Title
Essays in Labor Economics

Author
Li, Nicholas

Publication Date
2019

Peer reviewed|Thesis/dissertation

Essays in Labor Economics

by

Nicholas Y. Li

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Card, Co-chair
Professor Frederico Finan, Co-chair
Professor Patrick Kline
Professor Jesse Rothstein

Spring 2019

Essays in Labor Economics

Abstract

Essays in Labor Economics

by

Nicholas Y. Li

Doctor of Philosophy in Economics

University of California, Berkeley

Professor David Card, Co-chair

Professor Frederico Finan, Co-chair

This dissertation applies tools developed in labor economics to empirically study questions in labor, development, and urban economics. Each chapter attempts to decompose a problem into competing explanations. The first decomposes racial segregation in US cities. The second decomposes differences in wages between agricultural workers and and non-agricultural workers. And finally, the last decomposes the heterogeneous responses of workers to a new monitoring technology.

In the first chapter, I revisit the question of whether residential segregation in US cities emerged in the mid-twentieth century as a consequence of decentralized location choices in combination with white antipathy toward black residents or whether it reflected institutionalized constraints on the availability of neighborhoods that black families could access. The chapter analyzes rich population data from the 1930 and 1940 censuses to disentangle these channels. I first lay out a simple discrete choice model of residential choices by white and black families that depends on the local price of housing and on the fraction of black residents in each neighborhood. I show how the preferences of both race groups can be identified using information on the impacts of exogenous inflows of white and black residents to different neighborhoods. White and black rural inflows constituted a major source of inmigration to major cities during this time period; I construct a pair of novel instrumental variables for these inflows by connecting the distributions of white and black surnames in rural areas to earlier migrants living in different census tracts in 1930. The resulting structural estimates confirm that white families had a relatively high willingness to pay to avoid black neighbors, consistent with an important role for preferences in the evolution of neighborhood segregation. Combining white and black preferences, however, I also find strong evidence that black residents faced supply side constraints on their neighborhood choices. I conclude that about one half of the overall degree of neighborhood segregation observed in 1940 was due to the different preferences of white and black families, while a comparable share was due to implicit or explicit constraints on which neighborhoods black families could move into.

While the first chapter interpreted incumbent residents' responses to migrants as reflective of their racial preferences, the second chapter, based on joint work with Marieke Kleemans, Joan Hamory Hicks, and Ted Miguel, directly studies the migrant experience itself. Recent research has pointed to large gaps in labor productivity between the agricultural and non-agricultural sectors in low-income countries, as well as between workers in rural and urban areas. Most estimates are based on national accounts or repeated cross-sections of micro-survey data, and as a result typically struggle to account for individual selection between sectors. We use long-run individual-level panel data from two low-income countries (Indonesia and Kenya). Accounting for individual fixed effects leads to much smaller estimated productivity gains from moving into the non-agricultural sector (or urban areas), reducing estimated gaps by over 80%. Estimated productivity gaps do not emerge up to five years after a move between sectors. We evaluate whether these findings imply a reassessment of the conventional wisdom regarding sectoral gaps, discuss how to reconcile them with existing cross-sectional estimates, and consider implications for the desirability of sectoral reallocation of labor.

Finally, the third chapter is based on joint work with Ernesto Dal Bó, Frederico Finan, and Laura Schechter and empirically studies models of task assignment within organizations in a developing country context. Standard models of hierarchy assume that agents and middle managers are better informed than principals about how to implement a particular task. We estimate the value of the informational advantage held by supervisors (middle managers) when ministerial leadership (the principal) introduced a new monitoring technology aimed at improving the performance of agricultural extension agents (AEAs) in rural Paraguay. Our approach employs a novel experimental design that, before randomization of treatment, elicited from supervisors which AEAs they believed should be prioritized for treatment. We find that supervisors did have valuable information—they prioritized AEAs who would be more responsive to the monitoring treatment. We develop a model of monitoring under different allocation rules and roll-out scales (i.e., the share of AEAs to receive treatment). We semi-parametrically estimate marginal treatment effects (MTEs) to demonstrate that the value of information and the benefits to decentralizing treatment decisions depend crucially on the sophistication of the principal and on the scale of roll-out.

To my Pucci.

Contents

List of Figures

List of Tables

Acknowledgments

I owe a debt of gratitude to all the people who helped save me from myself.

Dave: *you* may never understand why I can't draw a straight line from point A to point B—me neither. *But, I* may never understand what inspired you to rescue me from spinning my wheels and doing everything *except* the obvious thing. Needless to say, I would not have rolled across the finish line without your guidance and support.

