting neural network predictions; we focus on saliency mapping, which is simple and widely used (Simonyan et al., 2013; Zeiler and Fergus, 2014; Samek et al., 2016). Saliency maps typically take the form of a heat map showing which pixels in a particular image most strongly influenced the network's prediction. They provide qualitative assurance that the network utilizes "reasonable" features of the image.

The saliency map is generated by calculating the derivative of the score of a class of interest S_c with respect to the input $I \in \mathbb{R}^{r \times c \times d}$ at any image I_0 (Simonyan et al., 2013). In our case, the problem is regression rather than classification. To adapt saliency maps to this setting, we generate the saliency map $M \in \mathbb{R}^{r \times c}$ by

$$M_{ij} = \sum_{c=1}^{7} \left| \omega_{h(i,j,c)} \right|,$$

$$\omega = \left. \frac{\partial f}{\partial I} \right|_{I_0},$$

where f is the entire model for prediction, $\omega_{h(i,j,c)}$ is the i-th row, j-th column and c-th channel of ω. In this way, the saliency map will show which features increase the output most across all the channels.

Figure 1.5 shows several saliency maps for images in urban, suburban, and semi-rural environs and for which model predictions of income in levels are accurate and inaccurate. Examining cases in which the model performs well and poorly at each population density level gives context

on the types of land cover features that are being captured accurately in our models and those that are not. We caution that interpreting saliency is challenging—the motivation for using a CNN is that the relevant image features are unknown and thus one would not expect saliency maps to have in each instance a visually obvious and precise interpretation. Nonetheless, it may be possible to extract some lessons from their examination. Reassuringly, the network ignores water and tends to focus on developed regions in images. For instance, in the second row of the column titled "Rural, Accurate," the network is focusing on the small developed region at the top of the image. Similarly, in the first row of the column "Suburban, Accurate," the model is focusing on the developed region at the lower left. However, in the first and second rows of the column "Urban, Innacurate" the network seems to prioritize undeveloped regions. This is not necessarily a concern, as in some contexts green space is predictive of income. Taken with our quantitative results, which show relatively little evidence of overfitting, the saliency maps suggest that our model is extracting relevant economic information from the images.

Results Appendix

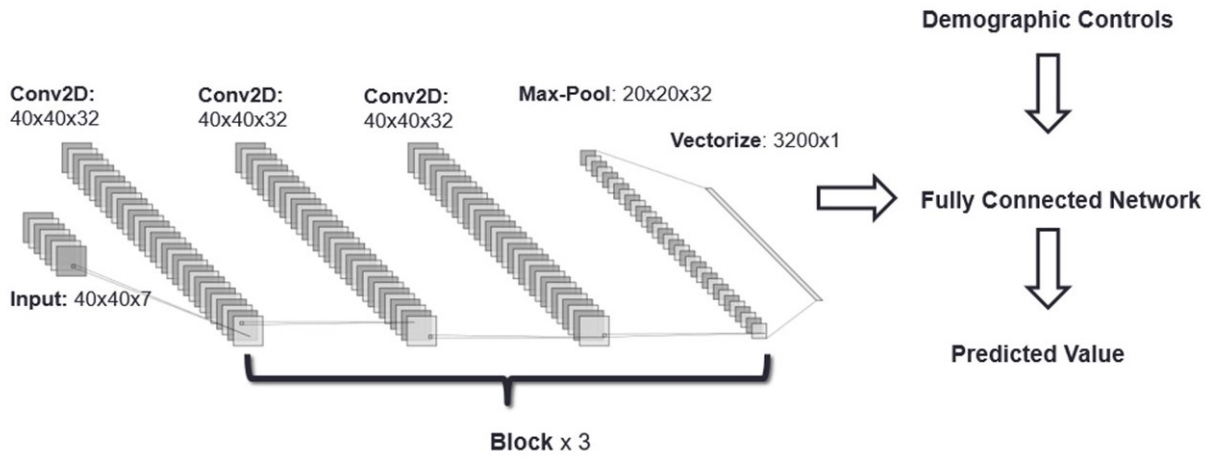


Figure 1.4: Convolutional Neural Network, Landsat Imagery Model Architecture

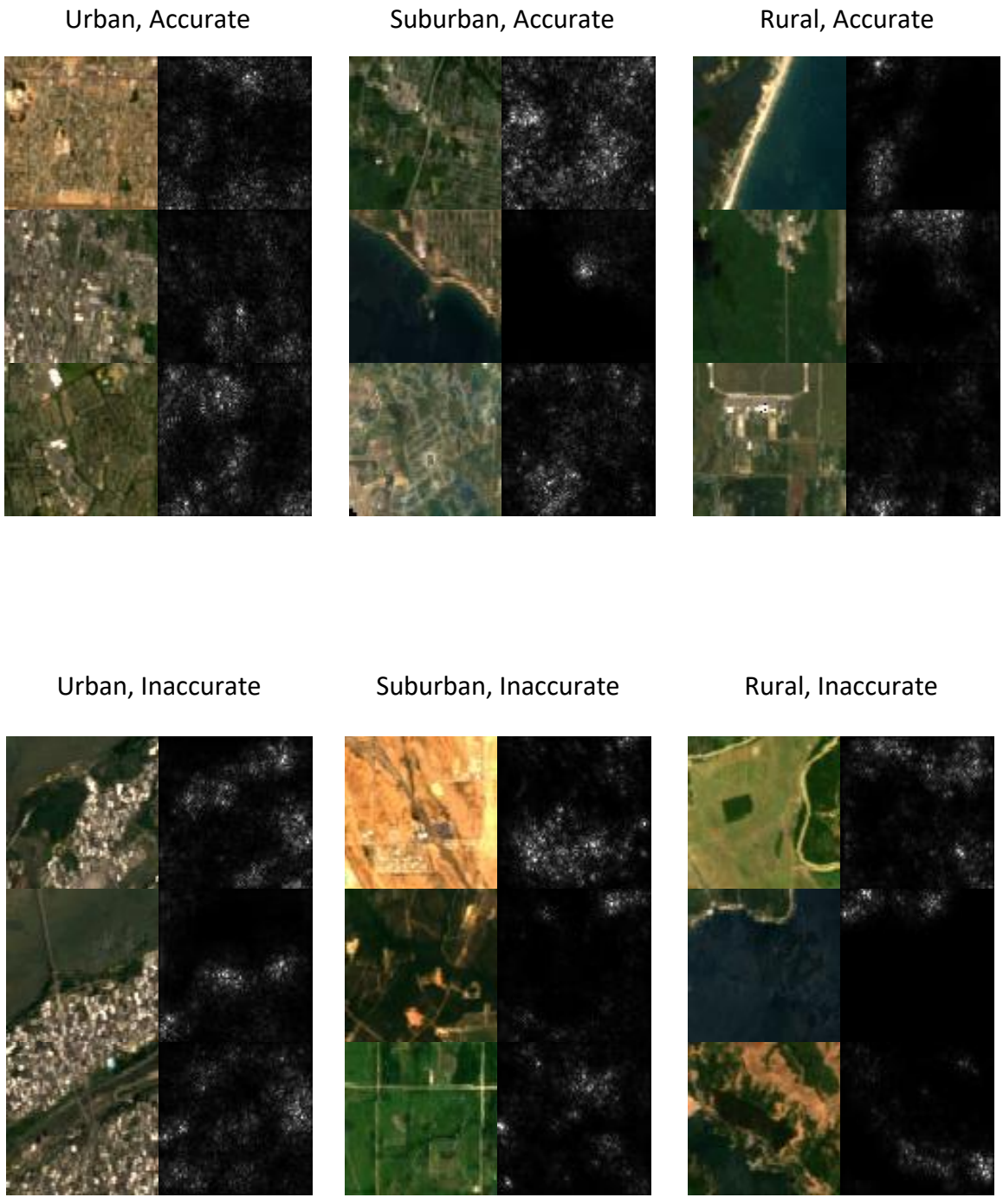


Figure 1.5: Selected Saliency Maps

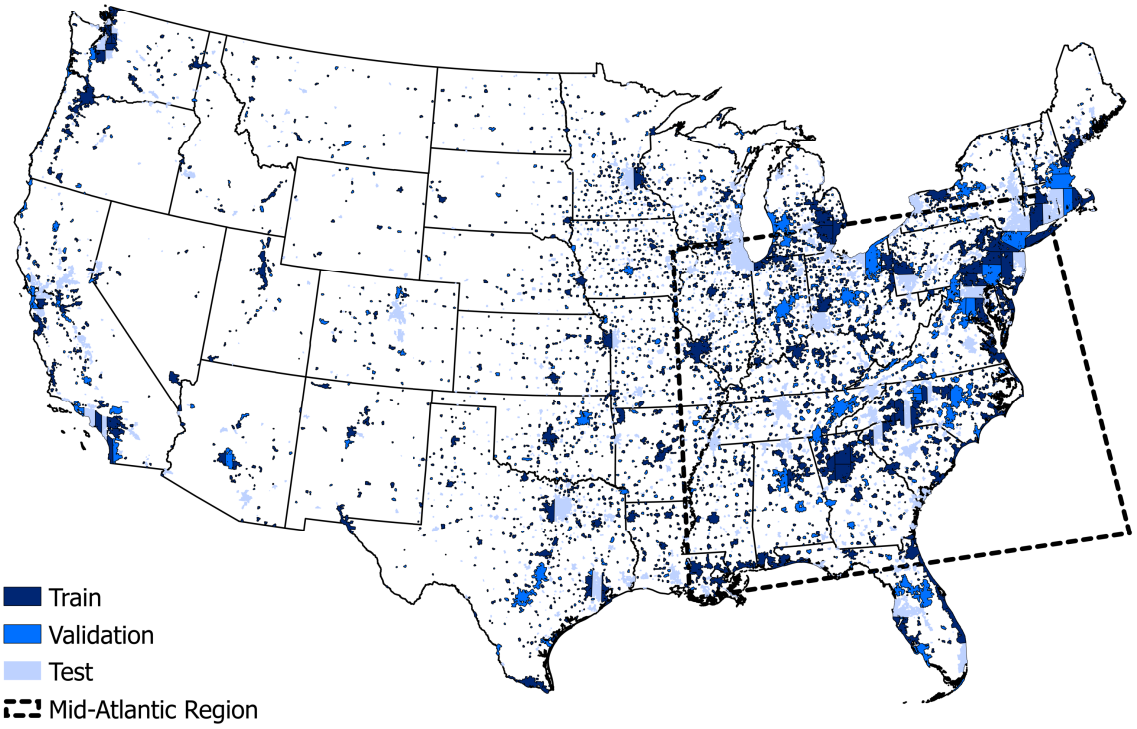


Figure 1.6: Spatial Extent of Urban Areas and Model Development Subsets

Note: This map shows the urban areas of the contiguous Unites States used to assign images into training, validation and testing subsets for our CNN models. The black dotted line represents the Mid-Atlantic region used for our high-resolution imagery robustness tests discussed in Section 1.4.3. The sample subsets represented in this map are randomly generated and used in training the model for large images; separate randomization is conducted to subset the areas for small images and the high-resolution robustness check. Blank space on this map represents low population density regions (approximately 7% of total population), which are not included in our analysis of urban areas.

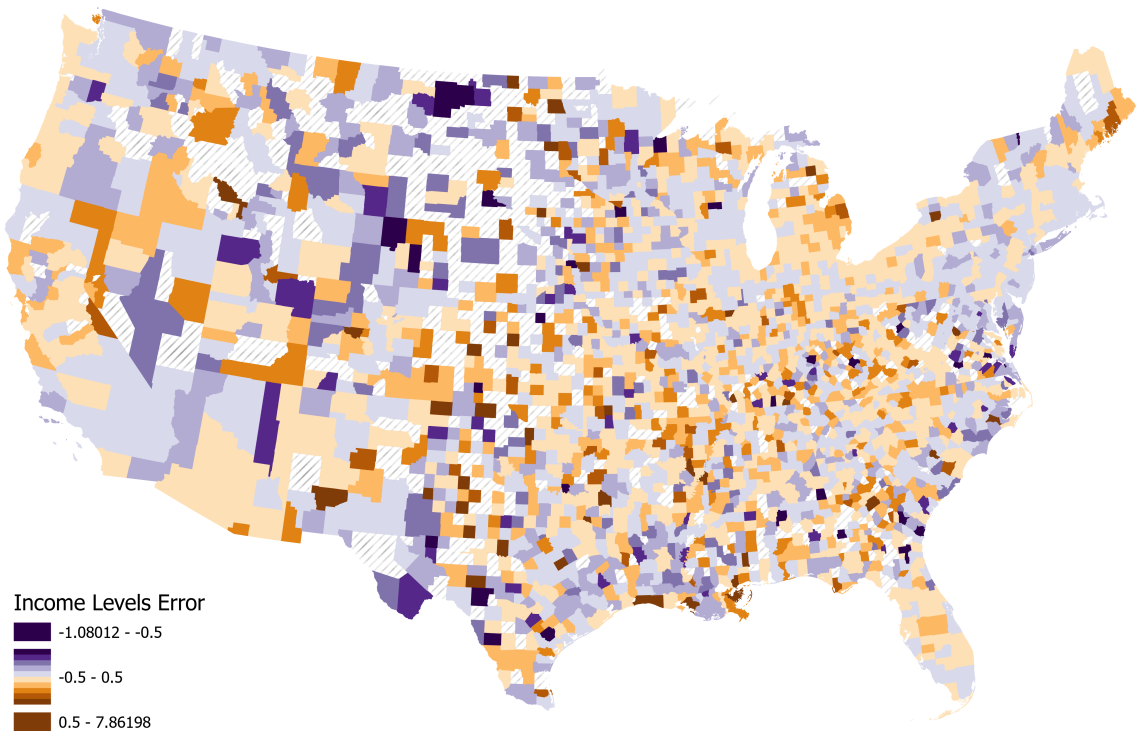


Figure 1.7: Levels Model Average Prediction Error across Counties

Note: This map shows the spatial distribution of our modelling error across US counties. We compute image-level prediction error as the average of predicted log income minus actual log income in 2000 and 2010. The color of each county in the map represents the average of these prediction errors across all images in the county.

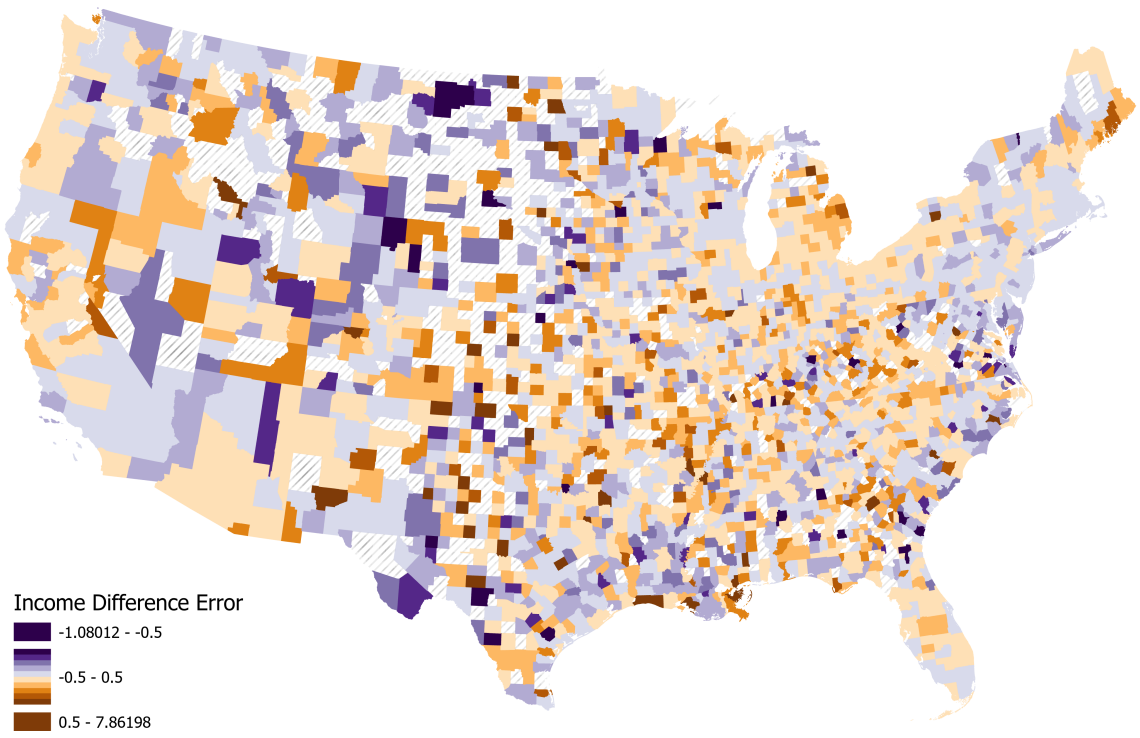


Figure 1.8: Differences Model Average Prediction Error Across Counties

Note: This map shows the spatial distribution of our modelling error across US counties. We compute image-level prediction error as predicted income change from 2000 to 2010 minus actual income change for the same period. The color of each county in the map represents the average of this prediction error across all images in the county.

Table 1.3: Prediction Error Correlations with Covariates and Geography

	Income Level	Income Difference	Population Level	Population Difference
Female	–0.0925	0.0409	–0.0620	0.0591
Emp in Business Services	–0.0750	–0.0026	–0.0268	–0.0227
Emp in Accommodation & Food Services	0.0562	0.0570	0.0659	0.0446
Emp in Wholesale Trade	–0.0521	–0.0336	0.0013	–0.0450
Log Income, County	–0.0520	–0.0049	–0.0011	–0.0230
Emp in Admin/Support/Waste/Remediation Services	–0.0499	0.0247	–0.0127	0.0166
Emp in Production, County	–0.0467	0.0211	–0.0294	0.0387
White	0.0428	–0.0022	0.0101	–0.0052
Emp in Non-Business Services	0.0420	0.0358	–0.0076	0.0408
Log Population, County	–0.0417	–0.0038	0.0015	–0.0179
Hispanic	–0.0415	0.0088	–0.0054	–0.0010
Emp in Business Services, County	–0.0369	0.0173	–0.0426	0.0101
Emp in Construction	–0.0369	0.0169	–0.0470	0.0054
Emp in Professional/Scientific/Technical Services	–0.0364	–0.0008	–0.0288	–0.0077
Emp in Real Estate, Rental & Leasing	–0.0358	0.0042	–0.0273	–0.0046
Emp in Production	–0.0337	–0.0000	–0.0112	0.0189
Emp in Finance & Insurance	–0.0336	–0.0123	–0.0035	–0.0397
Emp in Non-Business Services, County	0.0335	0.0211	–0.0403	0.0451
Emp in Public Administration	0.0293	0.0074	–0.0132	0.0045
Black	–0.0266	0.0052	–0.0266	0.0245
Emp in Information	–0.0251	–0.0000	0.0048	–0.0247
Emp in Transportation and Warehousing	–0.0243	–0.0012	0.0125	–0.0145
Group Quarters	–0.0230	0.0010	–0.0090	0.0161
Emp in Mining/Quarrying & Oil/Gas Extraction	0.0226	–0.0431	–0.0060	0.0080
Emp in Manufacturing	–0.0224	0.0030	0.0165	0.0175
Emp in Retail Trade	–0.0178	0.0060	–0.0353	0.0153
Emp in Agriculture, Forestry, Fishing, & Hunting	–0.0153	0.0010	–0.0227	–0.0004
Emp in Arts, Entertainment & Recreation	–0.0142	0.0258	–0.0062	0.0030
Emp in Health Care & Social Assistance	0.0131	0.0081	–0.0170	0.0352
Emp in Management	–0.0128	–0.0017	0.0067	0.0029
Emp in Utilities	0.0122	–0.0060	0.0049	0.0002
Emp in Educational Services	0.0062	–0.0077	–0.0380	–0.0020
Emp in Other Services	–0.0023	0.0146	–0.0112	0.0123
Working Age	0.0023	–0.0555	0.0020	–0.0718
Urban Area Fixed Effects	0.1067	0.0803	0.0890	0.0669

Note: The table reports correlation coefficients between covariates and prediction errors in each of the four prediction exercises: log income in 2000 and 2010, the change in log income from 2000 to 2010, and the corresponding values for population. The final row shows the R^2 coefficient of an OLS regression on fixed effects by contiguous urban areas (as shown in Figure 1.4). All covariates measure an initial value (2004 for employment, 2000 for the rest) at the image-level, and all but the initial county income and population columns represent shares of the relevant image population. These values are spatially interpolated to images from Census Block labels, with the exception of rows listed as County. Residential employment shares are broken down by two-digit NAICS manufacturing industries as well as the aggregates Business Services, Non-Business Services, and Production. Rows are sorted from highest to lowest correlation for income levels. Prediction errors are constructed based on models which include initial conditions.

Table 1.4: R^2 Values for Income Per Capita in Large and Small Images

	2000 and 2010 Levels			2000 to 2010 Difference		
	Train	Valid	Test	Train	Valid	Test
National 2.4*km* Imagery						
With Initial Conditions	0.7049	0.6795	0.6533	0.1220	0.0624	0.0674
Without Initial Conditions	0.5077	0.4276	0.3884	0.0984	0.0407	0.0461
National 1.2*km* Imagery						
With Initial Conditions	0.7011	0.6166	0.6091	0.0838	0.0621	0.0653
Without Initial Conditions	0.4502	0.3037	0.3317	0.0534	0.0360	0.0306

Note: The table shows R^2 values computed on each subset of the images with 2.4*km* and 1.2*km* sides. The total sample size of spatially unique images in training, validation and test subsets is 112,932 for larger images and 320,880 for smaller images. Income per capita measures the log of total personal income per person. 2000 and 2010 levels represent a model predicting levels for images in the two years together, while the differences columns show the result predicting the change from 2000 to 2010. Initial conditions included in the model are gender and racial composition, employment shares and county level population and income, all measured in 2000. The results show that predictions on income per capita are less accurate than those on income or population separately, particularly when predicting differences and excluding initial conditions.

Table 1.5: Model R^2 for National 2.4km Imagery: All Bands vs RGB Only

	2000 and 2010 Levels			2000 to 2010 Difference		
	RGB Only	LS Bands	LS + NL	RGB Only	LS Bands	LS + NL
Income						
With Initial Conditions	0.8580	0.9018	0.8949	0.3330	0.3962	0.3917
Without Initial Conditions	0.7502	0.8374	0.8429	0.2815	0.3702	0.3827
Population						
With Initial Conditions	0.8781	0.9132	0.9025	0.3467	0.4573	0.4408
Without Initial Conditions	0.7952	0.8684	0.8571	0.3197	0.4202	0.4538

Note: The table shows R^2 values computed on the test set of images with 2.4km sides. The total sample size of spatially unique images in training, validation and test subsets is 112,932. Income measures the log of total personal income, while population is the log of total population. 2000 and 2010 levels represent a model predicting levels for images in the two years combined, while the differences columns show the result predicting the change from 2000 to 2010. Initial conditions included in the model are gender and racial composition, employment shares and county level population and income, all measured in 2000. RGB Only refers to the red/green/blue Landsat 7 bands, LS refers to all 7 Landsat Bands, LS+NL Refers to all 7 Landsat bands plus the DMSP-OLS nightlight band. The results show that including the non-visible Landsat bands improves the model performance, particularly in predicting differences. Further including nightlight data does not improve models (with initial conditions).

Table 1.6: Model R^2 in Mid-Atlantic Region: 30m vs 15m Resolution RGB Imagery

	2000 and 2010 Levels		2000 to 2010 Difference	
	30m RGB	15m RGB	30m RGB	15m RGB
Income				
With Initial Conditions	0.7997	0.7970	0.2499	0.2320
Without Initial Conditions	0.6644	0.6683	0.2014	0.1773
Population				
With Initial Conditions	0.8167	0.8189	0.2749	0.2492
Without Initial Conditions	0.7159	0.6995	0.2545	0.2265

Note: The table shows R^2 values computed on the test set of images with 1.2km sides. The total sample size of spatially unique images in training, validation and test subsets is 163,250. Income measures the log of total personal income, while population is the log of total population. 2000 and 2010 levels represent a model predicting levels for images in the two years combined, while the differences columns show the result predicting the change from 2000 to 2010. Initial conditions included in the model are gender and racial composition, employment shares and county level population and income, all measured in 2000. 30m RGB refers to the same Landsat 7 RGB bands used in the rest of the analysis. 15m RGB refers to pan-sharpened RGB bands which are refined from 30m to 15m resolution using the panchromatic Landsat band. All results in this table are based on the Mid-Atlantic subset of the national imagery, shown in Figure 1.6, to address the additional computation of analyzing imagery with double resolution. Results show that the extra information of 15m resolution images does not meaningfully improve model accuracy relative to equally sized images with 30m pixels.