Fred: I also do not know what you saw in me when you "allowed" me to hand merge two datasets of almost a million names in undergrad. But, working with you gave me a chance, and you have had my back since the beginning throughout my studies. I would not have gotten here without you.

I also owe thanks to Chris, Enrico, Ernesto, Jesse, Pat, and Ted for thoughtful conversations and advice. Beyond the always-constructive feedback that I clung to, the diversity of your perspectives taught me to not only articulate my own ideas using different rhetoric to different audiences, but clarified my own understanding of my own points. You deserve none of the blame for the economist I've become, but I have exemplars to aspire to from seeing you teach and absorbing your thoughts from seminars.

Thank you also to my other collaborators Joan, Laura G., Laura S., and Marieke. I learned so much from working with you and hope you'll continue to tolerate me in the future.

Thanks also to my past and present officemates: Caitlin, Carla, Caroline, Francis, Jonathan, Maxim, Murillo, and Peter for enduring the sprawl of my mess, my crazy ideas, and the persistent smell of brewing coffee.

I need to also thank my contemporaries, stronger evidence of peer effects than any regression analysis. Daniel, I admire how you see insight and nuance and beauty in models that are beyond me. Evan, you have given me so much from intramural sports, to surfing, to data entry assignments. You've been a true friend with the rare ability to turn my incoherence into some semblance of logic. Juli, you are a force of nature and a prominent economist in the making. I feel blessed to have learned from you, and your friendship has been a source of stability throughout the volatility of the PhD. Yotam—another kindred spirit in left-handed thinking as Dave would say—thank you for all the questions-turned-debates about statistics and econometrics as well as sharing with me all your other milquetoast and uncontroversial opinions. To all of you, I only ask that you remember us little people.

Thanks also to Avner, David S., Erin, Greg, Jon L., Jon S., Jonas, Julien, Kevin, Raffa, Zarek, and many others whose absence would have made my PhD a profoundly different and frankly worse experience.

To all the friends outside of the economics bubble, I ask for forgiveness for being absent and locked in my office for the last few years. Alicia and Stephen: thank you for persistently inviting me out to beer, bike rides, and ski trips, and I regret not joining for more of them. Chris, David, and Ryan: my visits to the south (and north) bay became less and less frequent, but I always enjoy seeing you for our game nights.

I need to thank my parents and my sister Andrea who have continued to support me throughout the years. If anyone has endured years of me not doing something which they consider obvious, it's you. Mom and dad: while you deserve at least a tiny bit of the blame for my inherited stubbornness, I have tried to follow your examples in being generous and maintaining perspective on the most important aspects of my life. And Andrea: I live vicariously through your adventures and am inspired by your toughness.

Finally: Ale, if anyone deserves credit for saving me from myself, it's you. For the past several years, you have endured my quixotic pursuit of a miracle. You have resisted my tendencies to redirect my obsessive compulsions toward your work and your problems. You forced me to seek help when I was trying to flounder alone. You have prevented me from burning out completely by forcing me to take breaks. And, you have taken care of me. All the while, you coolly managed your own research and responsibilities. I stand in awe of you. You are more brilliant than you realize. I don't know how to fully express both my gratitude for everything you do and my confusion for why you bother with me. I'm so excited about our next adventure, and I will try not to seek out more windmills.

What I did to deserve all of you and your generosity is beyond me. But, you have my thanks all the same.

Chapter 1

Housing Market Channels of Segregation

1.1 Introduction

Higher rates of residential segregation are associated with worse educational outcomes for black children and persistent problems through adulthood, including lower employment and earnings.[1] Despite the wealth of findings on this correlation, researchers have been cautious in their recommendations for policy. Cutler and Glaeser (1997), for example, conclude, "[i]t may be that widespread social changes in attitudes toward minorities and housing choices will be required before equality of outcomes can finally be achieved."

Implicitly, this caution derives from the prevailing view that segregation is an indelible feature of cities driven by the preferences of whites to avoid neighborhoods with a substantial presence of minorities (Schelling, 1971, 1978). The role of white preferences in driving segregation has been corroborated by "white flight" following school desegregation efforts (e.g., Coleman, Kelly, & Moore, 1975; Reber, 2005) and by studies of rapid "tipping" of neighborhood racial shares in response to minority inflows (Card, Mas, & Rothstein, 2008). Recent research on the impacts of the Great Migration (Boustan, 2010; Shertzer & Walsh, 2016) suggest similar reactions throughout the twentieth century. Nonetheless, in their landmark study of the rise of black-white residential segregation over the twentieth century, Cutler, Glaeser, and Vigdor (1999) argued that segregation first arose as a coordinated effort to constrain the housing supply to black residents.[2] It was only reinforced in the latter half of the century by the decentralized decisions of white residents who fled inner city neighborhoods with growing minority shares.

Although detailed case studies have documented the existence of formal and informal constraints on the housing market options available to black families (see e.g. Ondrich,

[1]See e.g., Cutler and Glaeser, 1997; Massey and Denton, 1993; the 1966 Coleman Report; Chetty and Hendren, 2018a, 2018b; Chetty, Hendren, Kline, and Saez, 2014.

[2]Recently, Rothstein (2017) has underscored the role of government policies in promoting and enforcing racial segregation.

Stricker, & Yinger, 1998, 1999; Yinger, 1986), to the best of my knowledge, there has been no systematic attempt to quantify the separate contributions of "demand-based" explanations for segregation (i.e., explanations based on individualistic choices made by white and black families) from "supply-based" explanations (i.e., explanations based on institutions as well as extralegal threats of violence that restrict the choices available to black families). In part, the exercise has been hampered by the absence of credible identifying information that can make it possible to separate the effects of housing prices and neighborhood racial composition on the housing demands of white and black families. A related requirement is a modeling framework rich enough to specify the choices of black families *in the absence of any (non-price) constraints.* Indeed, state of the art models of housing choices (see e.g. Bajari & Kahn, 2005; Bayer, Ferreira, & McMillan, 2007; Bayer, McMillan, & Rueben, 2004) fit to contemporary data typically assume that each family can freely choose among housing units, given their incomes and preferences.

This paper attempts to provide a credibly identified, quantitative summary of the contributions of supply and demand based explanations for the patterns of racial segregation in large US cities in 1940, just before segregation became a permanent, entrenched fact of the American urban landscape. Over the previous decade, many cities had experienced large influxes of white and black rural migrants—many of whom followed the path of earlier migrants from specific origin counties (e.g., counties near the Mississippi River Delta) to specific destinations (e.g., the South Side of Chicago).[3] I use the predictable component of these migrant inflows as a source of identifying variation that allows me to estimate simple structural models of the neighborhood preferences of white and black families, separating families into broad occupation groups. I then use the resulting estimates and the structure of the model to identify counterfactual neighborhood demands of black residents in the absence of non-price constraints. This allows me to quantify the components of observed segregation attributable to the differential preferences of whites and blacks (demand side) and the non-price constraints faced by black residents (supply side).

I begin my analysis by setting up a multinomial logit model of neighborhood choice where families have preferences over the local price and black share of the neighborhood. This model is the basis of my empirical analysis. I show that under some additional assumptions, exogenous migrant demand shocks affect prices and the neighborhood black share. Crucially, the model also predicts that these effects materialize differently for more and less black neighborhoods, heterogeneity that ultimately provides part of my identifying variation.

The model provides a theoretical basis for using migrant demand shocks to perturb existing sorting equilibria and recover the preference parameters. Because migrants' endogenous neighborhood choices may reflect unobserved neighborhood changes, I develop shift-share instruments for inmigration from rural counties built on the fact that migrants are attracted enclaves of past migrants (Altonji & Card, 1991; Card, 2001). To overcome the lack

[3]Lieberson (1980), Wilson (1987) argue that rural black migrants received focused animosity from whites in the early part of the 20th century. Anti-immigrant settlement subsided following following the legislated curtailments of migration from Europe and Asia, but competition from black migrants from rural counties drew the ire of urban white residents.

of origin county information in the data, I connect migrants living in census tracts in 1930 to origin counties on the basis of their last name. I show that surname distributions are highly clustered and provide a strong signal of one's county of birth within a state. The surname-predicted based on *pre-1930* migrant settlement patterns are highly predictive of actual county-to-census tract flows in the 1935–40 period.

Instruments for rural migration in hand, I take the model to the data. I control for unobserved neighborhood heterogeneity by estimating a series of first-differenced regressions by census tract.[4] I show that the instruments' reduced form effects on white and black populations replicate predictions that one would expect from the model. These population effects similarly trace out corresponding changes in prices and neighborhood black share, which become the first stage estimates for the model.