Table 1.7: Model R^2 for National Imagery By Year

	2000			2010			Diff		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Panel A: National 2.4*km* Imagery									
Income									
With Initial Conditions	0.9287	0.8981	0.9029	0.9221	0.8887	0.9006	0.4863	0.4126	0.3962
W/out Initial Conditions	0.8672	0.8330	0.8373	0.8577	0.8248	0.8375	0.4951	0.3960	0.3702
Population									
With Initial Conditions	0.9610	0.9061	0.9146	0.9613	0.8996	0.9119	0.5410	0.4839	0.4573
W/out Initial Conditions	0.9186	0.8620	0.8669	0.9189	0.8652	0.8700	0.7004	0.4496	0.4202
Panel B: National 1.2*km* Imagery									
Income									
With Initial Conditions	0.9032	0.8729	0.8615	0.8883	0.8512	0.8470	0.3819	0.3061	0.3216
W/out Initial Conditions	0.7988	0.7604	0.7482	0.7949	0.7591	0.7507	0.2959	0.2609	0.2690
Population									
With Initial Conditions	0.9149	0.8788	0.8650	0.9052	0.8645	0.8548	0.4217	0.3401	0.3559
W/out Initial Conditions	0.7815	0.7602	0.7452	0.7867	0.7623	0.7532	0.3924	0.3051	0.3036

Note: The table shows R^2 values computed on each subset of the images with 2.4*km* and 1.2*km* sides. The total sample size of spatially unique images in training, validation and test subsets is 112,932 for larger images and 320,880 for smaller images. Income measures the log of total personal income, while population is the log of total population. Results for 2000 and 2010 are shown separately here, while the differences columns show the result predicting the change from 2000 to 2010 as in Table 1.1. Initial conditions included in the model are gender and racial composition, employment shares and county level population and income, all measured in 2000. The results show high accuracy in predicting both levels and differences in income and population; there is not strong evidence of over-fitting in the training set. Model fit is lower using the smaller images.

Chapter 1 has been accepted for publication at the American Economic Review: Insights. Khachiyan, A.; Thomas, A.; Zhou, H.; Hanson, G.; Cloninger, A.; Rosing, T.; Khandelwal, A. The dissertation author was the primary investigator and author of this material. This project was funded through the generous support of the Russell Sage Foundation program on Computational Social Science.

Chapter 2

The Impacts of Fracking on Micro-Spatial Residential Investment

Arman Khachiyan[1]

2.1 Abstract

As fracking has become the dominant method of oil and gas extraction in the US, the population living within 1 mile of a well has quadrupled to over 10 million. While environmental externalities of this extraction are most concentrated within 1 mile, profits predominantly flow out of the host county. Measuring the intensity and distribution of fracking impacts on adjacent neighborhoods requires outcomes with high spatial resolution. I study changes in total neighborhood income and population which are derived from machine learning models trained to identify urban growth in daytime satellite imagery. Coupled with a precise shale geology instrument, my microspatial approach identifies that fracking exposure as far as 20 miles away leads to a 2 percent decline in neighborhood income. The spatial gradient and associated mechanisms of this effect

[1]Support was provided by the Russell Sage Foundation Project Grant on Computational Social Science. I thank Gordon Dahl, Gordon Hanson, Judson Boomhower, Alexander Gelber, and James Fowler for patient and insightful advising, Anthony Thomas, Huye Zhou, and Luke Sanford for guidance and technical assistance, and Julie Cullen, Becky Fraenkel, Katarzyna Bilicka, Adam Storeygard, Wes Howden, Sam Krumholz, Gordon McCord and many seminar participants for thoughtful comments. All errors are my own.

indicate that it is driven by local industrialization rather than direct environmental externalities. While this effect is exacerbated by more extraction, it completely attenuates in areas with strong environmental protections or employment specialization in relevant sectors.

2.2 Introduction

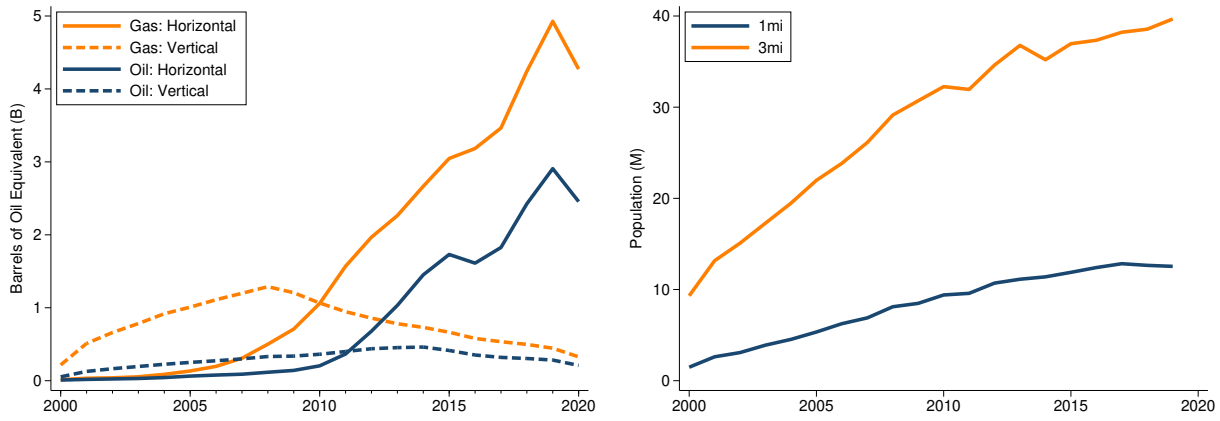
Across a broad class of localized economics shocks a general question arises: how will exposed neighborhoods be impacted? Consider the case of a coal plant replacing an empty field. For the impacted county this implies an increase in employment and public revenue. However, a family living across the street bears the negative externalities from air pollution and degraded neighborhood aesthetics. Many ongoing industrial transitions such as infrastructure investment, renewable energy farms, distribution centers, retail zones, corporate offices, and even natural disasters all exhibit similar patterns: different mechanisms generate the dominant impacts at the neighborhood, county, national, and global scales. As such, inference on aggregated administrative areas cannot be assumed to apply to smaller geographies contained therein.[2] Indeed major policy decisions around industrial transitions are often decided largely based on the perceived risks for the most exposed communities.

For example the State of New York justified its decision to ban fracking in 2014 based on "significant uncertainties about the kinds of adverse health outcomes that may be associated with [fracking]" (Zucker and Dreslin, 2014). In many shale-rich parts of the country without outright bans, a debate on setback distances between wells and occupied structures is currently unfolding.[3] The first source of local disamenity is the initial drilling and fracking of a well, a weeks long process which involves regular truck traffic, noise and light pollution, and in some cases many small earthquakes each day. During well completion and in the years of extraction following, air pollution, ground water contamination, and wastewater disposal represent persistent environmental

[2]This phenomenon is referred to as the ecological inference fallacy in public health.

[3]The Denver Post highlights one such debate in the article: *Colorado regulators signal support for 2,000-foot buffers between fracking sites, homes.*

risks.[4] Compounding these localized impacts is the high dosage associated with common drilling of dozens or even hundreds of wells in concentrated clusters. Figure 2.1 shows the magnitude and timing of the fracking revolution in the US and the degree of residential exposure. In 2.1a we see that fracking greatly expanded the total amount of domestic extraction, becoming the dominant method of gas extraction by 2010. Figure 2.1b shows a quadrupling of the population living within 1 or 3 miles of an active well over this period, the latter threshold including approximately 40 million residents.



(a) Expansion of Fracking Extraction

(b) Population Exposure To Active Wells

Figure 2.1: Scale and Scope of the US Fracking Revolution

Note: Panel (a) shows the rapid expansion of fracking as a share of US oil and gas extraction. Solid lines show extraction via horizontal wells, a proxy for fracking. Dashed lines show extraction using conventional drilling methods. Panel (b) shows that the residential population exposed to nearby active wells (1 or 3 miles away) has expanded fourfold since fracking began. Population statistics are constructed in Khachiyan et al., 2021 and reflect similar estimates in Gold, 2014. Well data are sourced from Drillinginfo.

While the economics literature has identified a few margins of spatially concentrated negative fracking effects in Pennsylvania, the preponderance of work has studied county-level economic outcomes. Currie et al., 2017 show that in utero exposure to fracking within 1 kilometer leads to small decreases in infant birth weight and health index. Related work finds that the perception of ground water contamination risk devalues ground water dependant homes within 1 mile of a well by 13% (Muehlenbachs et al., 2015). This evidence of highly localized externalities contrasts the broader literature, which documents positive county-level impacts of fracking on employment,

[4]The Natural Resource Defense Council provides an excellent online resource reviewing the process and local environmental implications of fracking.

income and public revenue across many shale formations and states (Bartik et al., 2019; Feyrer et al., 2017). Of key relevance from this county-level literature is the result that more of the income generated by fracking extraction flows out of the source county than stays in it (Feyrer et al., 2017).

This spatial decoupling of the harms and benefits of fracking–narrowly concentrated residential hazards and wide dispersion of economic profits–characterizes an important gap in our understanding of fracking. County-level analyses pool neighborhoods with high and low direct exposure and thus may mask spatial inequality in the economic impacts from fracking. While a host county may experience economic growth as a result of this shock, residential quality could precipitously decline for the neighborhoods within the county which are inundated with drilling, truck traffic, and environmental risks.

I inform this puzzle by studying the mile-by-mile spatial gradient of fracking impacts on residential investment and population. This allows me to empirically estimate the spatial horizon of fracking impacts on residential areas and characterize the within-county distributional consequences. My sub-county approach further enables me to ask: i) what neighborhood characteristics are associated with near exposure to fracking?, ii) what mechanisms correlate with the effects of near and far exposure?, and iii) which neighborhood traits (if any) characterize heterogeneity in these effects? These questions build off of each other to directly address unfolding policy debates around how to regulate fracking.

I study these highly localized questions by applying newly developed outcome measures of microspatial residential investment and population based on daytime satellite imagery (Khachiyan et al., 2021). These data are generated by a prediction algorithm trained to identify neighborhood total income and population changes using medium resolution daytime satellite imagery. I interchangeably refer to the neighborhood level change in total residential income as residential investment. These outcomes accurately capture economic conditions in 2.2 square mile neighborhoods at a temporal scale otherwise unavailable. To account for endogenous conditions impacting local extraction amounts, I implement a spatially precise version of a commonly used instrument

measuring geological shale reserves (Bartik et al., 2019). By estimating a long-difference model centered on play-level first fracking year, my identifying variation is generated through the interaction of local adoption timing and neighborhood proximity to shale reserves.

My primary finding is a large and robust negative effect of proximity to drilling activity on population and residential investment of neighborhoods. Exposure to any instrumented extraction leads to a 2 percent decline in total income for neighborhoods up to 20 miles away.[5] A standard deviation increase in extraction within 3 miles leads to a relative decline in residential investment of 6.4 percentage points. These binary and continuous effects of neighborhood income mirror those of local population, in each case dissipating toward a null effect as the exposure threshold expands. Restricting this exercise to years in which Census data is available confirms that my source of outcome data does not drive these results. Instead, I show that my estimated negative effects deviate from existing county-level positive findings because of my more localized unit of analysis and extended post-shock period.

By investigating localized effects across all major US shale plays and incorporating data on employment, pollution, and housing, I am able to add economic context and interpretation to my primary findings. The leading mechanism consistent with residential investment declines in exposed neighborhoods is a marginally smaller decline in population, indicating that the extensive margin (depopulation) captures the first-order of observed changes. In contrast, other mechanisms such as direct air pollution and home prices are not strongly associated with the treatment. Treatment effects do vary substantially, however, based on prior employment specialization and environmental protection policies. The robust environmental regulatory environment in Pennsylvania relative to neighboring Ohio is associated with a complete attenuation of my measured negative exposure effects. Because existing evidence of economic effects from near exposure relies heavily on data from Pennsylvania, my finding of muted local impacts in Pennsylvania relative to other states indicates that the existing literature may be substantially understating the negative impacts of fracking on adjacent neighborhoods.

[5]Going forward I will refer to instrumented extraction simply as extraction and will discuss the non-instrumented value as raw extraction.

In perspective of the broad and well established literature studying the impacts of industrial development on the spatial distribution of economic activity and residential settlement, my work is tightly linked both by motivations and methods. Currie et al., 2015 and Barrows et al., 2019 offer examples of the local health and real-estate implications from plant openings in developed and developing countries. Work on Superfund sites (Kiel and Williams, 2007; Greenstone and Gallagher, 2008) and air pollution exposure (Currie and Walker, 2011; Smith and Huang, 1995; Chay and Greenstone, 2005) document the degree of associated neighborhood disamenities, and the implications for urban land use patterns (Duranton and Puga, 2015). Directly informing the social planners decision, a long tradition of work has attempted to balance the associated costs and benefits of such shocks into a local welfare analysis (Greenstone and Moretti, 2003; Bartik et al., 2019). The implications of the spatial distribution of labor on urban agglomeration and associated residential sorting underlies such exercises (Glaeser and Kahn, 2001; Moretti, 2004). Long-term consequences of local labor specialization in the context of mineral extraction are examined in Allcott and Keniston, 2017. Finally, my work expands on several recent papers applying machine learning techniques to imagery data for the study of localized economic shocks (Donaldson and Storeygard, 2016; Glaeser et al., 2018; Currie et al., 2020; Fowlie et al., 2019).

This paper proceeds in the following sections. Section 2.3 details each of the data sources and conducts interpretation exercises for my neighborhood outcome measures. I describe my empirical methods in Section 2.4. In Section 2.5 I present and review my primary regression results. Section 2.6 covers dimensions of exposure and the identifying assumptions. In Section 2.7 I examine mechanisms and margins of heterogeneity for my primary results. I conclude with a discussion of the results and implications for literature and policy in Section 2.8. Robustness results are presented and discussed in the Appendix.

2.3 Neighborhood Exposure and Outcomes Measurement

The central empirical innovation of this project is to collect data for both exposure and outcomes at the neighborhood rather than county-level. The primary challenge here is to measure an outcome which captures a general notion of residential well-being at this geographic scale. This is because the primary dimensions of exposure–proximity to well sites and shale reserves–are spatially precise and thus can be defined at any scale. My primary outcome of interest is a measure of the change in income of neighborhood residents ("residential investment") derived from satellite imagery in Khachiyan et al., 2021. As defined by the construction of this outcome measure, the geographic unit for my analysis will be 2.2mi^2 grid cells (square "neighborhoods" with side length 2.4km) which cover approximately 90% of the US population.[6]

2.3.1 Fracking Exposure Data

I measure gas and oil production using well-level data from the Drillinginfo platform of the firm Enverus. Using this dataset I export precise geographic coordinates of each well, coupled with annual oil and gas production amounts. These data represent a comprehensive accounting of onshore oil and gas extraction in the US. I construct my endogenous variable **Extraction** as a standardized aggregation of annual oil and gas produced,[7] scaled by inflation and the price of crude oil from the relevant year. I transform this variable into a standard normal distribution at each distance threshold so that coefficients can be interpreted as the impact of a standard deviation increase in produced value.

In addition to this extraction data, I am also interested in the spatial extent and quality of shale reserves for use as an instrument. Rystad Energy constructs these data based on detailed geological maps for each major play in the US. Rystad prospectivity maps aggregate relevant and

[6] Areas with very low population density in 2000 are excluded for reasons of computation and interpretation.

[7] Extraction amount is defined as Barrels of Oil Equivalent which is the sum of barrels of oil and natural gas. Natural gas is divided by 6 to convert from thousands of cubic feet of gas (Mcf) to a unit of energy approximately equivalent to a barrel of oil.

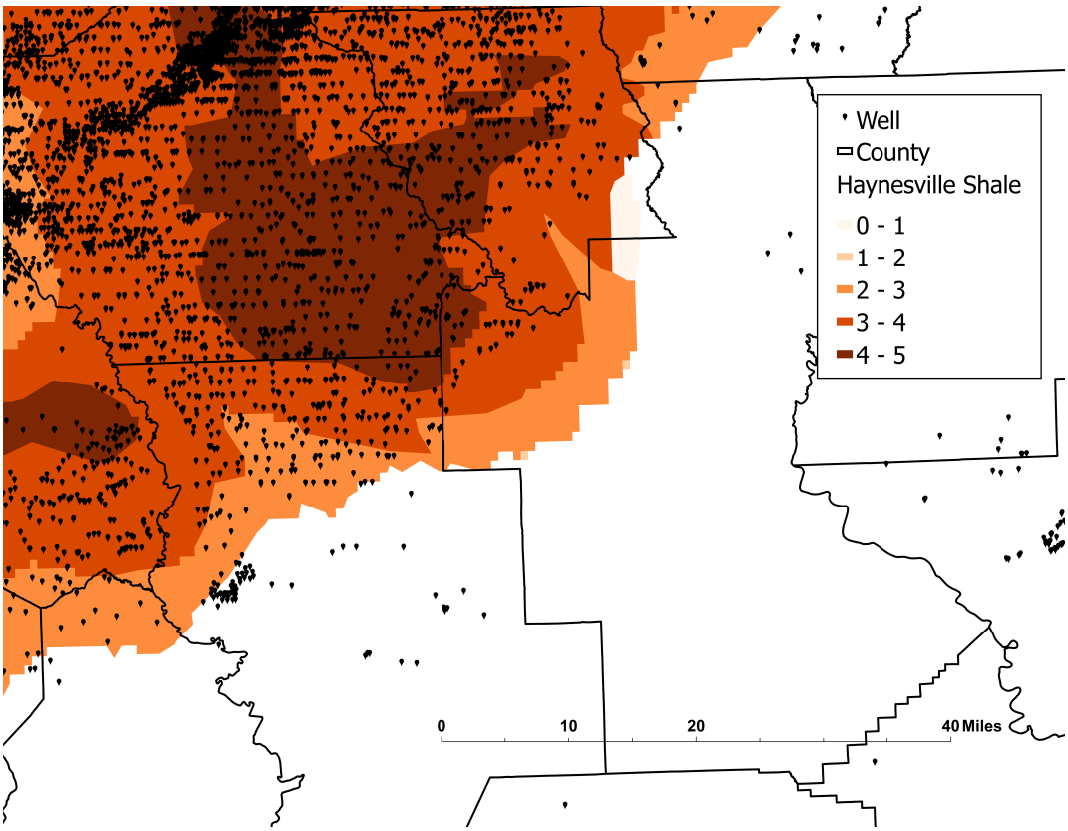


Figure 2.2: Sub-County Exposure to Wells and Shale

Note: The map shows that the degree of exposure to nearby fracking can be very different for neighborhoods on opposite ends of an exposed county. Pins show active individual well sites and orange shading shows the level of shale reserve prospectivity. County boundaries are shown in black. Drill locations are sourced from Drillinginfo and shale data is collected by Rystad Energy.

available geological data to geolocate the areas with the most potential for fracking.

The final piece of my fracking exposure definition relies on initiation dates from Bartik et al., 2019. These dates represent the first year that fracking begins in each shale play, and are based on a detailed analysis of local news and firm reporting on fracking activity. I use these dates to set the event year, with which I define the play-by-play event-window for my regression analysis.

The geographic scale of these extraction and shale data are shown in Figure 2.2 alongside county boundaries. This region of the Haynesville Shale Play in North West Louisiana showcases the important role of spatial resolution when measuring the impacts of fracking exposure. Two counties intersect the Southern edge of the shale play in the center of the image, both overlapping a

small portion of the highest prospectivity region of the play (the darkest shade of orange indicating prospectivity score above 4). Yet the scale bar shows that each of these counties stretches over 50 miles from corner to corner. As such, some neighborhoods are directly adjacent to an active fracking region while others are 50 miles away from one. This example illustrates how large and irregularly shaped county boundaries can undermine measurement of near effects of fracking using county-level outcomes.

2.3.2 Predicted Residential Investment

As outlined in Donaldson and Storeygard, 2016 the use of satellite imagery in economics presents three broad advantages: 1) information that is not easily accessible or may not exist otherwise, 2) high spatial resolution, and 3) global coverage. This has generated a wide range of economic applications of these data in realms such as national accounts, urban boundaries, natural disasters, deforestation, agriculture and economic growth. Improvements in sensor and satellite technology have driven steady advances in the spatial, spectral, and temporal resolution of these data. Coinciding with these advances, free access to expansive pre-processed data sets through platforms such as Google Earth Engine, as well as the computational techniques and hardware needed to analyze them have rapidly expanded the possible applications for economists. Starting with a single, discrete measure of nightlight intensity Henderson et al., 2012 are able to track economic growth at the national and regional scale across many countries. While this approach loses predictive power for smaller spatial units, the spatial resolution and multi-spectral nature of daytime satellite imagery offer an alternative.[8]

In Khachiyan et al., 2021 we advance this literature by training a convolutional neural network to predict the level of total income in a given image (neighborhood) and the difference in total income between two images. The benchmark for training is aggregate income of all residents at the block-level as measured by the decennial Census in 2000 and 2010. The input data

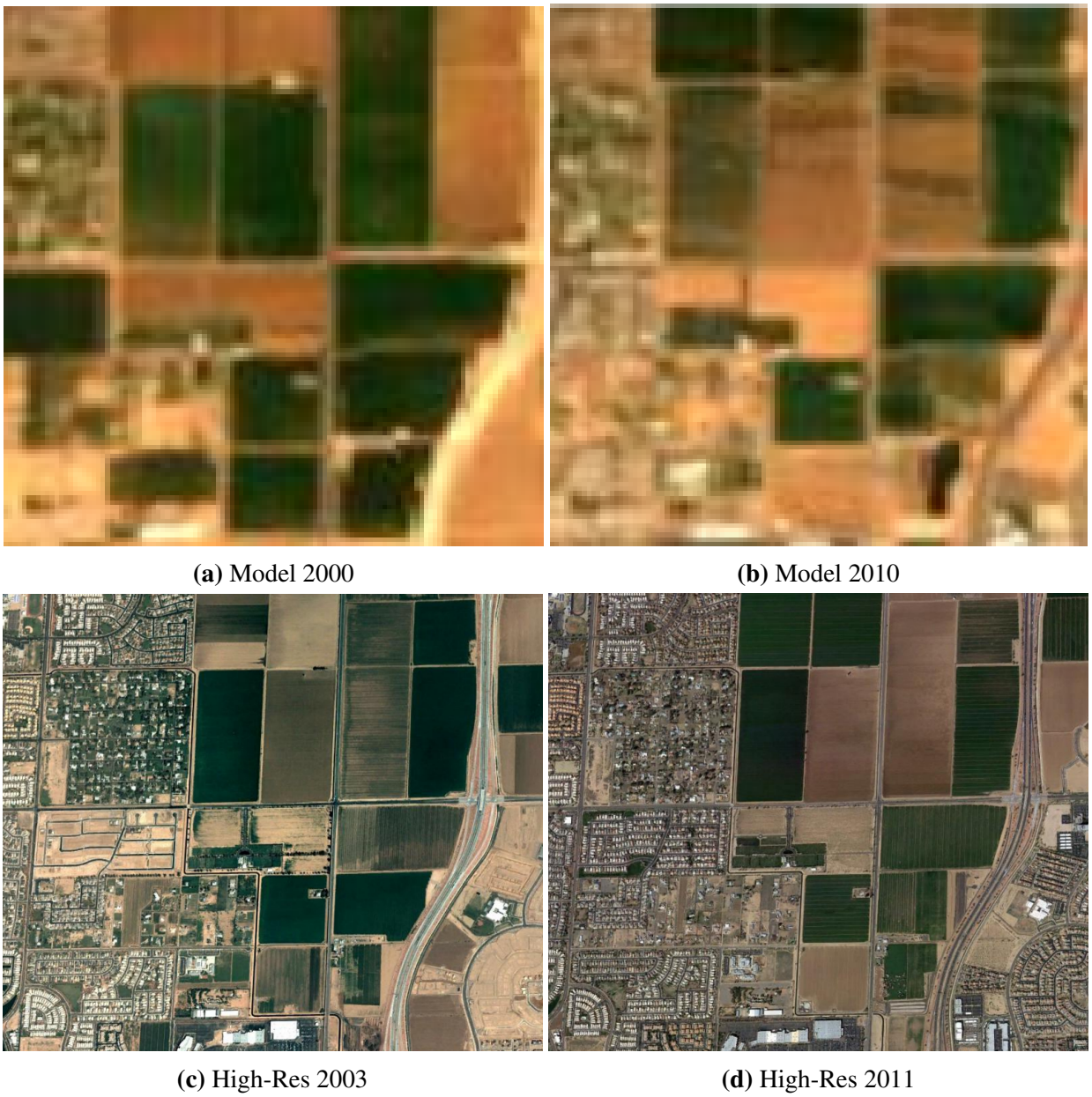
[8]Spatial resolution of satellite imagery measures the pixel size: traditional nightlight measures have pixels with side length of 1km while Landsat 7 daytime imagery have pixel size of 30m. For each pixel, daytime satellite images contain measurement of many different bands (variables) across the light spectrum, which is the spectral resolution.

used to form these predictions are annual composites of Landsat 7 satellite imagery from 2000 to 2019. Each annual image contains 6400 pixels (30m resolution) with continuous measurement of 8 spectral wavelengths (Red, Green, Blue, Infrared (3), Thermal, and Panchromatic). Convolutional Neural Networks are a powerful and proven Machine Learning tool for identifying latent patterns in image data. In a hold-out test set of the images, our model captures approximately 90% of income variation in a given year and 40% of the variation in changes from 2000 to 2010; this performance on predicting differences is more than twice as good as other benchmark methods such as nightlights at this geographic scale. We find very similar results when modeling neighborhood population. The usefulness of these models is that they can be used to predict local income or population changes between any pair of years with Landsat 7 imagery (2000 to present). Because the initiation year of fracking in each play ranges from 2005 to 2012 in my sample, this allows me to consistently measure the neighborhood change from several years before the initiation to 7 years after. Further details on the data used and the modelling exercise can be found in Khachiyan et al., 2021.

To build intuition on the imagery characteristics which drive these predictions, Figures 2.3 and 2.4 present example images and the associated model predictions. Figure 2.3 shows a rural-urban interface in Western Phoenix that represents one of the fastest growing parts of the nation over this time frame. Panels A and B show the Landsat images of this neighborhood that went into the CNN model. Panels C and D show higher-resolution Google earth images from the corresponding available years, to offer more clarity on the changes occurring in the medium-resolution Landsat images. In this example we see several large agricultural plots (notably along the Western edge) developed into residential neighborhoods over these 10 years.

Figure 2.4 shows the complementary images for an area in North Forth Worth which features high exposure to fracking in the Barnett play (initiated in 2000) and experiences a decline in local income. The decline in this case is characterized by an expansion of industrial infrastructure, which is visible as highly reflected roofs which span the distance of several city blocks. In this case the model over-predicts the degree of decline as –41% rather than the actual change of -23%

measure by the Census over this decade.



(a) Model 2000 **(b)** Model 2010

(c) High-Res 2003 **(d)** High-Res 2011

Figure 2.3: Positive Residential Investment in West Phoenix

Note: The above images show landcover changes in a West Phoenix neighborhood that are associated with positive changes in total income of residents. Panels a and b on the top show the 2000 to 2010 change using Landsat 7 30m resolution imagery; this is the image product used to construct my outcome measure of predicted income change. Panels c and d on the bottom show the same area in the nearest years with high-resolution google earth imagery; this offers more insight on the detailed neighborhood changes that the model is picking up on. In this case that is the replacement of agricultural fields with neighborhoods, particularly in the bottom left quadrant of the neighborhood. The model predicts an income change of 99% over this period; the true income change is 101%. A long run satellite timelapse of this area can be found here.

Because this measure is being used as an outcome in my regression analysis, classical mea-

(a) Model 2000 **(b)** Model 2010

(c) High-Res 2003 **(d)** High-Res 2011

Figure 2.4: Residential Decline in North Fort Worth

Note: The above images show landcover changes in a North Fort Worth neighborhood that are associated with negative changes in total income of residents. Panels a and b on the top show the 2000 to 2010 change using Landsat 7 30m resolution imagery; this is the image product used to construct my outcome measure of predicted income change. Panels c and d on the bottom show the same area in the nearest years with high-resolution google earth imagery; this offers more insight on the detailed neighborhood changes that the model is picking up on. In this case I observe an expansion of bright industrial rooftops which corresponds to growth in oil services firms in this area. The model predicts an income change of -41% over this period; the true income change is -23%. A long run satellite timelapse of this area can be found here.

surement error (i.e. uncorrelated with the causal relationship of interest) will not bias the results. To mitigate this concern, we show in Khachiyan et al., 2021 that the prediction errors are only weakly correlated with image-level demographic characteristics and geographic fixed effects. Figure 2.5 further shows the correlates of predicted residential investment relative to the correlations with census measured income. Confirming the intuition that this measure is based on visible land-cover changes associated with residential settlement, we see a strong association with changes in home values and residential construction and weaker associations with local government accounts. I expect this is due to the fact that variation in government accounts is high and tightly linked to business cycle fluctuations, unlike physical infrastructure which often persists for decades.

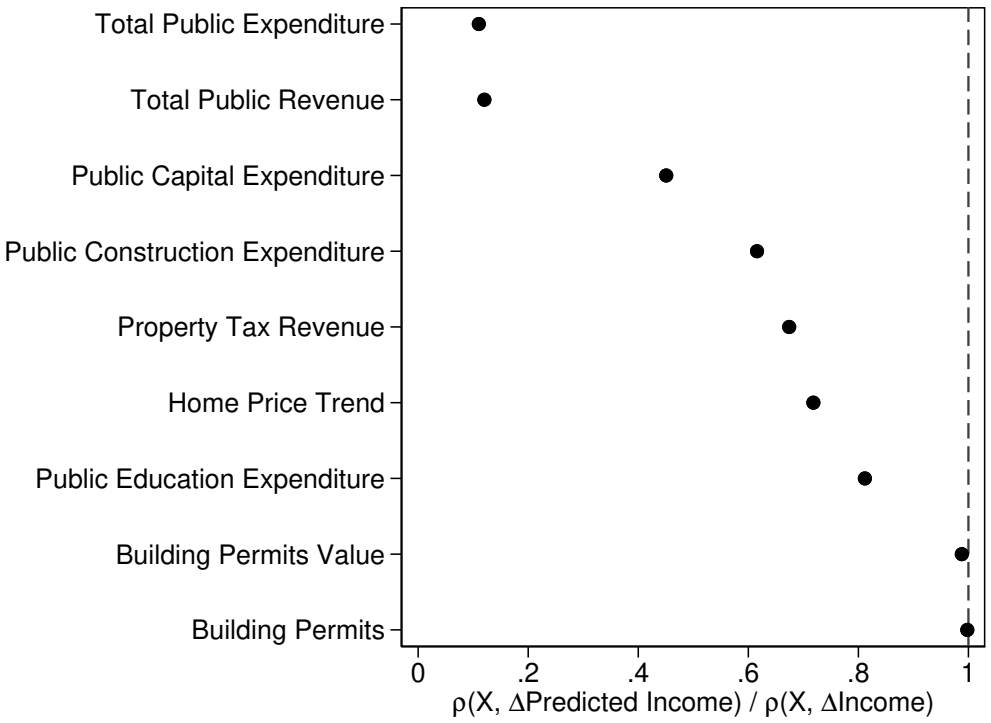


Figure 2.5: Correlates of Predicted Income

Note: The graph shows that my predicted measure of local income change and the census measure are equally correlated with residential infrastructure covariates, while the census more accurately captures local public finance accounts. Each covariate on the y-axis is a county-level measure of neighborhood changes from 2000 to 2010. The x-axis show the correlations between each of these and my predicted measure of income change over this period, scaled by the corresponding correlation with census measured income change.

Table 2.4 further shows that at the image-level, the satellite-based measure of residential

investment yields an almost identical treatment effect as Census-measured income, restricted to the 2000 to 2010 sample period. The details of the model in Table 2.4 are expanded on in Section 2.4. This contrasts with more positive impacts using Census-measured income at the aggregated county-level, which are similar to results found in the literature at this scale.

2.3.3 Neighborhood Exposure Assignment and Sample Definition

To construct my analysis sample I measure extraction and shale characteristics at the neighborhood level. Each neighborhood is defined based on the grid of 2.2mi^2 images used to generate my remotely sensed neighborhood income measure. Using the distance between the centroid of each image and all active wells from 2000 to 2019 within 100 miles I compute annual extracted value in distance bins from 1 mile to 100 miles. Assigning neighborhood event year and exposure to a shale play requires assignment to the nearest play. Each neighborhood is assigned to the top quartile play[9] which is closest. The only exception to this assignment is the case in which another nearby play starts in an earlier year: if the neighborhood is less than 10 miles farther from the earlier top quartile play then it is assigned to that one. A visual representation of this assignment is given in Figure 2.6, which shows a map of the national sample of neighborhoods color-coded by play assignment on top of shale prospectivity maps. This play assignment establishes the event year for each neighborhood and allows me to measure the measure of total exposure to the assigned play for the same expanding distance bins as used with extraction.

My analysis sample is thus restricted first based on being exposed to a shale play within 100 miles. I further restrict the sample to neighborhoods with predicted income and population values between the 1st and 99th percentile in all years and prediction errors within the same percentiles in 2010. Finally I exclude images which are part of the Northeast Corridor, a large contiguous urban area in which many large East Coast cities are concentrated. Robustness results in the Appendix show that these sample inclusion criteria do not alter the baseline regression results, they merely add structure around the interpretation of findings and rule out influences by outliers of the

[9]The top quartile of a play is the geographic area of the play which contains the top quartile of its shale prospectivity value as measured by Rystad.

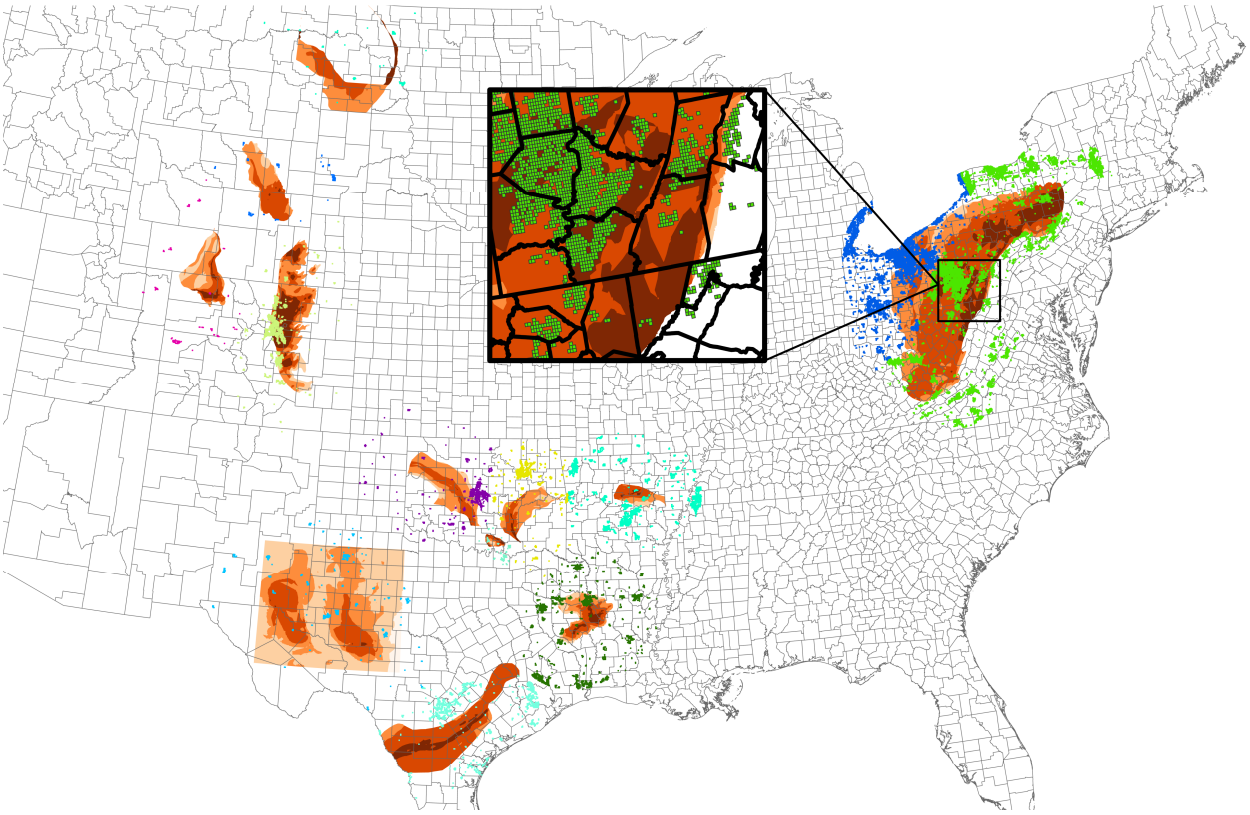


Figure 2.6: Analysis Sample of Neighborhoods Over Plays and Counties

Note: The map shows my analysis sample of neighborhoods as squares overlaying shale play and county boundaries. Images are color-coded by play assignment and plays are shaded such that the highest prospectivity parts are darkest. The pop-out panel shows a magnified perspective on the scale of each dataset in Western Pennsylvania.

outcomes. The robustness results discussed in the Appendix also show that further restrictions such as excluding neighborhoods near multiple plays or those with active wells inside their boundaries also do not change the baseline results.

2.3.4 Secondary Data Sources

To investigate the composition of compliers and conduct mechanisms and heterogeneity analysis, I attribute several additional variables to each neighborhood. Baseline demographic data from 2000 are assigned to images based on geographic overlap with Census Blocks. These variables include shares of racial groups, population density, and income per capita. To investigate the pollution channel of my measured treatment effects, I incorporate local PM2.5 estimates con-

structed using a combination of remote sensing and spatial modelling in Di et al., 2019. Annual employment statistics across the 20 NAICS sectors and overall are collected from the block-level Census LODES Residence Area Characteristics. Data on local home price trends are collected from the FHFA Census Tract-level annual home price index. For interpretation exercises I also study county-level data on public finance accounts collected in The Government Finance Database by Willamette University Atkinson Graduate School of Management. I finally incorporate building permit data from the Census Residential Building Permits Survey, also as county-level interpretation covariates.

2.4 Empirical Strategy

My analysis is focused on the distribution of long-term fracking impacts across neighborhoods. A long difference specification allows me to exploit the spatial precision of my data. In particular a long difference is more parsimonious than the natural alternative of an event-study model with by-year observations, interactions and fixed effects. By estimating fewer parameters and focusing on long-term and generally more statistically distinguishable effects, the long difference approach retains sufficient identifying variation and degrees of freedom in the estimating equation to investigate detailed heterogeneity across spatial units.

$$\Delta log(Income)_{i,t} = \beta \widehat{Extraction}_{i,d,t} + \alpha_{play_i} + \varepsilon_{i,t} \qquad (IV)$$

$$Extraction_{i,d,t} = \omega Prospectivity_{i,d} + \rho_{play_i} + \nu_{i,d,t} \qquad (FS)$$

To set the relevant time-frame for this long difference analysis, I directly apply the findings of Bartik et al., 2019 on the first year of fracking in each play. For each neighborhood (i), event year 0 is set as this first year of fracking, and regression variables are calculated for years relative to this start year. In the instrumental variables regression (IV), long-run residential investment is defined as the average of log total income in event years 5 through 7 minus the same average for

event years -3 through -1. This averaging approach mitigates the impact of any noisy predictions in a single year and reduces the variance in estimated treatment effects.

Endogenous treatment exposure is similarly defined as the total value of barrels of oil equivalent extracted between event years -1 and 7. Successive models estimate the impact of this treatment at expanding spatial horizons (d) of 1, 3, 10, 20, 40, 60, 80 and 100 miles. To simplify the interpretation of this treatment, I normalize extraction value within each distance horizons by play-level mean and standard deviation. In this sense, the causal interpretation of a 0.01 coefficient on "Extraction 10mi" would be that a 1SD increase in extraction value within 10 miles leads to 1 percentage point higher residential investment on average. Using raw extraction directly as the exposure measure (OLS) may lead to endogeneity bias. Of particular concern is that the amount of oil extracted in a given location may be closely tied to endogenous economic conditions. For example, areas with a history of conventional oil development may offer a higher supply of the capital and labor necessary for fracking than those without. If these places also exhibited a relative positive trend in outcomes over this period, this would create positive bias in my estimated treatment effects.

Hence, inspired by the instrumental variables models in Feyrer et al., 2017 and Bartik et al., 2019, I use the geographic distribution of underlying shale reserves as an instrument for nearby oil and gas extraction. For each neighborhood, this instrument is computed as the total prospectivity of the assigned play within the distance (d) corresponding to the distance threshold for the Extraction in a given model. For example if the endogenous regression examines the residential investment associated with extraction within 10 miles, I use the sum of shale prospectivity within 10 miles as the instrument for extraction within 10 miles. Equation (FS) shows this first stage regression. As discussed in Section 2.4, these prospectivity data measure the geologic distribution and intensity of shale reserves. The results in Section 2.5 show that this instrument is highly relevant for fracking extraction. The argument for validity is that these shale reserves accumulate over millennia and are only detectable on the land surface with specialized prospecting equipment. Therefor the instrument should have no impact on local economic conditions aside from the direct impact on where

fracking takes place. Further details and results on accessing the IV assumptions in this empirical setting can be found in Section 2.6.

It is necessary to use play fixed effects in this IV approach because the measure of prospectivity is only ordinal within each play, not across plays. Including these play fixed effects also adjusts treatment and outcome means for fixed differences by play: namely any discrepancies in outcome measurement across regions and business cycle differences based on initiation year, region, or both. A discussion of the trade-offs of using more spatially narrow fixed effects (i.e. state or county) can be found in the Appendix robustness discussion.

In spatial analyses like this one, correlated errors between nearby observations can lead to a substantial underestimation of treatment effect variance(Conley, 2010). To account for this I apply the simplest correction technique, standard errors in all regressions are clustered at the county level. Figure 2.10 shows that this approach is more conservative than using Conley standard errors with arbitrary spatial correlation within any specified distance horizon. Each line represents the standard error on the impact of 1SD increase at the 10 mile threshold in the IV income change models presented in Table 2.1 and discussed in Section 2.5. In particular the solid line shows how the standard error changes as the spatial correlation threshold increases: expanding the threshold increases the variance of the estimator until 50 miles. Accounting for spatial correlation beyond 50 miles only leads to modest shrinkage in the calculated error. Beyond offering a conservative estimate, county clustered standard errors also require a fraction of the computational power and are widely used and understood in economics.

2.5 Baseline Impacts of Nearby Fracking

In this section I present and discuss my baseline regression results measuring the absolute effect of any fracking exposure and relative effect of more nearby extraction. I also present related event-study graphs which show the time-path of these impacts. Section 2.6 further accesses and interprets these results by testing underlying assumptions, characterizing the groups most exposed

to this shock, and decomposing the contributions of scale and my outcome measure. Section 2.7 goes on to explore the mechanisms and margins of heterogeneity underlying these baseline effects.

2.5.1 Relative Impacts of Continuous Exposure Amount

Table 2.1 displays coefficients on the instrumented measure of total hydrocarbon production. Columns labelled “IV” show the IV results of estimating equations (IV) and (FS), while “OLS” columns give the results for the equivalent model without an instrument. The final column displays the first stage coefficient of the IV models. Results are given for three outcomes: change in neighborhood income, change in population, and change in income per capita. In this table, each descending row shows results using a wider bandwidth (d). Taking as an example the first entry in the "Income IV" column, I find that a 1 standard deviation (SD) increase in extraction value within 1 mile leads to a 10 percentage point (pp) reduction in the long-run income change. This effect shrinks as the bandwidth increases, with a SD more extraction within 100-miles yielding a -2.6pp treatment effect.

Additional columns showing results on population and income per capita should be thought of as the first mechanism results. Here I find that population results are similar but smaller in magnitude than the total income results; this leads to income-per-capita effects which remain negative but are much smaller and lack statistical power. I interpret this comparison across outcomes to indicate that the negative results on total income are primarily capturing a slowdown in the extent of residential coverage in a neighborhood, rather than an intensive deterioration of existing neighborhoods. Given that it is challenging for the satellite based outcome measure of neighborhood income change to identify negative intensive margin changes (Khachiyan et al., 2021), in this sense the results may be considered an upper bound on the full magnitude of negative neighborhood impacts.

Table 2.1: IV And OLS Baseline Impacts by Exposure Distance

	Δ Income		Δ Population		Δ Income/Population		First
	OLS	IV	OLS	IV	OLS	IV	Stage
Extraction 1mi	0.002	-0.102***	-0.003	-0.080***	0.005**	-0.022	0.184***
	(0.002)	(0.033)	(0.002)	(0.029)	(0.002)	(0.016)	(0.033)
Extraction 3mi	-0.004	-0.064***	-0.007**	-0.052***	0.003*	-0.012	0.327***
	(0.004)	(0.019)	(0.003)	(0.017)	(0.002)	(0.009)	(0.050)
Extraction 10mi	-0.005	-0.048***	-0.009**	-0.038***	0.003*	-0.010	0.464***
	(0.004)	(0.014)	(0.004)	(0.013)	(0.002)	(0.006)	(0.065)
Extraction 20mi	-0.007*	-0.041***	-0.010**	-0.032***	0.003	-0.009*	0.577***
	(0.004)	(0.011)	(0.004)	(0.010)	(0.002)	(0.005)	(0.064)
Extraction 40mi	-0.013***	-0.037***	-0.013***	-0.030***	-0.000	-0.007*	0.707***
	(0.004)	(0.009)	(0.004)	(0.009)	(0.002)	(0.004)	(0.053)
Extraction 60mi	-0.020***	-0.035***	-0.020***	-0.029***	-0.000	-0.006	0.773***
	(0.004)	(0.008)	(0.004)	(0.008)	(0.002)	(0.004)	(0.046)
Extraction 80mi	-0.023***	-0.031***	-0.023***	-0.026***	-0.000	-0.005	0.828***
	(0.005)	(0.007)	(0.005)	(0.007)	(0.002)	(0.004)	(0.045)
Extraction 100mi	-0.023***	-0.026***	-0.021***	-0.023***	-0.002	-0.004	0.897***
	(0.005)	(0.007)	(0.005)	(0.006)	(0.003)	(0.003)	(0.039)
Observations	18,202	18,202	18,202	18,202	18,202	18,202	18,202

Note: The table shows my baseline regression result of negative fracking impact on total income concentrated in neighborhoods most exposed to extraction. These effects are driven by a reduction in population rather than a major reduction in income per capita. Coefficients show the impact of a 1 standard deviation increase in the value of oil and gas extraction within a set radius of each neighborhood in the sample, as specified in equation 1. Outcomes are long differences of log variables designated in column headers. The first stage coefficient for all IV models is shown in the final column. Each row shows impacts of extraction at increasing distance thresholds. Coefficients should be interpreted as the percentage point change in outcome resulting from a 1 SD increase in extracted value. Standard errors clustered by County are shown in parentheses; *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

2.5.2 Absolute Impacts of Binary Exposure

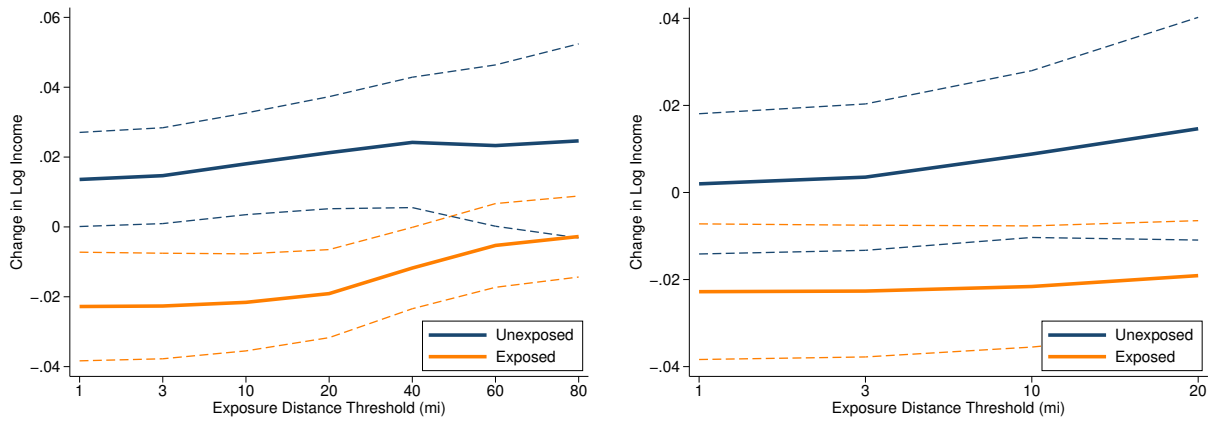
The spatial gradient of impacts presented in Table 2.1 shows the continuous effects of a standard deviation increase in extraction as the radius of inclusion expands from 1 mile to 100 miles. Perhaps more relevant for policy interpretation is a binary version of this spatial gradient, which is shown in Figure 2.7.[10] In these graphs the exposed group at each distance threshold is defined as having an instrumented extraction value greater than 0, and the unexposed group comprises all neighborhoods that have no instrumented exposure at that distance threshold.[11] Panel (a) shows that neighborhoods exposed to nearby fracking extraction (i.e. within 20 miles) exhibit a statistically significant 2 percentage point residential decline, while those without near exposure see 2 percentage points of investment on average. The negative impacts of exposure dissipate approximately linearly to 0 as the threshold for exposure increases above 20 miles. Conversely, the positive impacts of not being exposed remain constant and become less precise as the sample size decreases at higher cutoffs. Panel (b) specifically shows that if the unexposed group is limited to be neighborhoods with exposure within 40 miles, i.e. within the approximate size of a county, the positive effects of being unexposed are largely erased. In particular, neighborhoods that have exposure within 40 miles but none within 3 miles exhibit no residential growth on average, and do not show statistically different outcomes when compared to those with this level of near exposure.