I then estimate the choice parameters separately by broad occupation group using the linear instrumental variable approach developed by Berry (1994). These two-stage least squares (2SLS) estimates provide estimates of white and black willingness-to-pay for more or less black neighborhoods. Consistent with past findings of white aversion to more black neighborhoods, a typical white household would have to be compensated by a 1% lower house price for a 1 p.p. increase in the black share of the neighborhood to hold utility constant. At the same time, blacks seem to have no or weak preferences toward more black neighborhoods.

Nonetheless, for black families, these estimates reflect choices made among the restricted set of tracts where they were allowed to live. To quantify the effect of any non-price rationing on the allocation of black families to different neighborhoods, I have to specify the demand of black families in the absence of those restrictions, a task that is fundamentally unsuited to using only within-variation from the first-differenced regressions. To make progress, I assume a correlated random effects (CRE) structure that maintains the instruments' identification of the preferences over price and neighborhood black share. The CRE model imposes the assumption that the static component of neighborhood choice is a linear function of observable characteristics and an orthogonal unobserved component. Importantly, I allow this unobserved component to be correlated between races. This model allows me to predict counterfactual demand using the observed prices and white choice probabilities in all-white neighborhoods.

Finally, I compare the actual distribution of black and white demand by extending decomposition methods of Kullback and Liebler's (1951) relative entropy, a measure of statistical divergence between two distributions.[5] Specifically, I decompose segregation between blacks and whites by first comparing black families' actual neighborhood choices to the counterfactual choices that would arise if neighborhood constraints had been removed—quantifying the

[4] Importantly, I include as one of my controls the sum of the population shares of rural migrants in each census tract as of 1930. The addition of this control variable addresses a commonly-raised concern that shift-share instruments may be inadvertently picking up differences in the shares of earlier migrants that are correlated with unobserved determinants of subsequent migration.

[5] Mora and Ruiz-Castillo (2010, 2011) develop methods for decomposing KL divergences between nests of groups (e.g. city vs. school district segregation) and along different dimensions (e.g income vs. race).

contribution of supply-side explanations for segregation—and then comparing the counterfactual black demand to actual white demand—quantifying the contribution of demand-side explanations based on the responses to the prices and black resident shares observed in each neighborhood. This decomposition suggest that roughly half of segregation is explained by divergent preferences over the neighborhood's black share, and the remainder is driven by constrained supply.

In forthcoming work, I turn to aggregate evidence of housing market segmentation and its consequences for black residents. Abstracting from heterogeneity within cities, I compare the effects of an exogenous change in black population and white population on median black and white prices, respectively. Two facts emerge that are consistent to the within city analysis. First, an exogenous increase in black inmigration has a quantitatively large, statistically significant effect on average housing prices paid by blacks. In contrast, inflows of whites have no large or significant effect on housing prices for blacks. These results suggest that housing supply constraints caused materially higher prices for black households, as suggested by Cutler et al. (1999).

This paper contributes to several strands of literature. Most relevantly, this paper relates to empirical papers following Epple and Sieg (1999) that relate household equilibrium location choice to neighborhood quality and public good provision. In particular, Bayer et al. (2007) (henceforth, BFM) estimate models of equilibrium sorting similar to the one I present in this paper, identifying preferences over neighborhood characteristics using variation driven by households sorting across school district boundaries.[6] Using a similar approach, Bayer et al. (2004) report black and white preferences for segregation, but are careful to note that their estimates "[combine] the difference that results from decentralized preferences... as well as any centralized discrimination that causes black households to appear as if they prefer black versus white neighborhoods." This limitation is shared by virtually all recent papers that characterize equilibrium neighborhood choices. My paper attempts to separate the contributions of preferences and non-price constraints that limited the choices of black families to a subset of neighborhoods in the late 1930s.

Second, this paper relates directly to the expansive literature studying localized effects of migrants on labor markets, particularly those which utilizes the "past settlement" instrument.[7] The bulk of these papers utilize variation in migrant flows from different countries of origin, or in the case of internal U.S. migration, the subject of this paper, different states of origin (Boustan, 2010; Shertzer & Walsh, 2016). This paper shows that internal migrants are drawn to the locations of past migrants defined at the much finer county geography, which is suggestive of the importance of social networks in driving migration in addition to access to similar modes of transportation. A similar conclusion is reached in a recent study

[6]One interpretation of their procedure is that they correct for unobserved selection in a hedonic model of house prices by (1) isolating the sample to school boundary areas and including boundary-specific fixed effects and (2) using the mean utility from a multinomial logit demand system as a control function, using interactions between housing and household characteristics and characteristics of houses in other neighborhoods as excluded instruments.