2.5.3 Time Path of Impacts

In order to investigate when the impacts of fracking are realized in exposed neighborhoods I implement the natural alternative to a long difference IV specification, an event-study. While the event study does not maximize the estimation power for the spatial gradient and heterogeneity analyses I focus on with the IV specification, the event study offers an intuitive way to think about

[10] Results in these figures are based on regressions of change in log income on instrumented exposure interacted with the binary exposed/unexposed group definitions. Play fixed effects and county clustered errors are applied, as in the main long-difference specification.

[11] Note that the sample is limited to neighborhoods within 100 miles of a play, which is why the highest comparison in these graphs is exposure within 80 miles vs no exposure within 80 miles (i.e. exposure only within 100 miles).



(a) Within 100 miles

(b) Within 40 miles

Figure 2.7: Absolute Residential Investment Impacts by Binary Exposure Distance

Note: The graphs show that neighborhoods exposed to instrumented extraction within 20 miles experience residential decline of approximately 2 percentage points over the course of the shock, while those without exposure within 20 miles see growth of about 2 percentage points. Lines show coefficients and 95% confidence intervals associated with treatment and control groups based on the binary treatment of having positive instrumented exposure within each distance threshold on the x-axis. Section 2.5.2 and footnotes 10 and 11 discuss details of the underlying regression model.

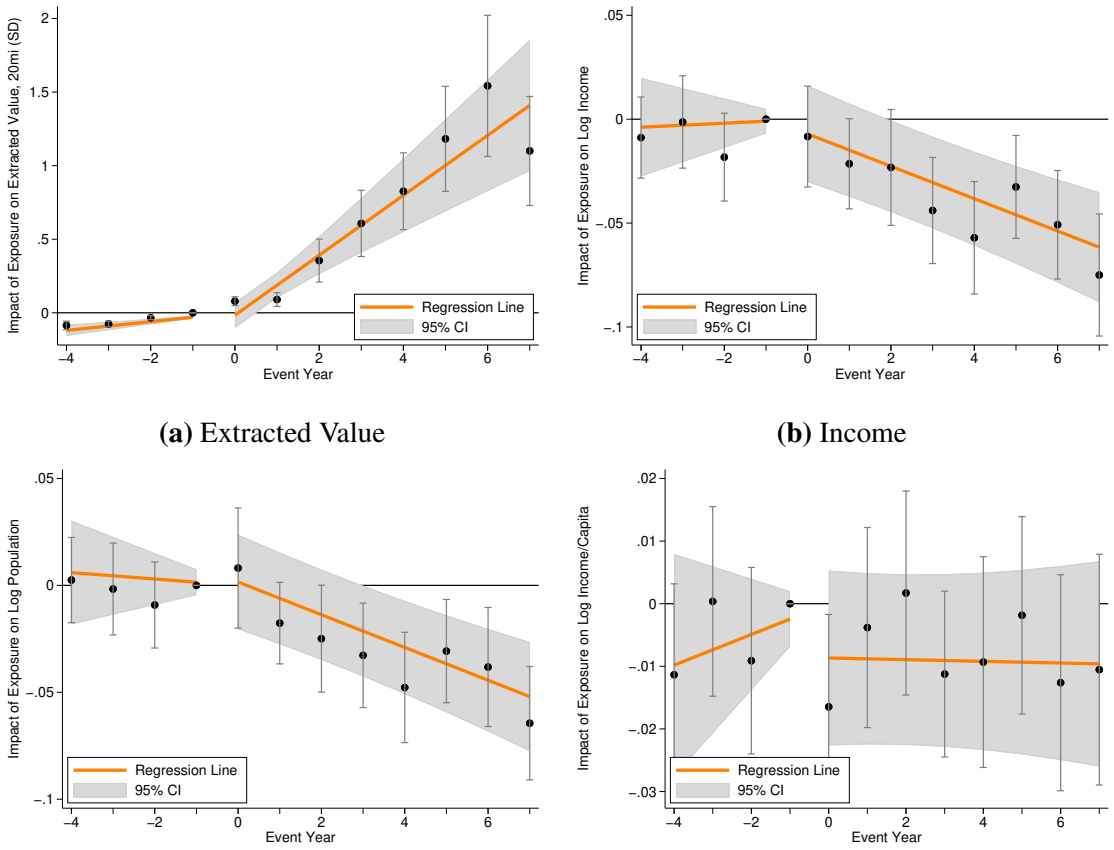
how trends diverge between treatment and control groups on a year-by-year basis. To conduct this analysis, I define the exposed group as all neighborhoods within 20 miles of shale reserves and the control as those with shale between 20 and 100 miles.[12] Figure 2.8 shows the results of event study regression using this exposure definition on the first stage of extraction and across the primary outcomes of total income, population and income per capita. These figures show both annual coefficients associated with exposure and a model of pre and post linear trends, in all cases using event year -1 as the reference year.[13] This approach is similar to the primary empirical strategy used in Bartik et al., 2019, with adaptations to my granular data setting.[14]

Figure 2.8 broadly shows results consistent with those in the IV specification. Given the differences between these approaches and the assumptions for unbiasedness, this finding gives ad-

[12]Note that 100 mile proximity to a shale play is also the inclusion criteria in the long-difference analysis.

[13]These models use neighborhood and play-by-year fixed effects and cluster standard errors by county. Dots show coefficients and 95% confidence intervals on an annual interaction model, while the orange line and shaded region show the coefficients from continuous interactions of event year with pre and post periods, controlling for main effects of event year and period.

[14]A key difference with the event-study approach is the assumptions necessary for identification, relative to the IV strategy. Broadly, unbiasedness in the event-study requires the assumptions of a difference-in-differences model, primarily that the measured effect would have been 0 had the shock not occurred.



(a) Extracted Value

(b) Income

(c) Population

(d) Income Per Capita

Figure 2.8: Event-Study Analysis of Fracking Impact Timing

Note: The graphs show evidence of a strong first stage increase in extraction and approximately linear negative impacts on total income and population persisting into year 7 of the rollout. Each panel shows the regression lines, annual coefficients, and 95% confidence intervals corresponding to event studies described in Section 2.5.3. Event year 0 is defined as the first year of fracking in each play and the baseline is set to event year -1 in all regressions. Standard errors are clustered at the county-level.

ditional confidence that the treatment effects I capture accurately represent the true causal effects. An insight from this figure is that observed effects on total income accumulate over time: effects continue to become more negative each year. This offers an additional piece of evidence explaining why my results are more negative than those in the existing literature, given that existing work focuses on effects up to 4 years after fracking starts. Specifically, panel (a) shows a clear positive spike in extraction (normalized and shown here again measured in standard deviations) starting in the impact year. While there is a slight positive trend in extraction in the years before fracking (perhaps from exploration activities), there are no such pre-trends observed in the outcomes of income

and population change. Finally, the income and population models show a total effect after 7 years of approximately -6%, which is similar to the -3% to -4% effects observed in the post period of the IV specification at the 20 mile threshold. Finally we observe here, again as with the IV models, little evidence of precisely estimated or economically meaningful effects on neighborhood income per capita.

2.6 Assessment and Interpretation of Baseline Results

I proceed with several exercises designed to investigate which groups are exposed to the baseline treatment effects studied in Section 2.5, how the instrument is operating, and to what extent the geographic scale and outcome measure are contributing to the results.

2.6.1 Assessing Relevance and Validity of the Instrument

To further interpret these instrumental variables results in Table 2.1 and Figures 2.7 and 2.8, it is necessary to access how well the shale prospectivity instrument satisfies the necessary assumptions for unbiased estimation. To assess relevance of the instrument, coefficients in the "First Stage" column of Table 2.1 show that the total geological prospectivity of shale at each distance threshold has a large and precisely measured association with total extraction in the same area. Specifically at the 10 mile bandwidth, a 1-SD increase in total shale prospectivity is associated with a 0.46 SD increase in extraction. This strong first stage across bandwidths establishes the relevance of the instrument. Reassuringly, the sign of the coefficients are all positive, confirming that increases in shale prospectivity are indeed associated with more oil and gas extraction.

To investigate validity in this setting–the exclusion restriction–Table 2.2 shows a counterfactual exercise on the pre-period. In particular, extraction and outcome changes are calculated from Event Year -4 to -1. If the effects observed in Table 2.1 accurately capture the impact of the boom in oil/gas extraction from fracking, then we should not see any meaningful pre-trends in local economic conditions using the same specification. Table 2.2 confirms this finding; I observe

only very small (i.e. 1pp) and marginally significant pre-trends for income effects at 40 miles and higher. Notable, the effects for smaller bandwidths are indistinguishable from 0, and the sign of all income effects in the pre-period is positive, again suggesting that the primary regressions in Table 2.1 may understate the true negative impacts. There are some marginal increases in mineral extraction employment in the pre-period; one explanation may be the geological exploration activities preceding the commencement of production in Event Year 0.

Another way to investigate validity is to conduct an overidentification test using multiple instruments for the endogenous variable of actual extraction. To do this, I add an additional instrument which is the distance from the neighborhood to the top quartile of shale in the assigned play. This variable is studied in Bartik et al., 2019 as both a binary distinction of exposure and a continuous measure for IV regressions. It captures a simple distance measure of how close the neighborhood is to the highest prospectivity (and consequently most active) part of each shale play. While I focus on my continuous measure of prospectivity within set distances as a more precise local measure, the distance to the highest quartile of prospectivity provides a useful alternative instrument to consider and test for overidentification. Taking the IV specification in column 2 row 4 of Table 2.1 on change in total income caused by a 1 standard deviation increase in drilling within 20 miles, adding distance to top quartile shale as a secondary instrument results in a p-value of 0.18 associated with the Hansen J-statistic testing overidentification. This indicates that I am unable to reject the null hypothesis that both instruments are valid, yielding additional confidence that the primary instrument is valid. Furthermore, to the extent that effect sizes using only this secondary instrument (distance to the top quartile of shale) differ from those on the primary instrument, they are larger. This is a further indication that the effect sizes I report in my baseline results may be a conservative accounting of the true negative impacts of nearby fracking.

2.6.2 Exposed Groups and OLS-IV Results Gap

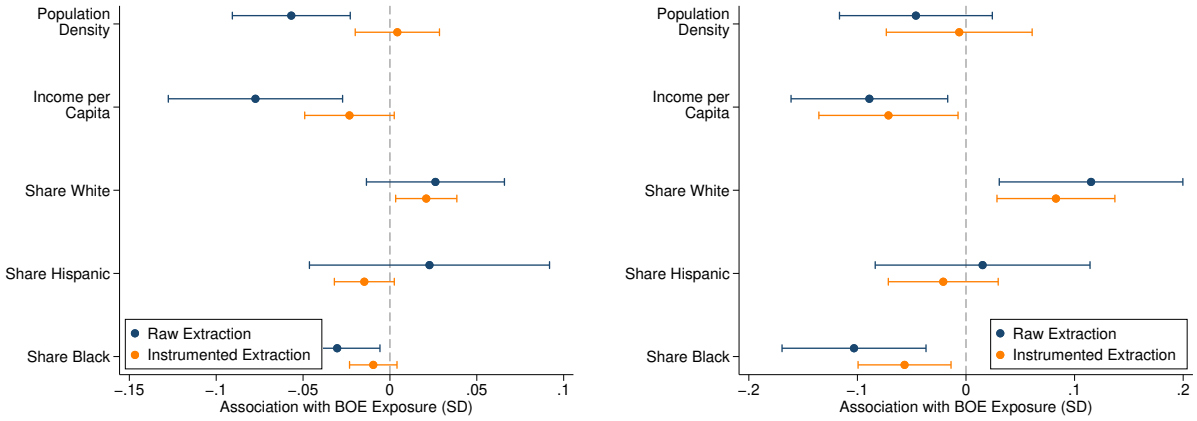
In order to interpret these baseline IV effects, it is necessary to study both which groups are most impacted by the exposure captured by the instrument and how this exposure is different

Table 2.2: IV Impacts of Extraction in Pre-Period

	Income	Total Emp	Mineral Emp
Extraction 1mi	0.030	-0.029	0.107
	(0.045)	(0.047)	(0.070)
Extraction 3mi	0.019	-0.024	0.071*
	(0.027)	(0.030)	(0.042)
Extraction 10mi	0.012	-0.016	0.041*
	(0.014)	(0.017)	(0.023)
Extraction 20mi	0.011	-0.013	0.030*
	(0.009)	(0.012)	(0.017)
Extraction 40mi	0.014*	-0.012	0.022
	(0.007)	(0.010)	(0.014)
Extraction 60mi	0.015**	-0.012	0.019
	(0.007)	(0.010)	(0.014)
Extraction 80mi	0.015**	-0.012	0.016
	(0.007)	(0.010)	(0.014)
Extraction 100mi	0.014*	-0.012	0.013
	(0.008)	(0.011)	(0.015)
Observations	18,202	16,504	16,504

Note: This table shows that there is little evidence of a counterfactual treatment effect occurring in the years before fracking started in each play, building confidence in the validity of the instrument. Coefficients show the impact of a 1 standard deviation increase in the value of oil and gas extraction within a set radius of each neighborhood in the sample, as specified in equation 1 but only over the pre-period. Outcomes are long differences of log income from 2000 to 2010. Each row shows impacts of extraction at increasing distance thresholds. Coefficients should be interpreted as the percent change in outcome resulting from a 1 SD increase in extraction. Coefficients should be interpreted as the percentage point change in outcome resulting from a 1 SD increase in extracted value. Standard errors clustered by County are shown in parentheses; *, ** and *** indicate significance at 10%, 5% and 1% levels respectively.

from raw extraction exposure. Figure 2.9 show how raw and instrumented exposure is associated with five demographic measures of interest. Panel (a) shows the associations with near exposure, within 3 miles, and Panel (b) for all exposure within 100 miles. The first notable result in these figures is that neighborhoods with more Non-Hispanic White residents, and those with lower income residents, face elevated exposure to the variation I study (Instrumented Extraction) relative to other racial groups and high income individuals. Because my results link exposure to negative neighborhood outcomes, this finding indicates that the fracking shock exacerbates local income inequality through the channel of reducing growth in low-income white neighborhoods.



(a) 3 miles **(b)** 100 miles

Figure 2.9: Dimensions of Exposure to Extraction

Note: The graphs show that neighborhoods with a higher share of White residents tend to be exposed to more fracking, and that the instrument focuses on near exposure among wealthier and more dense neighborhoods than exposure to raw extraction. Each bar and line show the coefficient and 95% confidence intervals on the continuous neighborhood characteristics shown on the y-axis, in a regression with either raw or instrumented extraction as the outcome.

A second insight of these exposure statistics is the difference between exposure to raw extraction and exposure to instrumented extraction. Notably, I find that the nearby exposure (within 3 miles) to instrumented extraction is concentrated in more dense and wealthier areas, relative to nearby exposure to raw extraction. This finding yields insight on the population of compliers in the IV analysis. Because IV regression estimates a treatment effect which is local to compliers–neighborhoods which increase extraction because of their proximity to shale–understanding who the compliers are motivates an interpretation of the difference between OLS and IV results in

Table 2.1. Table 2.3 tests this interpretation that the difference in OLS and IV results is driven by the local nature of the IV treatment effect. Columns (1) and (5) show the baseline OLS and IV results respectively (as shown in Table 2.1). Columns (2) through (4) augment the OLS results by narrowing the sample to subsets of compliers, first based on exposure to extraction prior to the rollout of fracking, then based on high population density and income per capita. Taken together these restriction show in column (4) that the OLS results at near exposure levels do approach the IV results once narrowing to the sample of compliers. While the coefficients in columns (4) and (5) maintain some quantitative differences, the direction and magnitudes of effect are qualitatively similar. This finding motivates the interpretation that the primary difference between my baseline OLS and IV results is the subset of variation on compliers which the instrument captures.[15]

2.6.3 The Role of Geographic Scale and Outcome Measure in Estimated Effects

The broad insight from my baseline long difference results is that more fracking leads to less residential investment. A natural question is: why do I find the opposite of the positive findings in existing leading work on this topic (Feyrer et al., 2017, Bartik et al., 2019)? Table 2.4 addresses this question empirically by comparing IV income change results using different data scale and outcome measure. The outcome measure I have which most closely resembles those used in the existing literature is Census measured total income of residents (the target used for the prediction models in Khachiyan et al., 2021). This outcome is only available from years 2000 and 2010, so I augment the analysis period for both oil extraction and income changes to be from 2000 to 2010.[16] In the first two columns of Table 2.4 we observe the treatment effects of extraction with county-level aggregated data on the outcome, endogenous and instrumental variables. Both the Census measure of income over this period and the satellite imagery based measure generate very

[15] I find no evidence in my sample of omitted variable bias on observable neighborhood characteristics, but it is certainly possible that bias on unobservables also contributes to the OLS-IV gap in estimated treatment effects.

[16] This also necessitates narrowing the sample of neighborhoods to those exposed to plays beginning production early on in the overall sample period. Namely, the sample for this table includes neighborhoods exposed to the Barnett, Permian, Fayetteville, and Woodford plays.

Table 2.3: Baseline OLS Regression on Subsets of Compliers Compared to IV

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	IV (5)
Extraction 1mi	0.002	0.011	0.000	-0.023*	-0.102***
	(0.002)	(0.008)	(0.005)	(0.012)	(0.033)
Extraction 3mi	-0.004	0.008	-0.012	-0.048**	-0.064***
	(0.004)	(0.011)	(0.008)	(0.020)	(0.019)
Extraction 10mi	-0.005	0.001	-0.009	-0.069***	-0.048***
	(0.004)	(0.011)	(0.008)	(0.020)	(0.014)
Extraction 20mi	-0.007*	-0.000	-0.009	-0.048**	-0.041***
	(0.004)	(0.012)	(0.009)	(0.023)	(0.011)
Extraction 40mi	-0.013***	-0.018***	-0.017**	-0.031**	-0.037***
	(0.004)	(0.006)	(0.007)	(0.014)	(0.009)
Observations	18,202	5,773	4,791	1,716	18,202
Sample Restrictions					
Prior 10mi Extraction<Median	No	Yes	No	Yes	No
Pop Density>Median	No	No	Yes	Yes	No
Inc/Cap>Median	No	No	Yes	Yes	No

Note: The table shows that OLS effects of near exposure on total income approach IV estimates when restricting to the sample of compliers. Coefficients show the impact of a 1 standard deviation increase in the value of oil and gas extraction within a set radius of each neighborhood in the sample. The outcome in all regressions is the long difference in total income. Each row shows impacts of extraction at increasing distance thresholds. Coefficients should be interpreted as the percentage point change in the outcome resulting from a 1 SD increase in extracted value. Standard errors clustered by County are shown in parentheses; *, ** and *** indicate significance at 10%, 5% and 1% levels respectively.

similar positive impacts of approximately 6pp at the county-level; this effect direction and size is also very close to findings in the existing literature, which also rely on county-level data. The two measure also show similar results in the third and forth columns of the table which are calculated at the image-level (as done in all other regression); however, here we see that the positive effect has dissipated into small and imprecise negative coefficients.

I take these results to indicate that the difference between my negative effects and the positive ones in the literature is largely driven by the smaller neighborhood scale of my analysis. The results of Table 2.4 show that the outcome measure can be interpreted directly as census measured income in this long-term analysis. Other sources of differences such as precise variable definition, sample inclusion, or estimating equation are minor; the very similar county-level results give confidence that these other factors are not shifting coefficients substantially. A mechanical but important difference between my overall baseline results (in Table 2.1) and those in the existing literature is that I am able to study a longer post period. As I show in the event-study analysis below, gaps from treatment begin slowly but accumulate in later years of the panel. Finally, one novel implication of the local scale of my analysis is that the geographic unit of an observation becomes a fixed size square, sidestepping concerns of endogenous county boundaries and sizes in existing aggregated analyses.

2.7 Mechanisms and Heterogeneous Treatment

In this section I augment the baseline regression model to produce mechanisms and heterogeneity results. The mechanisms exercises apply alternate outcomes to the same IV estimation, showing which economic factors shift alongside the primary results and may be important mechanisms generating them. Conversely, for heterogeneity exercises I interact treatment with indicators for subgroups of interest. Given the rich variation across the small geographies in my sample, this analysis can pick up on more nuanced differences in treatment than would be possible with more aggregated data.

Table 2.4: IV Results by Geographic Scale and Income Measure. 2000-2010

	County		Image	
	Census	Predicted	Census	Predicted
Extraction 1mi	0.054*	0.061***	-0.006	-0.002
	(0.030)	(0.022)	(0.044)	(0.032)
Extraction 3mi	0.056*	0.061***	-0.005	-0.006
	(0.030)	(0.022)	(0.038)	(0.027)
Extraction 10mi	0.061**	0.067***	-0.001	-0.004
	(0.030)	(0.022)	(0.031)	(0.021)
Extraction 20mi	0.059**	0.063***	0.002	-0.002
	(0.023)	(0.020)	(0.027)	(0.019)
Extraction 40mi	0.066***	0.070***	0.005	-0.000
	(0.022)	(0.020)	(0.027)	(0.018)
Extraction 60mi	0.089***	0.092***	-0.000	-0.002
	(0.026)	(0.021)	(0.028)	(0.019)
Extraction 80mi	0.132***	0.135***	-0.011	-0.010
	(0.040)	(0.032)	(0.032)	(0.020)
Extraction 100mi	0.191***	0.194***	-0.026	-0.021
	(0.070)	(0.058)	(0.036)	(0.021)
Observations	211	211	4,020	4,020

Note: The table shows that results using my predicted income change measure are statistically indistinguishable from the county measure over the 2000 to 2010 period in which both are available. Also notable is that aggregating the data to the county level results in positive effects similar to those in the existing literature. Coefficients show the impact of a 1 standard deviation increase in the value of oil and gas extraction within a set radius of each neighborhood in the sample, as specified in equation 1. Outcomes are long differences of log income from 2000 to 2010. Each row shows impacts of extraction at increasing distance thresholds. Coefficients should be interpreted as the percentage point change in outcome resulting from a 1 SD increase in extracted value. Standard errors clustered by County are shown in parentheses; *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

2.7.1 Mechanisms

In Table 2.5 I estimate the baseline IV regressions across a set of outcomes which inform the mechanisms driving treatment effects on total income. These neighborhood-level outcomes include FHFA measured home price changes, particular matter 2.5 changes as measured in Di et al., 2019, and changes in employment overall and by related sectors.[17] The first clear finding in this table is that my primary treatment effects are not driven by distinct changes in local home prices or direct pollution. While home prices show modest but imprecise increases (between 1 and 5 percentage points) associated with a 1 SD increase in extracted value at each spatial horizon, pollution measures indicate that air quality may have even improved as a result of extraction. For example at the 3 and 10 mile horizons, effect sizes of 2 to 3 percentage point decreases are measured at the 10% confidence level. One possible explanation for this decline could be that fracking displaced outdated oil infrastructure which was operating based on out-of-date environmental standards.

The employment mechanisms displayed in the last 6 columns of Table 2.5 show the sectors which are helped and hurt most by a local fracking boom. Overall there is only a modest and imprecise impact of fracking on total local employment. In contrast, there is a clear decline in agriculture employment, most strongly felt from increased extraction activity within 1 mile of the neighborhood which is associated with a 20 percentage point decline in agriculture employment. This sectoral substitution is likely driven by both labor and capital contributions: land and low-skilled labor that was previously used for agriculture in these area seems to have been redirected towards fracking as it rolled out. Employment sectors which show increases complementary to the large increase in mineral extraction employment are utilities, construction and real estate. A 1 SD increase in extraction within 1 mile led to an employment increase of approximately 10 percentage points in each of these complementary sectors, with the association shrinking to approximately 3.5 percentage points at the 100 mile range. Each of these growing sectors has direct complementarities with fracking extraction: labor in utilities support the direct infrastructure transporting

[17] Employment sectors are selected for inclusion in this table based on having statistically distinguishable association with treatment.

Table 2.5: Mechanisms of IV Extraction Impacts by Exposure Distance

			Employment					
	Home Prices	PM2.5	Total	Ag	Mineral Ext	Utilities	Const	Real Estate
Ext 1mi	0.050	-0.043	0.024	-0.201***	0.858***	0.114*	0.116**	0.102**
	(0.037)	(0.029)	(0.033)	(0.061)	(0.199)	(0.060)	(0.049)	(0.046)
Ext 3mi	0.032	-0.030*	0.020	-0.127***	0.536***	0.076**	0.079***	0.068**
	(0.023)	(0.018)	(0.020)	(0.037)	(0.112)	(0.036)	(0.029)	(0.027)
Ext 10mi	0.022	-0.021*	0.012	-0.099***	0.394***	0.055**	0.056***	0.050**
	(0.015)	(0.012)	(0.015)	(0.027)	(0.078)	(0.026)	(0.020)	(0.020)
Ext 20mi	0.016	-0.017	0.009	-0.089***	0.331***	0.046**	0.046***	0.042**
	(0.012)	(0.010)	(0.012)	(0.022)	(0.054)	(0.021)	(0.016)	(0.017)
Ext 40mi	0.012	-0.015	0.009	-0.079***	0.282***	0.041**	0.038***	0.036**
	(0.009)	(0.009)	(0.010)	(0.017)	(0.039)	(0.017)	(0.014)	(0.016)
Ext 60mi	0.011	-0.013	0.012	-0.073***	0.267***	0.043***	0.036***	0.035**
	(0.009)	(0.009)	(0.009)	(0.015)	(0.036)	(0.016)	(0.014)	(0.015)
Ext 80mi	0.012	-0.011	0.013	-0.068***	0.259***	0.041***	0.035***	0.034**
	(0.009)	(0.008)	(0.008)	(0.013)	(0.036)	(0.015)	(0.013)	(0.015)
Ext 100mi	0.010	-0.009	0.012	-0.062***	0.248***	0.039***	0.032***	0.034**
	(0.010)	(0.008)	(0.008)	(0.013)	(0.035)	(0.015)	(0.012)	(0.014)
Obs	14,159	12,150	18,202	18,202	18,202	18,202	18,202	18,202

Note: The table shows that mechanisms closely associated with my negative total income results include increases in mineral extraction, utilities, construction, and real estate employment, as well as a decline in agriculture employment. Coefficients show the impact of a 1 standard deviation increase in the value of oil and gas extraction within a set radius of each neighborhood in the sample, as specified in equation 1. Outcomes are long differences of log variables designated in column headers. Each row shows impacts of extraction at increasing distance thresholds. Coefficients should be interpreted as the percentage point change in outcome resulting from a 1 SD increase in extracted value. Standard errors clustered by County are shown in parentheses; *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

and processing the extracted resources, construction work includes the construction of utility systems, and real estate employment includes leasing of both property and machinery which are the core capital inputs for fracking. Other sectors not shown in this table do not show a discernible association with the fracking shock.

2.7.2 Heterogeneity

Building on the interpretation exercises in Section 2.6 and the spatial gradient impacts in 2.5, I conduct heterogeneity exercises to break down how the treatment effects of nearby extraction on total income in Table 2.1 are distributed across groups in the sample. The granular nature of my data allow for a high degree of local variation in neighborhood characteristics, which grants more power to investigate heterogeneous treatment effects. In particular I estimate simple interaction models which add to the baseline model a binary indicator of group and an interaction between extraction value (treatment intensity) and the group. Coefficients on these interaction terms are shown in Table 2.6. Margins of heterogeneous effects explored here include income and population, race, proximity to shale reserves, state, play and dominant employment sectors.[18]

The first clear finding of these heterogeneity regressions is that neighborhoods which are wealthier in 2000 are much more negatively impacted by near exposure to fracking. Neighborhoods with average income above the median in 2000 ($26.3k) experience 10 percentage points less income growth than low income neighborhoods when exposed to 1 SD more fracking within 3 miles. The magnitude of this difference indicates that my estimated overall treatment effects are driven by negative impacts in relatively high income neighborhoods. Interestingly above median initial shares of white residents or population density do not show differential treatment effects. While these are important margins of the degree of exposure, for a fixed treatment intensity there are no differential impacts among these groups.

Another substantive margin of disparate impacts is proximity to the top quartile of shale. As in the validity analysis, this proximity is used by Bartik et al., 2019 as a simple metric for

[18]Note that only state, play, and employment categories with statistically distinguishable heterogeneous effects are displayed in Table 2.6.

Table 2.6: Heterogeneous Effects of Near Exposure by Subgroups

	1mi	3mi	10mi	20mi	40mi
High Inc/Cap	-0.164**	-0.096***	-0.049**	-0.039**	-0.029***
	(0.069)	(0.036)	(0.020)	(0.016)	(0.011)
High Pop Density	-0.061	-0.015	0.004	0.012	0.020*
	(0.059)	(0.025)	(0.015)	(0.012)	(0.011)
High White	0.018	0.012	0.003	0.004	0.005
	(0.052)	(0.028)	(0.019)	(0.014)	(0.010)
Far from Shale	-0.921	-0.628	-0.735**	-0.603**	-0.338**
	(0.701)	(0.410)	(0.363)	(0.234)	(0.131)
Arkansas	0.121***	0.079***	0.063***	0.064***	0.060***
	(0.038)	(0.024)	(0.019)	(0.019)	(0.020)
Colorado	0.275*	0.196	0.125*	0.097*	0.060*
	(0.160)	(0.128)	(0.072)	(0.057)	(0.031)
New York	-1.355	-1.109	-0.508**	-0.121***	-0.030*
	(1.208)	(0.909)	(0.230)	(0.038)	(0.018)
Pennsylvania	0.118*	0.072*	0.054*	0.047*	0.042
	(0.070)	(0.041)	(0.030)	(0.025)	(0.037)
Virginia	117.355	-4.615***	-1.814***	-0.374**	-0.233
	(217.902)	(1.523)	(0.463)	(0.153)	(0.154)
West Virginia	0.349	0.210**	0.122***	0.094***	0.089***
	(0.237)	(0.102)	(0.045)	(0.028)	(0.024)
Pennsylvania vs Ohio	0.193**	0.102*	0.079*	0.056*	0.044
	(0.097)	(0.053)	(0.041)	(0.032)	(0.041)
Fayetteville	0.115***	0.074***	0.059***	0.058***	0.049**
	(0.038)	(0.025)	(0.018)	(0.019)	(0.020)
Niobrara	0.285*	0.202	0.131*	0.102*	0.064**
	(0.163)	(0.125)	(0.070)	(0.054)	(0.027)
High Mineral Emp	0.230***	0.118***	0.082***	0.059***	0.032**
	(0.083)	(0.039)	(0.026)	(0.018)	(0.013)
High Construction Emp	0.155**	0.110***	0.069***	0.044**	0.021
	(0.062)	(0.039)	(0.025)	(0.018)	(0.013)
High Education Emp	-0.091*	-0.056*	-0.035*	-0.028**	-0.016
	(0.049)	(0.030)	(0.019)	(0.014)	(0.011)
High Hotel/Dining Emp	0.089**	0.054**	0.038**	0.031**	0.028**
	(0.041)	(0.024)	(0.018)	(0.014)	(0.011)
Observations	18,202	18,202	18,202	18,202	18,202

Note: The table shows differential treatment effects for splits of the overall analysis sample. Notably I find that high average income neighborhoods far from central drilling areas experience the largest negative effects, state policies in PA are associated with more favorable outcomes, and areas specializing in impacted employment sectors see more growth resulting. Coefficients show the impact of a 1 standard deviation increase in the value of oil and gas extraction within a set radius of each neighborhood in the sample, as specified in equation 1. The outcome in all regressions is the long difference in total income. In this table the columns show effects at increasing distance thresholds, while the rows show effects for the sample subgroups defined in Section 4.7. Coefficients should be interpreted as the percentage point change in income resulting from a 1 SD increase in extracted value. Standard errors clustered by County are shown in parentheses; *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

how close a neighborhood is to the central fracking region of each play. Neighborhoods in my analysis sample which are above the median here–more than 35 miles from the top quartile of the play prospectivity–exhibit a very large but noisily measured effect. It is important to qualify this magnitude by observing that mean extraction exposure in this group is approximately half a standard deviation lower than in the near group. So while an estimated difference in effect of -74 percentage points (in the 10 mile case) is very large, it represents an estimate far outside the range of exposure (1 SD) experienced by this group in the data. Given this limitation on interpreting the magnitude, the appropriate conclusion here is that neighborhoods further from the central play activity seem to be broadly more sensitive to increases in nearby extraction.

Administrative boundaries offer a unique opportunity to investigate how policy differences are associated with differential neighborhood treatment effects of fracking. I investigate the margin of state differences in effects in Table 2.6 and find that several states experience less severe negative local impacts. Arkansas, Colorado, Pennsylvania, and West Virginia all have clear positive interaction effects on treatment, showing evidence of positive impacts of local fracking which nullify the main negative effects to varying degree (West Virginia shows the largest positive magnitude). Conversely, New York and Virginia exhibit clear evidence of exacerbated negative fracking impacts relative to the rest of the sample.

Because these state differences in effects may be driven by correlations between state and other more causal margins of treatment heterogeneity, I investigate the case of Pennsylvania in more detail. I specifically compare it directly to the neighboring state of Ohio, as the border between the two is near a large mass of the sample: neighborhoods exposed to the Marcellus and Utica plays. The results of this subsample direct comparison are shown in the Pennsylvania vs Ohio row. I find that while neighborhoods in Ohio experience negative impacts slightly larger than those of the overall sample, those in Pennsylvania show no evidence of negative fracking impacts on local total income. State policy differences around fracking that may be driving this gap are discussed in detail in the State Impact Pennsylvania NPR series. Specific policies around local control of siting and a more general focus on environmental protection rather than resource extraction make

the regulatory environment in Pennsylvania much more protective of local communities than in Ohio. While these policies may contribute to differential impacts in parts of the states, Figure 2.12 shows that the gap in treatment effects shrinks to 0 when focusing narrowly on neighborhoods within 50 miles of the border between states. Therefor, without further investigation using random variation in policy regimes, the evidence of the role of environmental policy on fracking treatment effects remains circumstantial. However, regardless of the causal mechanism, these muted near effects in Pennsylvania imply that existing evidence of negative health and housing effects–both measured solely in Pennsylvania–are likely a substantial understatement of the true negative near effects nationwide (Currie et al., 2017, Muehlenbachs et al., 2015).

The next margin of heterogeneity examined in Table 2.6 is play. I find that the major plays do not exhibit major differences in treatment effects, but smaller plays (each less than 7% of the sample) do show some evidence of substantially positive treatment effects relative to the rest of the sample. Specifically, the Fayetteville play in Arkansas shows a clear and positive interaction with treatment which indicates that neighborhoods exposed to this play (i.e. those in Arkansas) do not exhibit the negative treatment effects seen in the overall sample. Once again here, interpreting whether this difference is causal or a correlation with other characteristics of these neighborhoods is not possible given the identifying variation I study.

Finally I show in Table 2.6 that neighborhoods with an above median share of employment in mineral extraction, construction or hospitality sectors in the 3 years prior to fracking starting experienced disproportionately positive treatment effects, while those with employment concentrated in education or health services saw disproportionate negative impacts. While the mechanism findings gave similar insight on which employment categories move with and against the mineral extraction employment from fracking, these heterogeneity results show that the composition of the local labor market prior to this shock is one channel through which treatment effects appear to be amplified or dampened. Areas which already specialize in the sectors which experience a positive direct or complementary shock are the ones that see the least negative (or most positive) impacts. Conversely, education employment declines. The vast majority of education employment is in

primary and secondary schools. Because I do not observe significant negative impacts on education employment coinciding with the shock, an alternative explanation of limited local labor force available to work in the rising extraction and adjacent industries may be a more likely explanation.

2.8 Discussion

In this paper I apply a remotely sensed measure of residential investment to study the effects of fracking on the neighborhoods with highest exposure. I show that the disaggregation from county to this microspatial neighborhood scale enables the distinction of heterogeneous within-county effects and reveals negative effects on neighborhoods nearest wells. This finding confirms and quantifies economic intuition in a setting which has been shown to have locally concentrated environmental and industrial negative externalities, in stark contrast to revenues which predominantly flow out of the producing area. On average neighborhoods closest to wells see a long-term decline in prosperity characterized by a decline in population and, to a lesser degree, decay in existing residential infrastructure. These negative effects are however mitigated in areas with strong environmental protections and neighborhoods with pre-existing employment specialization in positively impacted sectors.

Considering my findings alongside the breadth of existing evidence, the immediate policy implications is that residents and their representatives should be cautious about adopting fracking in or near their neighborhoods. Several large states including California and New York, and even several countries have banned fracking outright. However in many states the balance between employment and revenue benefits, and negative residential impacts is an active debate. In several states including Colorado and Pennsylvania there are initiatives to increase the minimum setback distances between active wells and residential structures; Colorado for example is considering increasing the setback from 1,000 feet to 2,500 feet.[19] My findings inform these decisions by documenting the degree of negative impacts on exposed neighborhoods, offering detailed conditions of proximity and neighborhood initial characteristics under which residential decline of varying

[19]Coverage of recent state initiatives on setbacks rules is offered by The Guardian.

degrees can be expected. I also point to examples such as the state of Pennsylvania where these effects have been effectively avoided, most probably in part due to strong environmental regulations and a high degree of local control.

I contribute to a very limited field of evidence on economic impacts of near-exposure to fracking. This is because prior work has generally been constrained by coarse county-level data. My remotely sensed, machine learning based outcomes allow me to measure overall residential development at a national scale over an extended panel. To my knowledge this is the first application of daytime satellite imagery in a causal identification framework. Given the proliferation of satellite imagery at ever increasing resolution and advances in computational techniques and machinery to identify patterns in this imagery, future work in this direction promises to expand our understanding of existing questions and raise entirely new ones. The particular strength of this approach is the high temporal and spatial resolution of the underlying satellite data; however, the success of related projects will also depend critically on the availability and quality of regular Censuses with which to train models of imagery on. Continued improvement in the ability to interpret machine learning models is also critically important. The ability of social scientists to interpret the mechanisms and output of these models will determine how quickly and widely they are adopted.

A natural next step in the study of local impacts of fracking would be to exploit sources of exogenous variation in policy. My results identify the spatial gradient of fracking and document associations between treatment intensity and neighborhood characteristics, including policy. Exploring these associations further will yield deeper insight and prescriptions for targeted policy. Furthermore, my approach can be directly extended to study the effects of volatility and decline in domestic oil and gas production as data following recent contractions becomes available. More broadly, there is wide scope to study localized shocks using my empirical approach. Familiar shocks such as the opening or closing of factories, warehouses, and plants offer similar opportunities to study the impacts of local industrial activity. Beyond this, current and future changes in domestic energy infrastructure such as the adoption of renewable energy sources represent a broad set of economic shocks with similar, scale-dependent mechanisms to explore.

Bibliography

Allcott, H. and D. Keniston (2017). Dutch disease or agglomeration? The local economic effects of natural resource booms in modern America. *The Review of Economic Studies 85*(2), 695–731.

Barrows, G., T. Garg, and A. Jha (2019). The health costs of coal-fired power plants in india. *Available at SSRN 3281904*.

Bartik, A. W., J. Currie, M. Greenstone, and C. R. Knittel (2019). The local economic and welfare consequences of hydraulic fracturing. *American Economic Journal: Applied Economics 11*(4), 105–55.

Chay, K. Y. and M. Greenstone (2005). Does air quality matter? evidence from the housing market. *Journal of political Economy 113*(2), 376–424.

Conley, T. G. (2010). Spatial econometrics. In *Microeconometrics*, pp. 303–313. Springer.

Currie, J., L. Davis, M. Greenstone, and R. Walker (2015). Environmental health risks and housing values: evidence from 1,600 toxic plant openings and closings. *American Economic Review 105*(2), 678–709.

Currie, J., M. Greenstone, and K. Meckel (2017). Hydraulic fracturing and infant health: New evidence from pennsylvania. *Science advances 3*(12), e1603021.

Currie, J., J. Voorheis, and R. Walker (2020). What caused racial disparities in particulate exposure to fall? new evidence from the clean air act and satellite-based measures of air quality. Technical report, National Bureau of Economic Research.

Currie, J. and R. Walker (2011). Traffic congestion and infant health: Evidence from e-zpass. *American Economic Journal: Applied Economics 3*(1), 65–90.

Di, Q., H. Amini, L. Shi, I. Kloog, R. Silvern, J. Kelly, M. B. Sabath, C. Choirat, P. Koutrakis, and A. Lyapustin (2019). An ensemble-based model of pm2. 5 concentration across the contiguous united states with high spatiotemporal resolution. *Environment international 130*, 104909.

Donaldson, D. and A. Storeygard (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives 30*(4), 171–98.

Duranton, G. and D. Puga (2015). Chapter 8 - Urban land use. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5 of *Handbook of Regional and Urban Economics*, pp. 467–560. Elsevier.

Feyrer, J., E. T. Mansur, and B. Sacerdote (2017). Geographic dispersion of economic shocks: Evidence from the fracking revolution. *American Economic Review 107*(4), 1313–34.

Fowlie, M., E. Rubin, and R. Walker (2019). Bringing satellite-based air quality estimates down to earth. In *AEA Papers and Proceedings*, Volume 109, pp. 283–88.

Glaeser, E. L. and M. E. Kahn (2001). Decentralized employment and the transformation of the american city.

Glaeser, E. L., S. D. Kominers, M. Luca, and N. Naik (2018). Big data and big cities: The promises and limitations of improved measures of urban life. *Economic Inquiry 56*(1), 114–137.

Gold, R. (2014). *The boom: How fracking ignited the American energy revolution and changed the world*. Simon and Schuster.

Greenstone, M. and J. Gallagher (2008). Does hazardous waste matter? evidence from the housing market and the superfund program. *The Quarterly Journal of Economics 123*(3), 951–1003.

Greenstone, M. and E. Moretti (2003). Bidding for industrial plants: Does winning a 'million dollar plant' increase welfare?

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American economic review 102*(2), 994–1028.

Khachiyan, A., A. Thomas, H. Zhou, G. Hanson, A. Cloninger, T. Rosing, and A. Khandelwal (2021). Using Neural Networks to Predict Micro-Spatial Economic Growth. Working paper.

Kiel, K. A. and M. Williams (2007). The impact of superfund sites on local property values: Are all sites the same? *Journal of Urban Economics 61*(1), 170–192.

Moretti, E. (2004). Human capital externalities in cities. In *Handbook of regional and urban economics*, Volume 4, pp. 2243–2291. Elsevier.

Muehlenbachs, L., E. Spiller, and C. Timmins (2015). The housing market impacts of shale gas development. *American Economic Review 105*(12), 3633–59.

Smith, V. K. and J.-C. Huang (1995). Can markets value air quality? a meta-analysis of hedonic property value models. *Journal of political economy 103*(1), 209–227.

Zucker, H. A. and S. Dreslin (2014). A public health review of high volume hydraulic fracturing for shale gas development. *New York State Department of Health 100*.

2.9 Appendix

I present two sets of results examining the robustness of my primary regressions. The first is presented in Table 2.7 in which I show the baseline regression results using well count instead of extraction value as the endogenous exposure variable. The motivation of this exercise is that the local negative externalities associated with fracking may be more directly related to the number of wells nearby than the total value of extraction occurring. If the biggest local disruptions happen during the well completion process rather than the production stage, then the simple number of wells nearby may offer a more precise measure of local treatment effects. Table 2.7 shows that the measured impacts of well exposure are quantitatively and qualitatively very similar to those of extracted value exposure. One notable difference in the results is that OLS models on well exposure at proximities of 20 miles and below show a much more clear negative association between more fracking and total income change. While the effect size remains much smaller than the IV models at these bandwidths, the estimation is precise enough to be statistically significant at the 1% level and consistently negative across bandwidths.

The second set of robustness results, shown in Table 2.8, tests small modifications of the estimation strategy or analysis sample used in the baseline IV income change regression. Across these specifications we observe that changes to the analysis sample (relative to the baseline model and sample in the first column) do not lead to meaningful differences in the results, however the data does not have enough power to include State or county-level fixed effects in near exposure specifications. The second column adds prediction error from the Satellite CNN model in 2010 as a control to address concerns that the spatial distribution of errors is driving observed results.[20] The third column include images in the North East Corridor region, showing that the exclusion of these images in baseline results is not driving the results. Next, the Full Sample column shows that further including income and population outliers also does not change the primary findings. The No Well Imgs column alternatively tests the baseline specification excluding neighborhoods

[20]Figure 2.11 further shows that the distributions of difference prediction errors in my analysis sample and the overall sample of images are almost identical and both have a mean very close to 0.

with a well located inside their boundaries. This sample helps to ameliorate the concern that the model was not trained specifically on fracking areas and thus may be incorrectly interpreting well infrastructure as negative local impacts. In fact here we see an almost doubling of the 1 mile treatment effect of extraction, suggesting that the opposite is likely the case: that the appearance of well infrastructure in the images of exposed neighborhoods is dampening the negative impacts across all neighborhoods with exposure at the 1 mile threshold. The sixth column of Table 2.8 omits all images exposed to shale reserves of multiple plays within 40 miles, addressing the concern that nuances of the well assignment strategy in these multiple exposure cases are not driving primary results.

Finally the last two columns use state and county fixed effects, respectively, in place of play fixed effects. The motivation of this exercise is to test if the data support controlling for fixed local economic characteristics when estimating localized treatment effects. We see here that both specifications make effects at near bandwidths statistically zero. This result is consistent with the interpretation that the analysis sample does not contain enough variation in 1-mile exposure within counties or states to precisely measure treatment effects at the range. This is likely the case because particularly the county fixed effects are absorbing the existing variation in very near-range exposure, limiting that available for estimation. At higher bandwidths this issue is resolved and clear effects consistent with the baseline models are observed in these local fixed effect models.