[7]For an inventory of such papers, see Jaeger, Ruist, and Stuhler (2018).

by Stuart and Taylor (2017) who use data on town of birth to study white and black migration flows over the course of the 20th century. While most of the research in immigration has focused on labor market effects, two studies in particular, Saiz (2003, 2010), utilize the housing demand variation driven by large inflows of immigrants to trace out housing supply curves.

Third, I relate to a smaller literature that studies housing supply and its determinants.[8] This literature has primarily sought to better understand the connections between supply and construction, government policy, and housing durability, but less is known about whether the determinants of housing supply are connected to race.[9]

Finally, this paper connects to the tradition across the social and biological sciences that investigates the signals hidden in one's name. The focal points of interests have diverged across disciplines: social scientists have taken particular interest in how names, often first names, are connected to labor market success (see e.g. Bertrand & Mullainathan, 2004; Clark, 2014; Olivetti & Paserman, 2015; Goldstein & Stecklov, 2016), while biologists and physical anthropologists trace divergences in gene distributions from the hereditary nature of surnames (see e.g. Zei, Guglielmino, Siri, Moroni, & Cavalli-Sforza, 1983; Piazza, Rendine, Zei, Moroni, & Cavalli-Sforza, 1987; Zei et al., 1993). This paper utilizes the latter to explore how highly localized nature of social networks transmits correspondingly into highly localized housing demand pressure by neighborhood.[10]

This paper is organized as follows. Section 1.2 presents the conceptual framework. It starts with a model of neighborhood choice where agents have preferences over the price of housing and the racial composition of the neighborhood, argues that model parameters can be identified using migration shocks, and shows how the predicted demands from a correlated random effects model can be decomposed. Section 1.3 develops the instrument by tracing the origins of migrants in the 1930 census using surname distributions and shows the predictive power of these instruments. Section 1.4 defines concepts in the full count census data that are crucial for the analysis. Section 1.5 presents preference parameter results and the decomposition results. Finally, section 1.6 concludes.

1.2 Conceptual Framework

In the first part of this section, I lay out a simple model of neighborhood choice, which I estimate directly in later sections. In this model, families have preferences that depend on

[8] DiPasquale's (1999) appropriately titled review "Why Don't We Know More About Housing Supply?" bluntly begins, "[v]irtually every paper written on housing supply begins with some version of the same sentence: while there is an extensive literature on the demand for housing, far less has been written about housing supply."

[9] A notable exception, Bayer, Casey, Ferreira, and McMillan (2017) use rich longitudinal data and find housing price premia for minorities.

[10] Massey, Alarcón, Durand, and González (1987), Munshi (2003) explore the strong ties that migrants retain with origin communities within states in Mexico.

the price of housing in a given neighborhood, the neighborhood black share, and a race-specific unobservable taste component that has both a fixed (time-invariant) component and a transitory (period-specific) component. In the empirical implementation described below, I define neighborhoods by census tracts and study the determinants of choice across tracts within different cities using a model with metropolitan area by time dummies, fit separately for families with a head in different occupation groups. For the purposes of describing the main features of the model, however, I begin by focusing on choices of white and black families from a single occupation group within a single city.

Decomposing the channels of segregation into preferences- and constraints-based explanations requires generating a counterfactual distribution of demand in the absence of supply constraints. To do so, I need to address three key issues:

1. identification of the key preference parameters parameters that govern choices of white and black families over neighborhoods with different housing prices and different shares of black residents,

2. specification of the unobserved component of black preferences for neighborhoods in which there were (essentially) no black residents in 1940,

3. linking a model of neighborhood choice by white and black families to overall segregation.

I discuss general identification issues in subsection 1.2. Then in subsection 1.2, I discuss a random effects specification of preferences that can be used to infer black preferences across all neighborhoods in a given city. Finally, in subsection 1.2, I show how I can use my fully specified model of preferences of black and white families to conduct a simple decomposition of the Kullback-Leibler divergence measure of segregation. This decomposition separates observed segregation into two components: one that summarizes the effect of constraints facing black residents (i.e., the effect of non-price rationing constraints); and the other that represents the effect of different preferences of white and black families.

A Model of Neighborhood Choice

The model presented in this section is related to the one presented by Bayer and Timmins (2005), Brock and Durlauf (2002) but with a specific functional form for the social interactions. I study partial equilibrium adjustments in this model rather than solve for the entire equilibrium, similar to the theoretical exercise in Cutler et al. (1999).