Table 2.7: IV And OLS Baseline Impacts of Wells by Exposure Distance

	Income		Population		Income/Population		First
	OLS	IV	OLS	IV	OLS	IV	Stage
Wells 1mi	-0.009***	-0.076***	-0.010***	-0.060***	0.001	-0.017	0.246***
	(0.003)	(0.021)	(0.003)	(0.020)	(0.002)	(0.012)	(0.046)
Wells 3mi	-0.012***	-0.062***	-0.012***	-0.050***	0.000	-0.012	0.336***
	(0.004)	(0.017)	(0.004)	(0.015)	(0.002)	(0.009)	(0.064)
Wells 10mi	-0.014***	-0.044***	-0.015***	-0.035***	0.001	-0.009	0.505***
	(0.004)	(0.011)	(0.004)	(0.010)	(0.002)	(0.006)	(0.073)
Wells 20mi	-0.016***	-0.037***	-0.016***	-0.029***	-0.000	-0.008*	0.646***
	(0.004)	(0.009)	(0.004)	(0.008)	(0.002)	(0.004)	(0.066)
Wells 40mi	-0.016***	-0.035***	-0.015***	-0.029***	-0.001	-0.007*	0.736***
	(0.005)	(0.009)	(0.004)	(0.008)	(0.003)	(0.004)	(0.049)
Wells 60mi	-0.019***	-0.038***	-0.020***	-0.031***	0.001	-0.006	0.712***
	(0.005)	(0.009)	(0.005)	(0.008)	(0.003)	(0.004)	(0.051)
Wells 80mi	-0.014***	-0.036***	-0.013***	-0.030***	-0.001	-0.006	0.717***
	(0.005)	(0.009)	(0.005)	(0.009)	(0.003)	(0.004)	(0.058)
Wells 100mi	-0.010*	-0.032***	-0.005	-0.027***	-0.005*	-0.005	0.746***
	(0.006)	(0.009)	(0.005)	(0.008)	(0.003)	(0.004)	(0.050)
Observations	18,202	18,202	18,202	18,202	18,202	18,202	18,202

Note: The table shows that coefficients on an alternative exposure measure of the number of wells yields very similar results to the baseline results reported in Table 2.1. Coefficients show the impact of a 1 standard deviation increase in the number of wells within a set radius of each neighborhood in the sample, as specified in equation 1. Outcomes are long differences of log variables designated in column headers. The first stage coefficient for all IV models is shown in the final column. Each row shows impacts of wells at increasing distance thresholds. Coefficients should be interpreted as the percentage point change in outcome resulting from a 1 SD increase in the number of wells. Standard errors clustered by County are shown in parentheses; *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

Table 2.8: Robustness Exercises for IV Impacts of Extraction by Exposure Distance

	Baseline Model	w Pred Error	w NE Corridor	Full Sample	No Well Imgs	1 Play w/in 40mi	State FEs	County FEs
Extr. 1mi	-0.102***	-0.104***	-0.099***	-0.093***	-0.197**	-0.084***	-0.054	0.109
	(0.033)	(0.033)	(0.028)	(0.026)	(0.076)	(0.027)	(0.034)	(0.115)
Extr. 3mi	-0.064***	-0.065***	-0.061***	-0.057***	-0.074***	-0.053***	-0.035*	0.065
	(0.019)	(0.019)	(0.016)	(0.015)	(0.025)	(0.015)	(0.021)	(0.060)
Extr. 10mi	-0.048***	-0.049***	-0.045***	-0.042***	-0.049***	-0.036***	-0.028*	0.023
	(0.014)	(0.014)	(0.012)	(0.011)	(0.015)	(0.010)	(0.015)	(0.029)
Extr. 20mi	-0.041***	-0.042***	-0.038***	-0.035***	-0.040***	-0.036***	-0.025**	0.004
	(0.011)	(0.011)	(0.009)	(0.008)	(0.011)	(0.009)	(0.012)	(0.019)
Extr. 40mi	-0.037***	-0.038***	-0.034***	-0.031***	-0.034***	-0.036***	-0.023***	-0.020
	(0.009)	(0.009)	(0.007)	(0.006)	(0.008)	(0.009)	(0.009)	(0.018)
Extr. 60mi	-0.035***	-0.035***	-0.034***	-0.031***	-0.032***	-0.035***	-0.024***	-0.050**
	(0.008)	(0.008)	(0.006)	(0.006)	(0.007)	(0.008)	(0.008)	(0.024)
Extr. 80mi	-0.031***	-0.032***	-0.032***	-0.029***	-0.028***	-0.032***	-0.020**	-0.064**
	(0.007)	(0.007)	(0.006)	(0.005)	(0.007)	(0.007)	(0.008)	(0.026)
Extr. 100mi	-0.026***	-0.027***	-0.030***	-0.027***	-0.023***	-0.028***	-0.013*	-0.054**
	(0.007)	(0.007)	(0.005)	(0.005)	(0.006)	(0.007)	(0.008)	(0.026)
Observations	18,202	18,202	24,683	28,343	14,383	15,994	18,202	18,122

Note: The table shows robustness exercises on the IV impact of a 1 standard deviation increase in total barrels of oil extraction value within a set radius of each neighborhood in the sample. The outcome in each is the long differences of log income. Columns designate robustness exercises. Baseline refers to the regressions shown in the main results section. w Pred Error shows the results of the baseline model with an additional control for the residential income prediction error in 2010. With NE Corridor shows results on the sample including neighborhoods in the North East Corridor region of the sample. Full Sample further includes images which are outliers for predicted income or population. No Well Imgs is a version of the analysis sample which excludes images with an active well located in the image area. 1 Play w/in 40mi alternatively limits the sample to images with at most 1 shale play within 40 miles, dropping images with multiple plays within that proximity. Finally, State FEs shows the baseline IV results further incorporating State-level fixed effects, and County FEs alternatively includes County-level fixed effects with the baseline play fixed effects. Each row shows impacts of extraction at increasing distance thresholds. Coefficients should be interpreted as the percent change in outcome resulting from a 1 SD increase in extraction. Standard errors clustered by County are shown in parentheses; *, ** and *** indicate significance at 10%, 5% and 1% levels respectively.

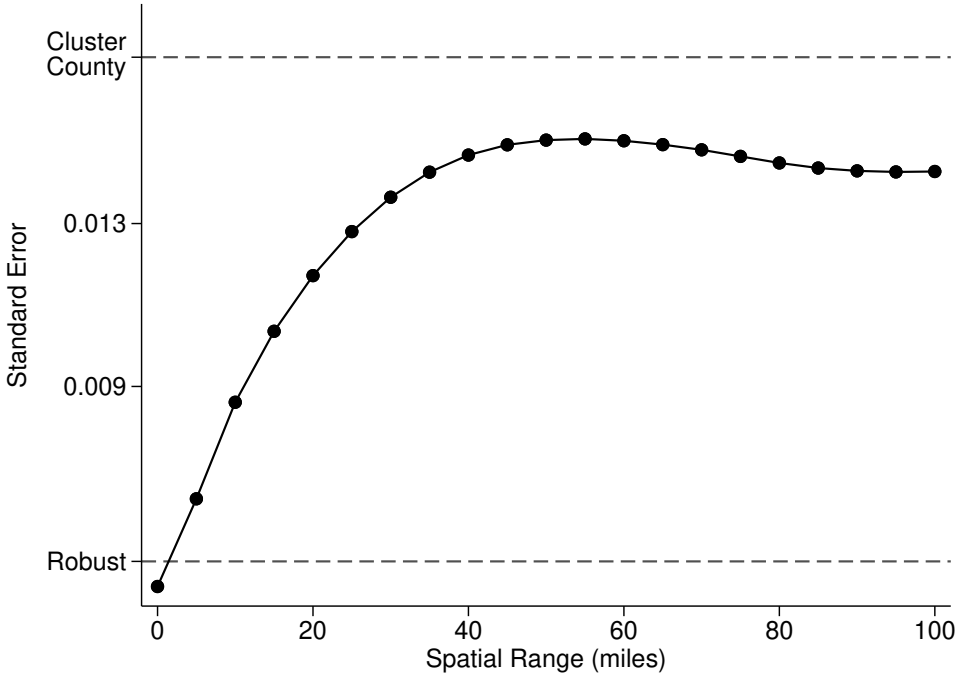


Figure 2.10: Size of Spatial Standard Errors Relative to Clustering

Note: This figure plots Conley standard errors for the coefficient in column 2, row 3 of Table 2.1 (IV total income impact of a 1 standard deviation increase in extraction within 10 miles). This approach allows for spatial clustering between sample units up to a set radius. By increasing this radius along the x-axis, I show that the standard errors with this approach fall between a naive heteroskedasticity robust approach and the county clustering approach I adopt, even when allowing for spatial correlation up to 100 miles.

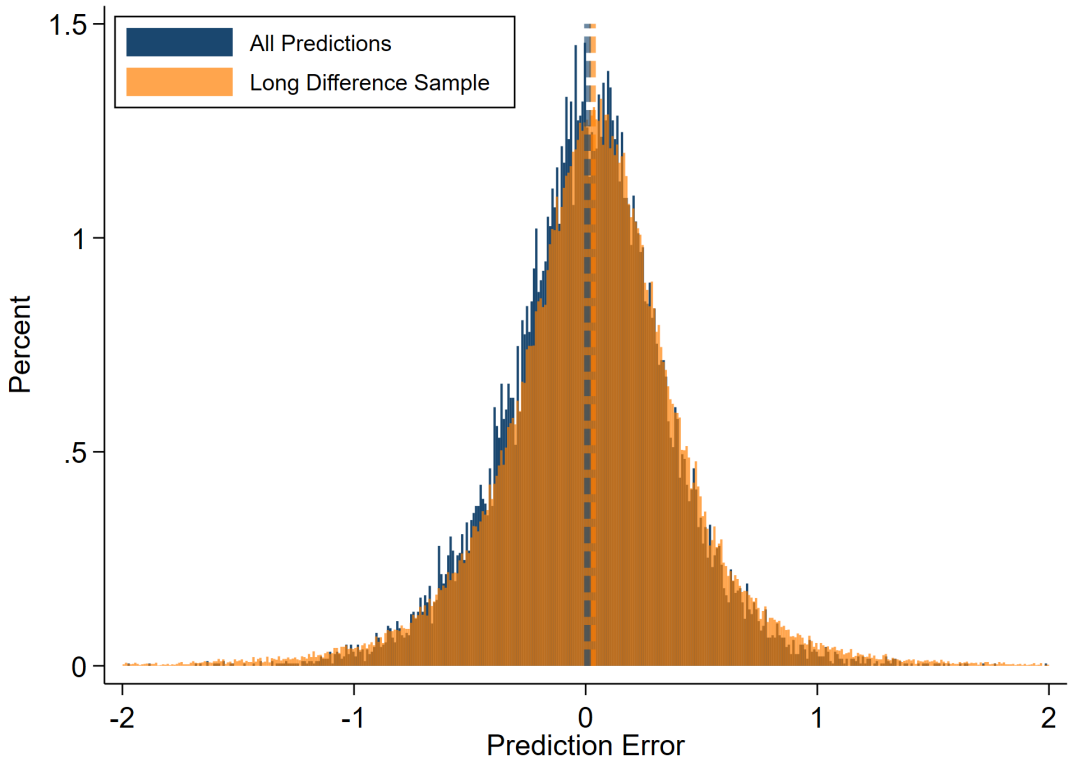


Figure 2.11: Difference Prediction Error Distribution

Note: This figure shows that income change prediction errors are distributed very similarly in my analysis sample and the overall sample of neighborhoods, indicating that my findings are unlikely to be driven by a correlation between income change prediction errors and sample inclusion. Predictions are of total income change from 2000 to 2010. Dashed vertical lines show the mean value of each distribution; in both cases it is very close to 0 which confirms that the predictions are unbiased on average both overall and in the analysis sample.

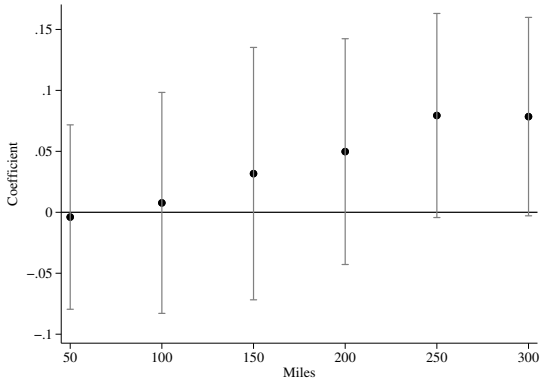


Figure 2.12: IV Interaction Impacts, PA vs OH by Distance to Border

Note: This figure shows that the interaction effect on Pennsylvania in the direct comparison to effects in Ohio shrinks to 0 when limiting to images near the border between the two states. The model used to generate these interaction effects on Pennsylvania is described in Section 2.7.2, and results on the unrestricted comparison between states is shown in the PA (v OH) row of Table 2.6.

Chapter 2 is currently being prepared for submission for publication of the material. Khachiyan, A. The dissertation author was the primary investigator and author of this material.

Chapter 3

Occupational Skill Portability: How Mobility Patterns Can Enhance Existing Skills Data

Arman Khachiyan

3.1 Abstract

Rich data on the multi-dimensional task requirements of each occupation has sparked a breadth of economic literature examining the portability of human capital across the labor market. A primitive in such analyses is constructing a norm over the vectors of occupational skills to create a continuous measures of skill distance between occupation pairs. While the existing literature has centered around factor analysis and angular separation as the leading norms, I show that using a regression framework derived from an Eaton, Kortum, Roy model of occupation switching directly implies a novel, empirical norm which is disciplined by observed occupation switching patterns. This approach relieves key limitations of existing measures, such as linearity and the inability to distinguish directional differences in skill portability, and allows for an analysis of which skill dimensions are critical in the portability of human capital, and which are not. Implications for

existing results on skill portability are discussed, along with immediate policy applications to alleviate adjustments costs of workers switching occupations mid-career. Skill portability measures are aggregated, showing that compositional changes in employment by occupation since 1976 have lead to increased overall skill portability. Finally, using this novel measure of skill portability, network analysis shows that the incidence of a recession on job loss across the occupation network is related to the severity and duration of employment effects overall and by occupation.

3.2 Introduction

Labor demand changes arising from both structural and cyclical fluctuations can shape employment experiences for impacted workers. As discussed in Charles et al., 2016, structural manufacturing decline over recent decades contributed to the steady decline in male labor force participation. The interaction of this trend with a uniquely volatile construction business cycle before and through the great recession, they argue, contributed to the severity of the ensuing labor market impacts. While the demands for labor adjust in real-time according to technological advances, trade policy, tastes, and many other relatively flexible factors, workers make career decisions throughout life which can be highly persistent and often involve substantial adjustment costs to change.

To help mitigate these inevitable worker adjustment costs, government programs such as unemployment insurance, re-employment service centers, retraining efforts and moving subsidies have been conceived and implemented widely. To inform these programs and others, the US Department of Labor collects information on the skill content of each occupation. These measures are immediately used by employment service providers to match unemployed workers with jobs demanding their skills, and they have also catalyzed a field of economic literature studying how the skill content of work is systematically changing over time.

Despite the substantial financial and intellectual resources that have rightfully been attributed to this topic, there is little consensus on what information these skill data hold, and how

applicable they are for a variety of applications ranging from occupational mobility to wage inequality and job security. A primitive in the study of the portability of human capital across jobs is constructing a metric of skill distance between them. While the literature has developed several norms on the skill data to measure this skill distance, no work has been done to validate, compare or improve on these measures. This leaves a wide opening in our understanding, which this paper attempts to fill, of how to measure the portability of human capital and how to best apply these measures to address pressing public policy questions.

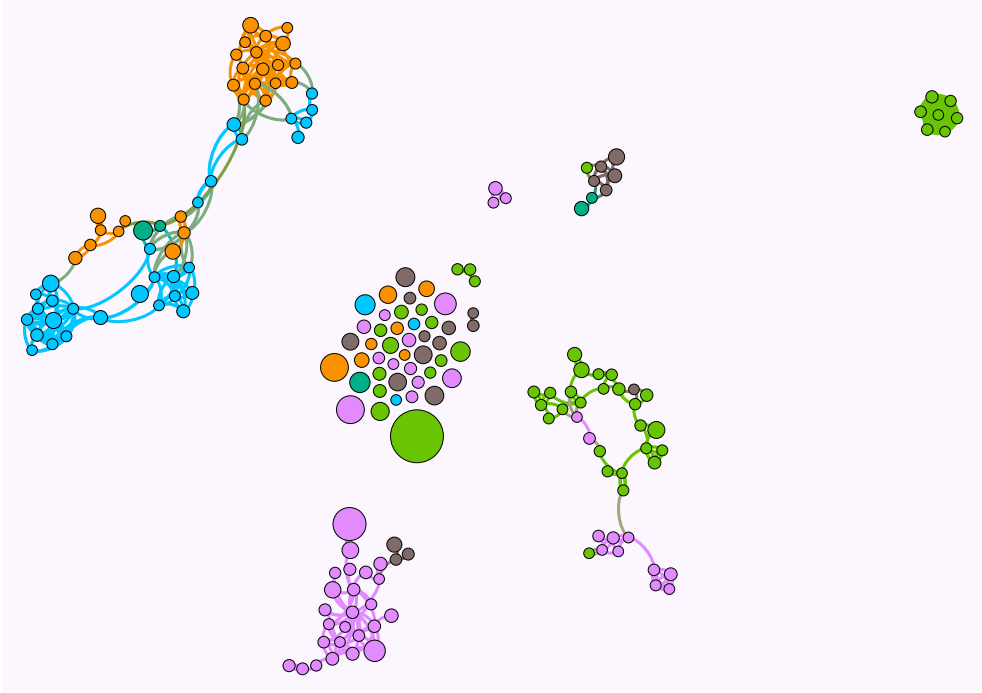
My first innovation is to develop a novel norm on existing skill data by targeting it to observed occupational mobility, and to conduct what I believe is the first analysis of the theoretical and empirical differences in properties of skill distance measures. I then construct a network of occupations using my new metric, shown in Figure 3.1, which guides several novel applications studying how recessions propagate between occupations. I further show how these methods can be used to characterize the outside options of workers in settings with more detailed data.

Key among my findings are that different measures of skill distance have substantially different distributions, and that the measure selected can qualitatively change leading results in the existing literature. I find some evidence that compositional changes in the labor market may in fact be driving higher skill portability, and that recessions disproportionately impacting central nodes of the occupation network tend to create substantially deeper labor market impacts.

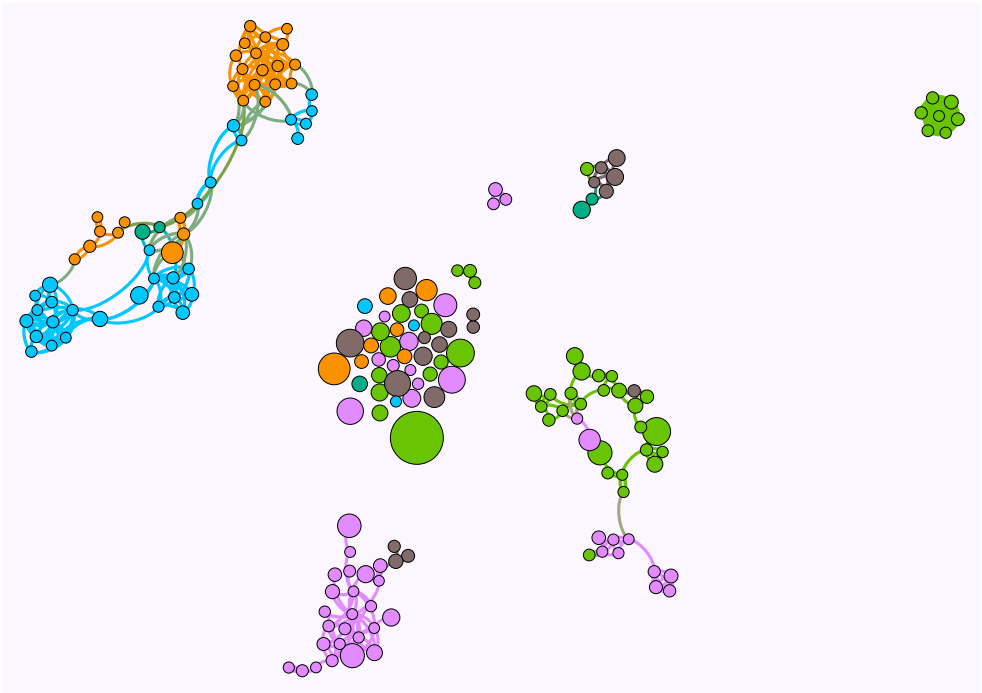
This work is informative both from the theoretical perspective of understanding the key drivers of occupational mobility in existing skills data, and from the practical perspective of designing novel analyses targeted to immediate policy concerns.

3.3 Literature Review

Policy applications with strong identification studying skill portability have focused on displaced workers. Because displaced workers are essentially defined as those who lose their job



A) 1976



B) 2018

Figure 3.1: Occupational Mobility Networks Over Time

Professional, Support, Service, Farm/Forest/Fish, Crafts, Operators/Laborers

due to labor demand factors, their job loss has been argued to be exogenous to the worker, helping to mitigate endogeneity concerns about worker quality leading to the separation. Furthermore, because the focus of this agenda is on alleviating worker adjustment costs induced by labor demand fluctuations, displaced workers broadly defined are the exact target group of the implied policy applications from such research.

Of course, thinking about labor frictions has been an important part of various macroeconomic models which have had meaningful impacts in the field. Most notably, Pissarides' framework for job-search incorporates labor frictions and has been widely adopted in a range of macroeconomic papers. More recently, Guvenen et al., 2019 studied the macroeconomic consequences of skill mismatch between workers and firms, and how this contributes to career growth within and between firms. My analysis follows Cortes and Gallipoli, 2018 in adapting some of these macroeconomic search models and concepts, along with work on gravity structures from trade, towards a mostly recent and growing sub-field of inquiry on these topics in applied labor economics.

In this area the core question has been how the skill overlap between occupations determines worker outcomes following displacement. Asking whether skill distance between the displacing and re-employing job is related to wage changes, Poletaev and Robinson, 2008 show that there is a significant and meaningful wage penalty to switching skill sets, which is distinct from the industry or firm switching penalty. Gathmann and Schönberg, 2010 study the extent to which human capital is task specific, and show that lifetime occupational transition data support a model of task-specificity and specific, meaningful income and time costs to switching skill sets by changing occupations.

While the existing literature is motivated by the key trends in skill demand identified in the skill-biased technical change literature, i.e. Autor et al., 2003, and work such as Charles et al., 2016, relatively little attention has been given to refining measures of skill distance, and studying how these measure can be aggregated into a systematic analysis of the complete labor market. This paper focuses on this goal by pushing forward the leading approaches in applied-microeconomics, redefining the core norm of skill data and rigorously comparing it to existing metrics on both core

properties and implications in a range of applications.

3.4 Data

Dictionary of Occupational Titles, 1991

Data on the skill composition of each occupation come from the Dictionary of Occupational Titles (DOT). Published in 1991, DOT represents the culmination of Department of Labor efforts to measure a set of skill demands common across US jobs. Using a combination of workplace surveys and on-site visits by occupational specialists, each occupation is assigned an intensity value across a wide range of pre-specified skill characteristics. This value is intended to capture the amount of each skill that is used in the regular execution of the given occupation. The primary purpose for creating the DOT, and the leading use case, was to assist job counselors in local employment offices to find suitable new occupations for unemployed workers. The idea being: when a worker arrives at a job center with a certain set of skills (for example as represented by her past experience in her resume), one of the counseling steps to help direct the job search is to open the physical DOT book and find occupations which align with her skill set.

This analysis specifically studies 66 dimensions of skill measured in DOT, all of which are listed in Figure 3.5. The core measures capture intensity of interaction with people, data and things, education in reasoning, math and language, general aptitudes including verbal ability, dexterity and coordination, a range of physical demands such as strength, balance and vision, personal temperaments such as performing repetitive work, expressing emotions, and making judgments, and finally environmental conditions such as exposure to extreme weather, loud noise, or hazardous materials. Further background, details and analysis of the DOT can be found in Miller et al., 1980.

While the DOT has since been revised as O*NET starting in 1998, I use the DOT for two primary reasons. First is for comparison, as this is the skill data used by the existing Labor Economics literature. Second is that the DOT was released in 1990, placing it roughly in the

middle of the existing Current Population Survey (CPS) micro-data on employment dynamics and minimizing concerns that skill requirements in a given year of study are different from those when DOT was measured. The 1990 release also enables a high fidelity merge between the DOT occupation codes and Census occupation codes from 1990, which IPUMS has created a crosswalk for spanning the full CPS history (starting in 1968). This long time series of a consistent occupation coding enables the analysis of changes over time, holding occupation definitions fixed.

Current Population Survey, 1976-2018

Occupation-level employment statistics are calculated using the Current Population Survey microdata (CPS), accessed through the Integrated Public Use Microdata Series (IPUMS). Annual measures are computed for each occupation, spanning the years 1976 to 2018. IPUMS constructs normalized occupation coding schemes to enable analysis across changes in official occupation schemes; for alignment with the DOT release timing and codes, I use the IPUMS scheme normalizing all occupation codes to the 1990 Census Bureau occupational classification scheme. In this way, each worker in each year is classified into exactly one occupation, and occupation definitions are fixed across time. While this does raise the concern that occupation definitions may be very unstable over time, it is a practical approach designed to handle the appearance of novel occupations and the grouping of existing occupations inevitable with structural changes in the labor market. Further details on the methodology used by IPUMS to construct this occupation crosswalk are available in the IPUMS documentation.

The key statistic calculated from the CPS is the flow of workers switching between occupations. In the monthly survey, each worker is asked their current occupation, and their occupation a year ago; a switch is defined as a difference between these two occupations, and is only defined for workers employed in both periods. For each directional pair of occupations (i.e. the switch from bus driver to truck driver is one directional pair), I compute the total number of workers who made the switch throughout the pooled 1976-2018 period. This switching matrix is the critical ingredient allowing a targeted measure of skill distance, in my novel measurement approach. Using the

CPS, I also compute annual occupation employment counts for various exercises on worker outside options and the occupation network dynamics of recessions. Finally, I also compute outcomes by occupation for displaced workers, such as re-employment rates, using the Displaced Worker Supplement (DWS) of the CPS. The biennial DWS tracks the experiences of workers whose job loss occurred as a result of firm downsizing.

Merging the DOT skill characteristics with employment counts results in a high-fidelity match representing approximately 98% of the labor force in every year of the sample. In total the analysis retains 317 unique occupation groups, representing the 3-digit occupation coding scheme of the Census Bureau in 1990 which were successfully merged with skill data from the DOT.

3.5 Methodology

3.5.1 Existing Measures of Skill Distance Between Occupation Pairs

In this paper I compare three commonly used measures of skill distance between pairs of occupations, and develop a novel one. You can think of each occupation as representing a point in a high-dimensional skill space. In other words, each occupation is essentially a unique combination of skill requirements. In this framing, determining the skill distance between two occupations is a fairly general problem of measuring distance in a high-dimensional space. As such, the most immediate and easily implemented measure of skill distance is the Euclidean distance (L2 norm) between the skill vectors of the two occupations. A conceptually minor variation on this is angular separation, a geometrically motivated measure of high-dimensional distance that has seen many recent applications in the skills literature. These measures are fairly similar because they both give exactly equal weighting to each dimension of skill, ruling-out the possibility that some skill dimensions (i.e. dealing with people) are more important determinants of a workers ability to transfer human capital across occupations than others (i.e. environmental conditions of the job).

To alleviate this important assumption, factor analysis has also been a popular dimension-

ality reduction method which has the benefit of giving varying loadings to different skill factors. This methods operates on the assumption that skill requirements of work can in fact be characterized by just a few underlying factors (often 4 are used when studying skills), and that the higher dimensional data represents various linear combinations of these underlying factors. By estimating these underlying skill factors for each occupation, factor analysis collapses together skill factors which are highly collinear and isolates the key dimensions of variation. This does allow for the differential weighting of different skill dimensions, but the basis for these weights is a clustering exercise with unclear external validity. In other words, the weights given to each skill dimension in factor analysis do not necessarily correspond to how important each skill dimension is in any specific overall sense of skill overlap or human capital portability; they simply correspond to the primary dimensions of variation in the skill data itself, which may be more a product of how the data are collected than a useful norm of this data in applications to specific economic questions. Other key limitations of all of these measures are: 1) they rule out the possibility of interactions or otherwise non-linear functions of skill dimensions, and 2) they by definition force skill distance to be directionally symmetric (distance from occupation A to B equals distance from B to A).

The existing distance measures are detailed below. Note $q_{j,o}$ is skill value q for skill dimension j and occupation o.

- Euclidean Distance: $\sqrt{\sum_{j=1}^{J}(q_{j,o}-q_{j,o'})^2}$
- Angular Separation: $\frac{\sum_{j=1}^{J} q_{j,o} \cdot q_{j,o'}}{\sqrt{(\sum_{j=1}^{J} q_{j,o}^2)\cdot(\sum_{k=1}^{J} q_{k,o'}^2)}}$
- Factor Analysis: $\sum_{i=1}^{4} f_i \cdot \omega_i$ for f = skill factor, ω = factor loading

3.5.2 Sorting Model of How Switching Behavior Relates to Skill Distance

In order to improve on existing measures of skill distance between occupations, observed worker flows can be used as a disciplining empirical measure. The idea being that observing

which occupation switches are commonly made carries empirical content about which skills are transferable. In order to understand the relationship between worker flows and occupation and pair specific factors (of which skill distance is one), existing theory provides a useful framework to build on.

I adapt the approach of Cortes and Gallipoli, 2018, who develop an Eaton, Kortum, Roy style model of occupation selection and switching. This is a static model focused on the partial equilibrium dynamic of occupation selection. Workers are heterogeneous in exogenously determined initial occupation, and decide whether to switch to another occupation. They face directional pair-specific iceberg costs of switching, and match quality shocks generated by a Frechet distribution determine the *offers* from other occupations. A worker makes a switch if the payoff of doing so dominates that of any other switch (including no switch); payoffs are based on individual characteristics, a measure of fixed utility of the potential occupation, the match quality shock and the iceberg cost of the switch.

In order to study the flows of workers across occupations, Cortes and Gallipoli, 2018 solve this model for the probability of each occupation transition; namely the probability that a given occupation switch dominates all other switches and not switching. Dividing this probability by the probability of staying (not switching occupation), the authors reach a reduced form equation relating the share of switchers to stayers for a given directional occupation pair to factors specific to the origin and destination occupations, and a pair specific switching cost factor, as shown in equation 1. Here, $s_{o,o'}$ is the number of switches from occupation o to o'. α and λ are occupation specific fixed effects, and $Skill\ Dist_{o,o'}$ is the part of the directional pair-specific cost that is generated by differences in human capital demands.

$$ln(\frac{s_{o,o'}}{s_{o,o}}) = \alpha_o + \lambda_{o'} + \beta \cdot Skill\ Dist_{o,o'} + \varepsilon_{o,o'} \tag{3.1}$$

To measure skill distance between occupations, Cortes and Gallipoli, 2018 use the angular separation measure of skill distance, discussed above. My analysis departs from Cortes and Gallipoli, 2018 by arguing that existing measures of skill distance do not adequately capture the

multi-faceted skill mismatch restricting human capital portability between occupations. I assert that some portion of the pair-specific error in this equation is determined by human capital mismatch, and that their theoretical structure and estimating framework is well-suited to determine exactly what that portion is.

3.5.3 A New, Predicted Measure of Skill Distance

To explain the methodology for constructing my novel skill distance measure, I will begin with an intuitive example of the simplest version and then discuss the detail of the precise estimation procedure implemented.

Taking the Cortes and Gallipoli, 2018 equation as a starting point, I propose replacing the constructed index of skill distance with the raw data on the underlying skill dimensions. One simple way to do this would be to estimate equation 2 below, in which I include differences on each skill dimension q between source occupation o and target occupation o', and allow a unique coefficient β_j for each skill dimension j. The unit of observation here is a directional occupation pair, meaning that there are two observations for each possible pair of occupations, one in each direction. Pairs in which the source and target are the same ($o = o'$) are omitted.

In this formulation, each β represents precisely the relevance of divergence on a given skill dimension for a worker's ability to transition from the source to target occupation. In other words, if a given β is negative and large this indicates that directional occupation pairs which diverge on this skill dimension have few workers making the given switch. Alternatively, a β which is near zero indicates that divergence on the related skill dimension does not relate to worker flows between occupations. In this way, the Betas are precisely interpretable as skill dimension weights empirically targeted to observed occupational mobility. Furthermore, the partial predicted value given by $\sum_{j=1}^{J} \beta_j (q_{j,o} - q_{j,o'})$ can precisely be interpreted as an empirically targeted norm on the underlying skill data, generating a continuous measure of skill overlap. This term is the component of switching behavior unexplained by occupation fixed effects which is correlated with the pair

specific skill data[1].

$$ln(\frac{s_{o,o'}}{s_{o,o}}) = \alpha_o + \lambda_{o'} + \sum_{j=1}^{J} \beta_j (q_{j,o} - q_{j,o'}) + \mu_{o,o'} \tag{3.2}$$

While OLS is a powerful demonstrative tool to explain the method, a machine learning approach is adopted for the implementation of this method, to allow for a rich accounting of the complex non-linear structure through which occupational skill mismatch may create frictions for occupational mobility. In particular, I adopt the following two-step estimation procedure in which I first residualize switching shares on occupation fixed effects using OLS, and then use a Random Forest algorithm to predict residual switching shares based on the skill characteristics of the two occupations.

Step 1: OLS regression of switch shares on occupation fixed effects, construct residuals[2]

$$OLS: \quad ln(\frac{s_{o,o'}}{s_{o,o}}) = \alpha_o + \lambda_{o'} + \varepsilon_{o,o'} \tag{3.3}$$

$$Yielding: \quad \hat{\varepsilon}_{o,o'} = \hat{\alpha}_o + \hat{\lambda}_{o'} - ln(\frac{s_{o,o'}}{s_{o,o}}) \tag{3.4}$$

Step 2: Estimate Random Forest model of residual switches using skill data

$$EstimateRandomForest: \quad \hat{\varepsilon}_{o,o'} = RF\{q_{j,o}, q_{j,o'}, q_{j,o} - q_{j,o'}\} \tag{3.5}$$

$$Predicted\ Skill\ Overlap_{o,o'} = \widehat{RF}\{q_{j,o}, q_{j,o'}, q_{j,o} - q_{j,o'}\} \tag{3.6}$$

There are two primary limitations to this approach. First is the assumption that all workers in a given occupation hold the same skill set, or in other words that your current occupation fully describes your current skill set. While this is inherently untrue, it gives tractability to the model

[1]Note that other pair specific factors such as geographic overlap may also be captured in my measure of skill distance. The implications of this and other potential limitations are discussed further, following the presentation of the method.

[2]Due to a high degree of sparsity in the switching matrix, and the model specification of log switching share, I replace switching shares of zero with the minimum non-zero value in the data. Forthcoming robustness exercises show that the specific nonzero value used for this replacement does not substantively alter the results.

and to the existing skill data at the occupation level. A mitigating factor is that the DOT skill measurements represent the skills that workers *use* in an occupation, rather than those they need, limiting the scope to which some occupations may have low skill "requirements" but in reality most workers are far above them. Furthermore, an intuitive way of conceptualizing this simplification is to think that hiring staff consider the most recent occupation as a sufficient statistic for a worker's skills.

A second, and likely more important limitation to my approach is that other pair specific factors which are correlated with both the skill data and the switching shares may be captured in my derived measure of skill distance. For example, geographic overlap between occupation pairs may be systematically related to both sides of equation 5. While the random forest algorithm should reduce this concern by reducing out-of-sample variance through bootstrapping methods, it is not possible to rule out the possibility that other unobservable non-skill factors are contributing to my derived measure of skill overlap. Given this, it is appropriate to think of my measure of skill overlap as the component of occupational mobility driven by pair-specific overlap of skills *and correlated pair-specific* factors.

3.6 Analysis

This section begins by directly comparing my predicted measure of skill distance to previous measures, and follows by conducting various applications to explore the potential for such measures and the implications of their differences.

3.6.1 Comparing Measures of Skill Distance

Re-scaling measures to a common unit range

Because the 4 measures of skill distance being compared all use different units, to conduct comparisons I first convert them all to a measure of skill overlap/portability in the unit range [0,1]. For measures that are constructed as distance (euclidean distance and factor analysis), I construct

$skill\ overlap = -skill\ distance$. Then I shift the range of each measure such that the minimum is 0 ($shifted\ skill\ overlap = skill\ overlap - min(skill\ overlap)$), and finally divide by the maximum to reach the unit range ($unit\ skill\ overlap = \frac{shifted\ skill\ overlap}{max(shifted\ skill\ overlap)}$).

Using this standard unit range for all measures, Figure 3.2 below shows how the distributions of the measures differs. My predicted measure is the only one that intuitively identifies that the vast majority of occupation pairs have very little meaningful skill overlap (for workers considering switching), and that there is a long tail of directional occupation pairs with meaningfully high skill overlap. Interestingly, the factor analysis and euclidean distance measures both yield an approximate normal distribution, while the angular separation measure asserts that most directional occupation pairs have highly similar skill sets.

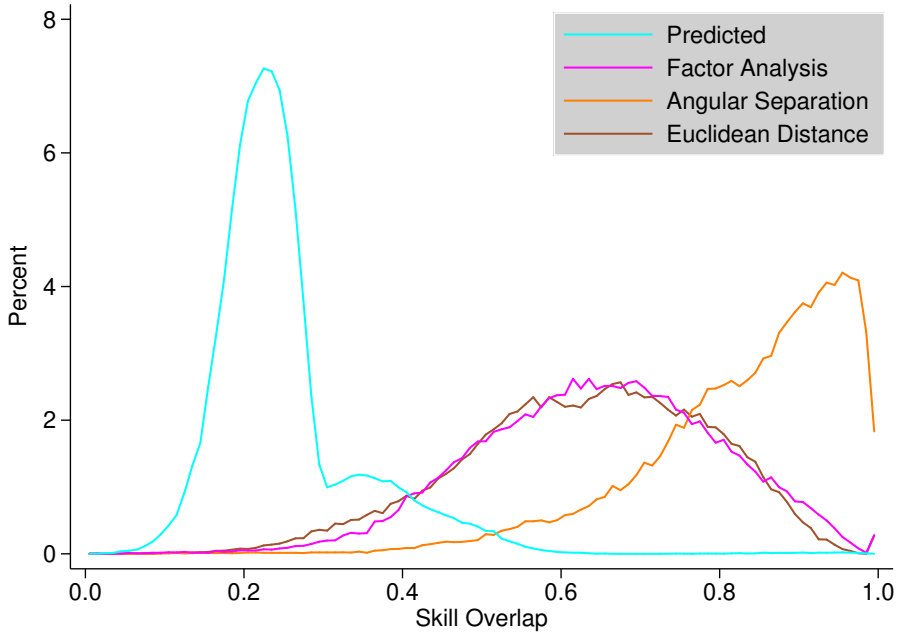


Figure 3.2: Distributions of Skill Overlap Measures on Directional Occupation Pairs

Furthermore, in simple univariate regressions of directional occupation pair switches on each skill distance measure, my predicted skill distance measure yields a much stronger fit (R^2 is up to 20 times higher). While this should be expected given that my measure is targeted to switches while others aren't, it is a supporting piece of evidence confirming the founding intuition. The second most predictive measure of switching behavior is the factor analysis skill distance,

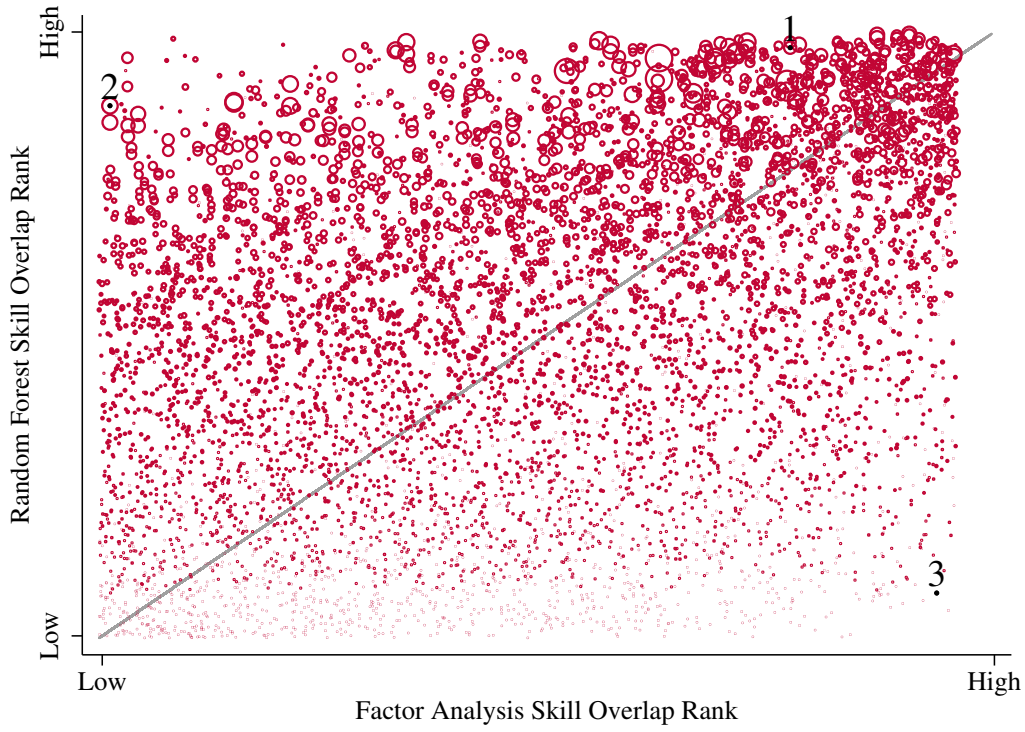


Figure 3.3: Alignment of Skill Overlap Rank Between Measures

Note: the y-axis shows the skill overlap rank from my predicted skill distance measure.

the only other measure which weights the individual skill dimensions. Figure 3.3 below shows a comparison of the predicted measure against the factor analysis measure, showing that when they disagree it is in fact because the number of switches is inconsistent with the factor analysis ranking. Each circle in this figure is a directional occupation pair, and the circle sizes correspond to number of switches (i.e. a large circle would indicate that many people made the switch from occupation A to occupation B).

A central feature on which these measures differ is the weights each assign to the different skill characteristics. Figure 3.5 plots the weights assigned to each dimension, by measure of skill distance[3]. Namely, angular separation and euclidean distance simply assign even weights to all skill dimensions, which is represented with the dotted vertical line. Weights from my predicted

[3]Weights for the predicted measure are constructed based on the standard importance measure of increase in node purity in a model on differences in skill dimensions. For the factor analysis, the weights are the weighted sum of each factor's loadings, weighted by the share of variation explained by the corresponding factor.

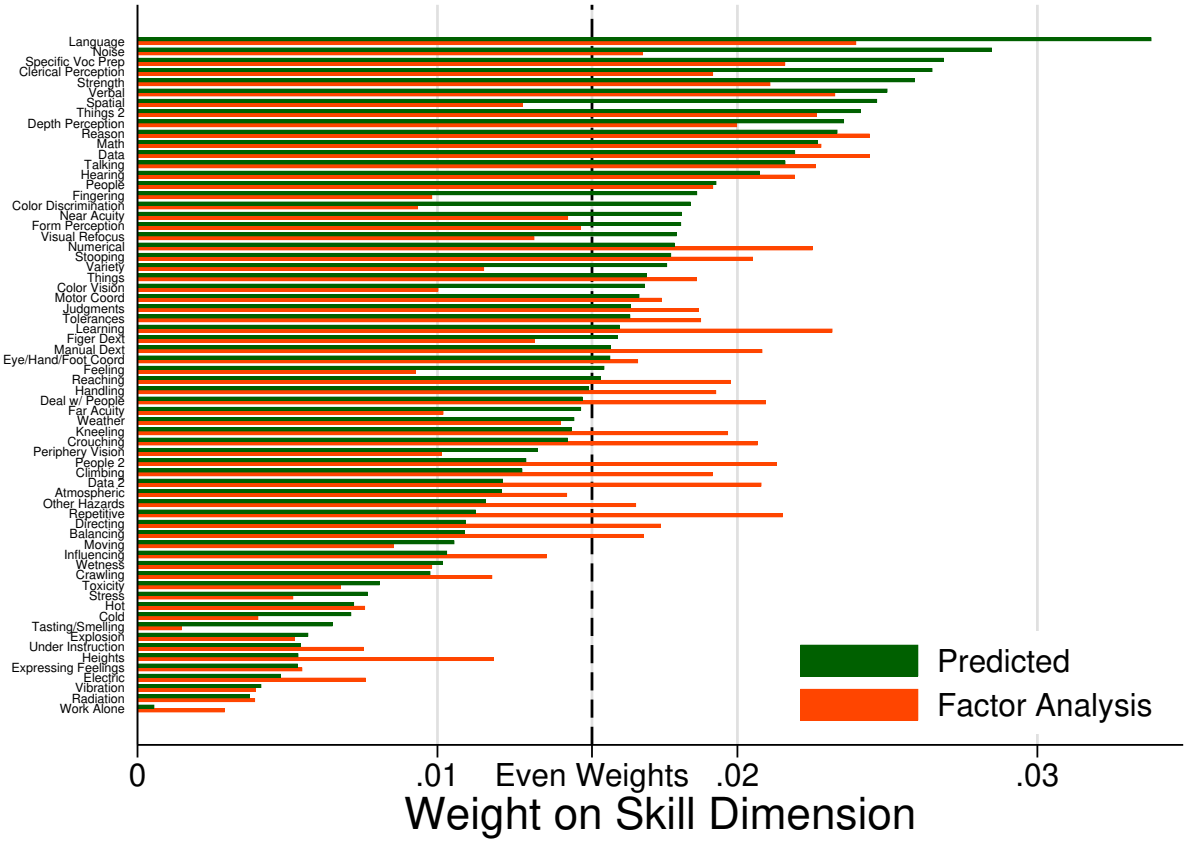


Figure 3.5: Weights of Skill Dimensions Across Measures

measure are shown in green, and skills are ranked in order of the weights I assign. The figure shows that language ability is the most important skill factor determining occupational mobility in my measure, while the ability to work alone is the least. Other highly important factors include duration of occupation specific training, physical strength, verbal and spatial skills, math and reasoning education, and communicating with others. Least important are environmental measures such as exposure to extreme temperature or hazardous materials, along with specific abilities such as taste, expressing feelings, and working alone.

Comparing my measure to those with even weights, we immediately see that observed switching behavior indeed implies that some skill dimensions are substantially more important than others. Wee see that the range of weights spans from very nearly 0 to approximately twice the weight if all are weighted equally. In comparing the predicted skill distance metric to the

factor analysis, the weights coarsely align with some notable exceptions. For example, while the predicted measure ranks verbal ability among the most important skills, factor analysis concludes that it's less important that the average dimension. An example in the opposite direction is the temperament to perform repetitive tasks; the predicted measure ranks this in the bottom third of skills, while factor analysis weighs it among the top third. While these comparisons do not impart an objective criteria to judge the measures against each other, they do show that empirical content implies non-equal weights, and particularly that adding observed switches to these data leads to substantial changes in the relative weights on skill dimensions.

3.6.2 Sensitivity of Prior Results to Skill Distance Measure

A leading result in both Poletaev and Robinson, 2008 and Gathmann and Schönberg, 2010 is that displaced workers who switch skill profiles experience significantly larger wage loss than those who switch to an occupation with similar skills. To examine this finding across different measures, and using the latest available data in the US, I construct a repeated time series of displaced workers in the DWS from 1986 to 2018. Focusing on those finding new work in an occupation other than that of their original displacing job, I regress the log change in wage from the displacing to the re-employing job onto an indicator variable for making a large skill switch using each measure[4]. Table 3.1 shows the results of this analysis. We see that despite the substantial differences between these measures of skill distance, in this coarse grouped analysis their implications are qualitatively very similar. The results are also similar across measures if regression on normalized skill portability, rather than this binary specification, suggesting that the coarseness is not driving the similarity.

The other primary result relating to the role of skill portability as an occupational mobility friction is one from Cortes and Gallipoli, 2018. In their analysis, the key parameter of interest is the coefficient on skill distance in the regression of switching share on occupation fixed effects

[4]For each measure, a value of 1 is assigned to those making switches above the 67 percentile of skill distance, and a value of 0 for those making switching below the 33 percentile. The middle third of skill distance switches are omitted.

Table 3.1: Wage Losses Associated with Skill Switching

	Predicted	Factor Analysis	Angular Separation	Euclidean Distance
Skill Switch	–0.07***	–0.05***	–0.06***	–0.06***
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	–0.07***	–0.09***	–0.08***	–0.08***
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	15028	15020	15019	15025

and skill distance. They argue that this coefficient is an indication of the extent to which skill factors could be causing mobility frictions. To test how robust their result is to the use of alternate definitions of skill distance, I reproduce their analysis with each of the four measures I consider[5]. This analysis is represented in Figure 3.6.

Overall these results show that the coefficient on skill distance has become more negative over time, across all measures. A key difference between measures is the magnitude of this coefficient, or intuitively the scope of skill switching frictions. This level is a central result of their analysis, in which they use the angular separation metric of skill distance. My results in this figure show that on this dimension, angular separation is a stark outlier, creating a much more negative coefficient than my predicted measure, factor analysis, and euclidean distance. This finding suggests that the results of Cortes and Gallipoli, 2018 may substantially overstate the importance of skill distance as a mobility friction.

3.6.3 Aggregate Trends in Skill Portability

Another way to test the hypothesis that decreases in overall skill portability may be exacerbating frictions to occupational mobility is to compute annual measures of overall skill portability. Combining the measures of skill distance (inversely, portability or overlap) by directional occupation pair, with employment counts by occupation, I strive to characterize the number of outside options a worker in a given occupation has. I define this as the number of jobs outside their occu-

[5]Note that the data I analyze is not identical to Cortes and Gallipoli, 2018 in several ways, including differing restrictions on sample inclusion based on age and occupation, as well as different formulation of what is considered an occupation switch in the data.

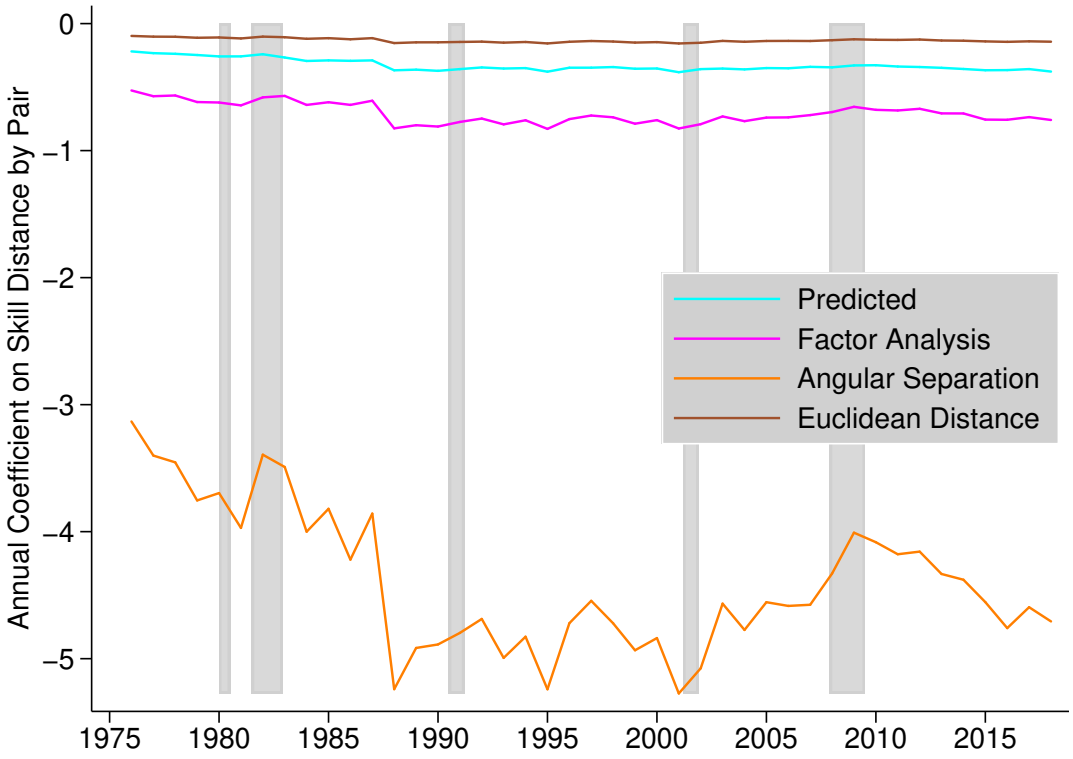


Figure 3.6: Trends in Regression Coefficient on Skill Distance

pation which require "similar" skills.

Having re-scaled the skill overlap measures to [0,1], they intuitively serve as a measure of this similarity. Taking a weighted sum of all outside jobs, in which the weights are the skill overlaps, outside occupations with very different skill sets (overlap near 0) will contribute little to nothing to the outside options of workers in the source occupation. Alternatively, for outside occupations with high skill overlap (near 1), nearly all of the jobs will be counted as outside options. The equation below illustrates this, with $P_{o,o'}$ as the unit skill overlap, and $employment_{o'}$ as the employment in each outside occupation. Note that employment and occupation-level skill portability are subscripted with year, re-iterating that this statistic is calculated by occupation and year, and that the variation across years is driven only by variation in employment levels across years.

$$\text{Occupational Skill Portability: } P_{o,y} = \sum_{o' \neq o} P_{o,o'} \cdot employment_{o',y} \tag{3.7}$$

With these measures of overall skill portability, or alternatively overall outside options, for each occupation, I use employment shares to aggregate up to a group of occupations. Taking the full labor market as the group, this amounts to a weighted sum of skill portability by occupation, in which the weights are the employment shares. Equation 8 shows this explicitly. The same procedure is used to construct skill portability by occupation groups smaller than the full labor market, the only difference being the the employment denominator varies depending on the group being considered. Equation 9 shows this.

$$\text{Labor Market Skill Portability: } P_y = \sum_o P_{o,y} \cdot \frac{emp_{o,y}}{emp_y} \tag{3.8}$$

$$\text{Occupation Group Skill Portability: } P_{y,g} = \sum_o P_{o,y} \cdot \frac{emp_{o,y}}{emp_{y,g}} \tag{3.9}$$

A central limitations of this aggregated analysis is that the only variation available is changes in occupational employment counts. While changes in the skill composition of each occupation over time may also substantially contribute to changes in the skill-based outside options of workers, because I only use the single iteration of the DOT from 1991 I cannot capture changes in skill content. Therefor, my aggregate results only reflect an unknown portion of overall shifts in outside options. A second important limitation is that demand in outside occupations is proxied for with employment count in that occupation; a measure of the number of vacancies would be preferred here. A further issue is that due to insufficient sample sizes, using the CPS to construct annual employment counts at the occupation level is an inherently coarse measure, particularly for smaller occupations. Finally, also due to limited sample and geographic coding in the data, this analysis is all done nationally; a similar analysis only considering outside options within state or region would likely better capture geographic components of overall occupational skill portability.

Keeping these important limitations in mind, Figure 3.7 shows the trend in P_y for each

underlying measure of skill distance. Consistent with increasing labor force size, we see that outside options have been increasing over time. In aggregate, the different measures agree that for those employed, outside options have increased faster than the general labor force growth. These measures of aggregate skill portability changes only disagree on the degree of portability growth, with my predicted measure yielding growth rates roughly 7.3 percentage points higher than the lowest measure (89% vs 81.7% growth) over a 43 year period. This difference is due to both different rankings of directional occupation pairs, and to a highly skewed distribution of predicted portability (a long right tail of portability) relative to the other measures. These findings show that compositional changes in the labor market are consistent with a *decrease* in the overall occupational mobility friction being generated by skill overlaps between occupations, indicating that this channel may in fact be improving the ability of workers to find re-employment after job separation.

By census occupation categories (6 total categories excluding not including military and unemployed), we see in Figure 3.8 that the outside options have been growing differentially by occupation group. Professionals have seen the largest growth, followed by Support workers; Farm/-Forestry and Fishing workers, Craft workers and Operators/Laborers have seen roughly equivalent changes, all experiencing the lowest outside-option growth. Because occupational mobility is generally much higher within occupation groups than between, a portion of these differential trends by category can be explained by changes in labor market composition shares between these categories over time. This is clearly a major contributor for Professionals, but for Support Occupations there appear to be outside option gains from growth in the closely linked Professionals group. Figure 3.9 shows these changes in occupation group employment shares over time.

Ultimately, these aggregated analyses of skill portability are limited by the quality and detail of the underlying data, both of which are substantial concerns with the current data. These aggregation methods would be much more informative in a setting with repeated measures of skill content by occupation, more direct measures of demand by occupation, and detailed geographic data. All of this ingredients appear to be present in the case of the German dataset used by Caldwell

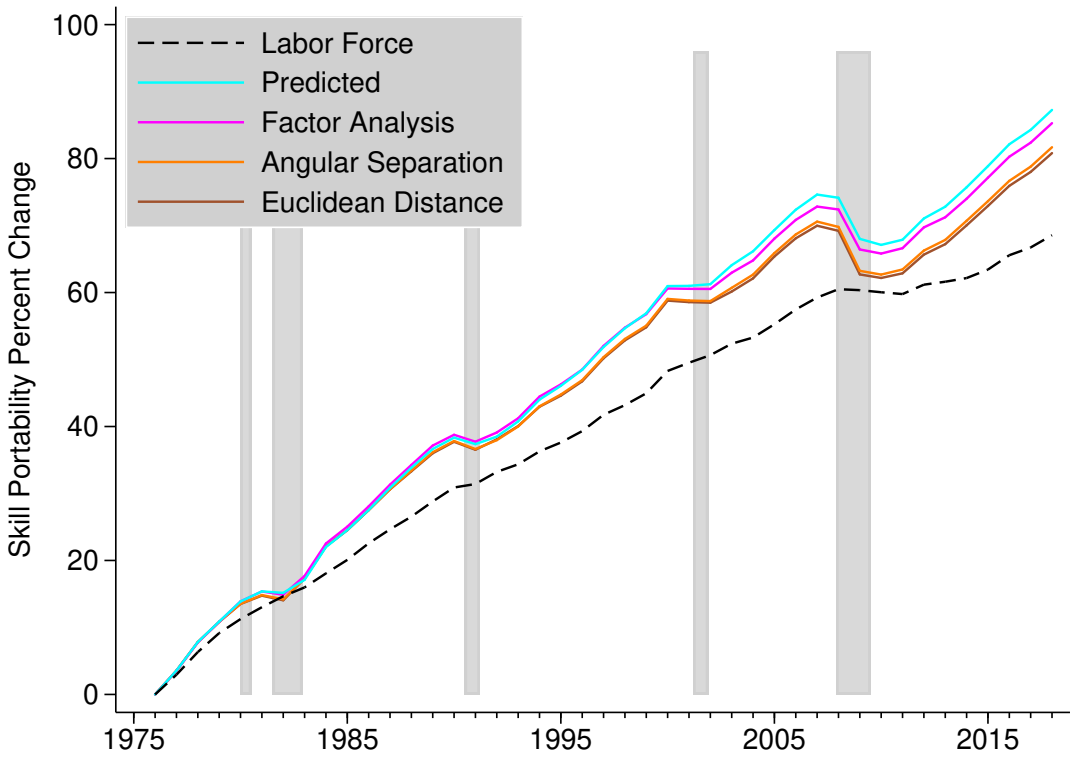


Figure 3.7: Aggregate Skill Portability Percent Changes, by Measure

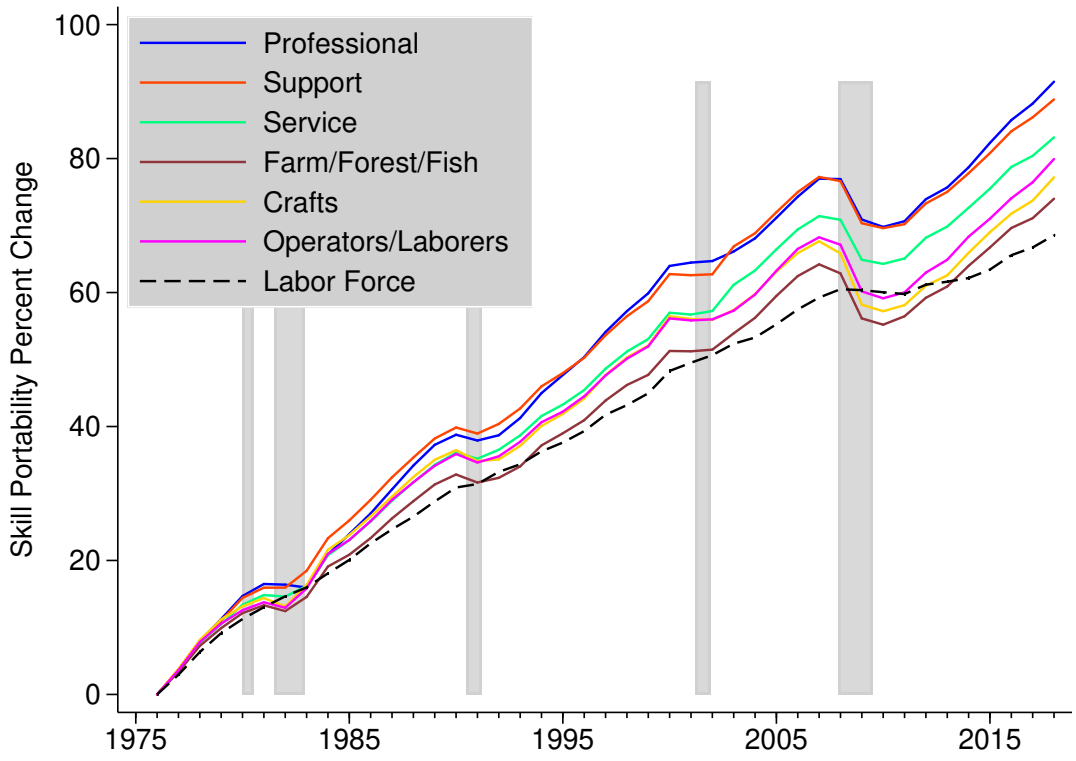


Figure 3.8: Occupation Group Skill Portability Percent Changes, Predicted Measure

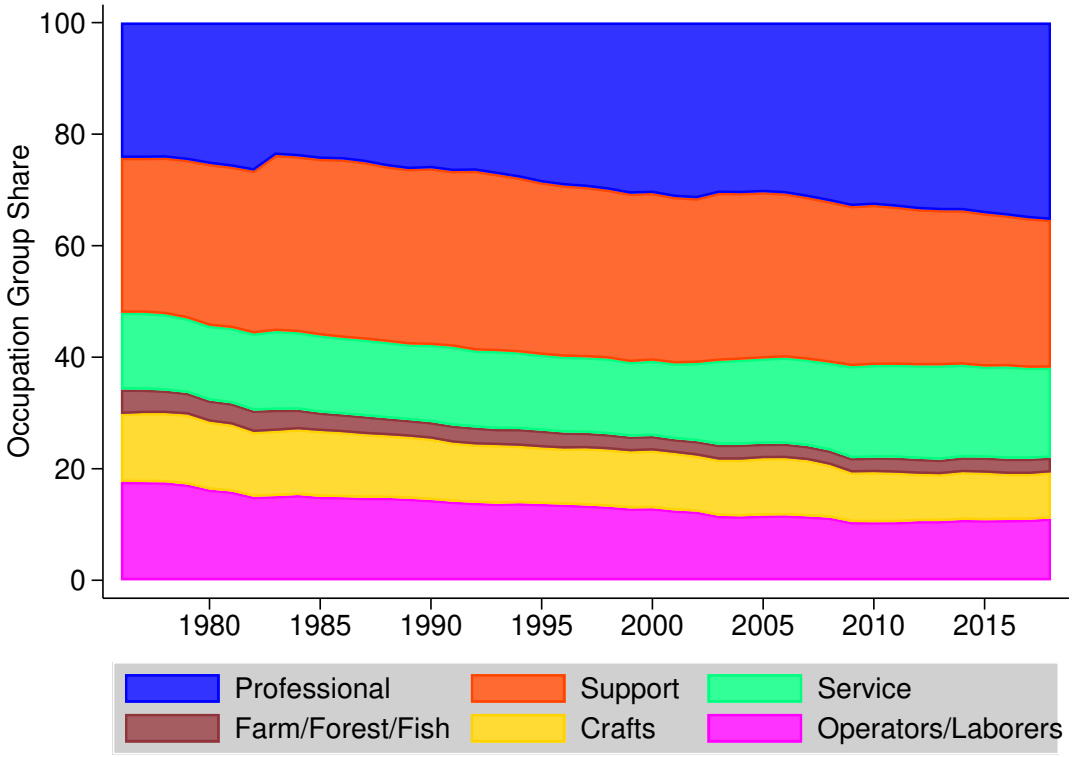


Figure 3.9: Trends in Employment Shares

and Danieli, 2019, which would serve as a promising avenue for further exploration.

3.6.4 Network Analysis of Skill Linkages

A final application of skill distance measures considered is the role of the occupation network structure on labor market outcomes by occupation. One network analysis hypothesis I test is whether recessions impacting central occupations in the network have disproportionately large overall negative impacts relative to those primarily hitting peripheral occupations. I define a central occupation as one with relatively high in-portability: for each occupation in-portability is an aggregate of the skill portability from all other occupations into the base occupation.The idea here being that if a recession impacts key receiving occupations, the labor mobility impact will be greater because more workers would have relied on this receiving occupation for work. This idea originates from Charles et al., 2016, who show that leading up to the Great Recession, some workers in manufacturing occupations were finding new work in construction roles. While construction served as a receiving occupation for those dislocated workers, when the recession hit construction employment plummeted, and this key transition opportunity disappeared for manufacturing workers.

To study these dynamics in a unified framework spanning recessions and occupations, I construct an index measure of the occupation network centrality of each recession. Weighting skill portabilities from each occupation into a base occupation by the employment in each source occupation, I first construct a measure of skill centrality for each occupation and recession, shown in equation 10. Then I aggregate these measures into a network centrality incidence measure by recession, by summing them up weighted by respective occupation shares of total employment losses from peak to trough, in the given recession, as shown in equation 11.

$$\text{Occupation Skill Centrality by Recession: } Centrality_{o,r} = \sum_{o' \neq o} P_{o',o} \cdot employment_{o',y} \qquad (3.10)$$

$$\text{Recession Skill Centrality Incidence: } Centrality_r = \sum_o Centrality_{o,r} \cdot \frac{emp\ loss_{o,r}}{total\ emp\ loss_r} \quad (3.11)$$

Plotting these by recession centrality incidence measures against the overall peak to trough percent change in employment, Figure 3.10 shows the results of this exercise. We see that recessions with job losses concentrated in central occupations have resulted in substantially larger total employment impacts. The correlation between these factors is approximately -0.8, indicating a strong negative relationship.

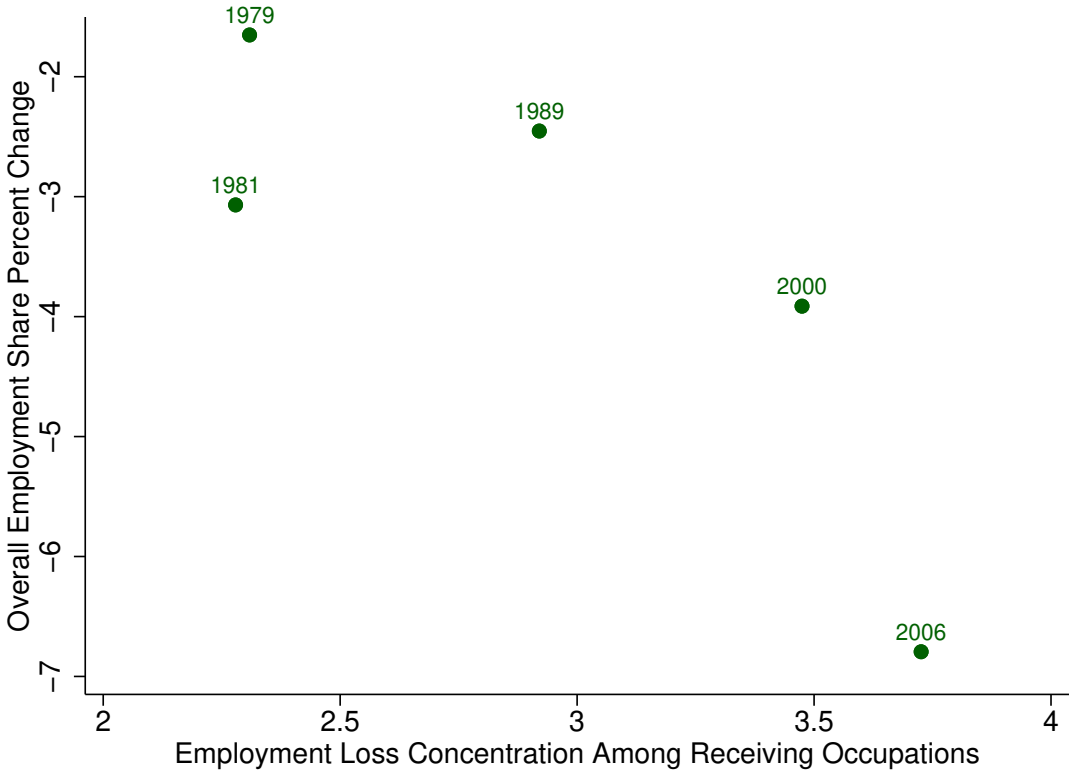


Figure 3.10: Recession Depth by Centrality of Occupations Impacted

A second network analysis exercise I conduct relates occupation level recessionary displaced worker outcomes to job losses in neighboring occupations. As shown in equation 12, the job losses in each neighbor are weighted by skill portability measures to construct an aggregate measure of neighborhood job losses for each occupation in each recession. I compare these exposure

measures to several re-employment outcomes of displaced workers in Table 3.2. These regressions show that recession incidence on neighboring occupations only appears to significantly relate to labor force dropout rates by occupation, and do not significantly correlate with non-employment rates, unemployment insurance exhaustion, or displacement shares.

$$\text{Occupation Neighbor Job Losses by Recession: } Centrality_{o,r} = \sum_{o' \neq o} P_{o,o'} \cdot \frac{emp\ loss_{o',r}}{total\ emp\ loss_r} \tag{3.12}$$

Table 3.2: Re-employment Outcomes by Occupation and Neighborhood Recession Incidence

	Dropout	Non-Emp	UI Exh	Displaced
Employment PC	–0.13**	–0.12	–0.31***	0.06
	(0.07)	(0.09)	(0.11)	(0.06)
Share Losses	3.04	–1.53	5.77	2.54
	(4.82)	(6.66)	(7.40)	(4.24)
Neighbor Share Losses	0.82*	–0.35	0.00	–0.17
	(0.42)	(0.58)	(0.69)	(0.37)
Share Losses × Neighbor Share Losses	–14.74	4.77	–27.43	–12.87
	(21.79)	(30.13)	(33.57)	(19.18)
Constant	–0.08	0.39***	0.41**	0.16*
	(0.11)	(0.15)	(0.18)	(0.09)
Observations	426	426	358	425

Taken together, these network analysis exercises suggest that high-quality skill distance measures are critically important for studying occupation network impacts of business cycles, and that there is some evidence that network factors matter for employment outcomes by occupation during recessions. It is also important to caution that these analyses are highly descriptive and lack a rigorous identification strategy, meaning that their results should be interpreted cautiously as merely suggesting future avenues of continued research.

3.7 Conclusion

This paper expands on methods for both constructing skill distance measures, and studying applications of them. I show that the previous literature has focused on a narrow set of measures with important limitations. Adopting a predictive framework and incorporating observed switching behavior as the target, I construct a measure of skill distance which resolves many of the existing limitations. This allows for directional measurement of skill distance in a non-parametric fashion, weighting dimensions of skill based on relevance to occupational mobility.

By comparing skill distance measures, I am able to study the empirical properties of each. Notably, my constructed measure accurately captures the notion that the vast majority of directional occupation pairs do not have a significant degree of overlap, and that there is a long right tail of occupation pairs which do. It further captures empirical weights for each skill dimension which are intuitively sensible and broadly align with the leading existing measure of factor analysis. In fact, the two are very similar except in cases when observed switching behavior strongly contrasts with the factor analysis measure, in which cases my predicted measures aligns with the switching patterns.

I study economic impacts using these measures in both existing and network analysis based applications. I find that my predicted measure and factor analysis capture similar economic impacts, while angular separation yields outlying results in a leading application. Testing two hypotheses of how an employment shock flows across the occupation network, I show that predicted skill distance has sufficient power to capture systemic and intuitive economic properties.

Much further work on these topics and methods is warranted. First, as all skill distance measures (in particular my predicted one) rely heavily on the underlying skill and employment data, collecting more detailed data to use in applications would be fruitful for all related inquiries. Ultimately, these measures would also be well-suited in a macroeconomic model on these topics, as is the approach in Guvenen et al., 2019. With these improvements in data and modelling, the leading application would be to work on predicting which workers (occupations) are most vulnerable

to negative demand shocks. By conducting such prediction, public policy and services can target workers facing the toughest re-employment prospects, essentially having the largest marginal impact on mitigating labor adjustment costs from cyclical and structural fluctuations in demand. One way to implement and test such an approach would be to experimentally deploy leading measures of skill portability to job-sites and study outcomes of match quality among treatment groups.

Bibliography

Autor, D. H., F. Levy, and R. J. Murnane (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics* (November), 1279–1333.

Caldwell, S. and O. Danieli (2019). Outside Options in the Labor Market. *Working Paper*.

Charles, K. K., E. Hurst, and M. J. Notowidigdo (2016). The Masking of the Decline in Manufacturing Employment by the Housing Bubble. *Journal of Economic Perspectives 30*(2), 179–200.

Cortes, G. M. and G. Gallipoli (2018). The Cost of Occupational Mobility: An Aggregate Analysis. *Journal of the European Economic Association 16*(2), 275–315.

Gathmann, C. and U. Schönberg (2010). How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics 28*(1), 1–49.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2019). Multidimensional Skill Mismatch. *AEJ: Macro, forthcoming*.

Miller, A., D. Treiman, P. Cain, and P. Roos (1980). Work, jobs, and occupations: A critical review of the Dictionary of Occupational Titles. *National Research Council*.

Poletaev, M. and C. Robinson (2008). Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984–2000. *Journal of Labor Economics 26*(3), 387–420.

Chapter 3 is currently being prepared for submission for publication of the material. Khachiyan, A. The dissertation author was the primary investigator and author of this material.