I begin by defining a city as a set of J neighborhoods $\mathcal{J}^*$, and householders i of race $r(i) \in \{B, W\}$ who have preferences over neighborhoods $j \in \mathcal{J}^*$ in time period t given by a linear indirect utility function:

$$
\begin{aligned}
v_{ijt} &= \delta_{r(i),jt} + \varepsilon_{ijt} \\
\delta_{rjt} &= \beta_r \ln P_{jt} + \gamma_r s_{jt} + \boldsymbol{\Phi}_{\boldsymbol{r}}' \boldsymbol{X}_{\boldsymbol{jt}} + \alpha_{rj} + \xi_{rjt},
\end{aligned}
$$

Here, the δ_{rjt} term is a race-specific population mean utilities for neighborhood j, $\ln P_{jt}$ is the price, s_{jt} is the black share, and $\boldsymbol{X_{jt}}$ is a vector of observable characteristics of the neighborhood. The α_{rj} and ξ_{rjt} terms represent time-invariant and time-varying unobservable components of race-specific demands for neighborhood j, respectively. Finally, ε_{ijt} is an i.i.d. error drawn from an extreme-value type I distribution, representing a component of demand that is specific to family i.

During the period of my data (1930–1940), there is extensive documentary evidence that certain neighborhoods in most cities were off-limits to black residents via formal prohibitions (e.g., steering and restrictive covenants) and also via informal threats of violence. I formalize this fact by allowing the choice set available to black families to be strictly smaller than the choice set available to whites. Specifically, let $\mathcal{J}_r \subseteq \mathcal{J}^*$ represent the set of neighborhoods "available" to a particular race and $J_r = |\mathcal{J}_r|$ be the size of the choice set. Households choose from available neighborhoods in each period that maximize their utility

$$D_{it} = \arg \max_{j \in \mathcal{J}_{r(i)}} v_{ijt}. \tag{1.1}$$

The independence of irrelevant alternatives (IIA) property is generic to logit models and allows analysts to obtain preference parameters among the set of choices available to agents. However, I am explicit about the choice set primarily because of its mapping into supply constraints. This representation allows households to have well-defined utility over all neighborhoods but can only make choices on those directly available to them, a distinction that will become useful in the counterfactual decomposition exercises.[11]

Because I do not include any within group heterogeneity in utility, the extreme value assumption on ε_{ijt} gives the choice probabilities the convenient and well-known functional form of a multinomial logit

$$\pi_{rjt} = \frac{\exp \delta_{rjt}}{\sum_{k \in \mathcal{J}_r} \exp \delta_{rkt}} \tag{1.2}$$

for $j \in \mathcal{J}_r$ and 0 otherwise. Taking logs of equation (1.2) yields

$$\begin{aligned} \ln \pi_{rjt} &= -\theta_{rt} + \delta_{rjt} \\ &= -\theta_{rt} + \beta_r \ln P_{jt} + \gamma_r s_{jt} + \boldsymbol{\Phi_r' X_{jt}} + \alpha_{rj} + \xi_{rjt}, \end{aligned} \tag{1.3}$$

where $\theta_{rt} = \ln \sum_{k \in \mathcal{J}_r} \exp \delta_{rkt}$ is the "inclusive value," the population mean utility, of agents given the available choice set and the prices and neighborhood black shares in each neighborhood.

[11] An alternative, observationally equivalent formulation is that black households have "access" to all neighborhoods but have very strong negative preferences for those neighborhoods $\alpha_{rj} \to -\infty$. As termed by Cutler et al. (1999), the threat of "collective action racism" in these neighborhoods serves as a de facto restriction on access. They make verbal note of the distinction between de facto and de jure constraints but ultimately the two are observationally equivalent in their model as in mine. Relatedly, some de facto constraints may not be so intense as to completely exclude black families from some neighborhoods. I am ultimately unable to capture these constraints.

Letting N_{rt} represent the total number of families in a given race group in a city in period t, the total number of housing units demanded by families in that race group is Q_{rjt}, where

$$Q_{rjt} = N_{rt}\pi_{rjt} \tag{1.4}$$

Summing across black and white families the total demand for housing in neighborhood j is:

$$Q^*_{jt} = Q_{Bjt} + Q_{Wjt}.$$

Finally, in the analysis below, I posit a generic upward sloping inverse supply curve that relates the price of housing in a neighborhood to the number of units supplied

$$\ln P_{jt} = S_{jt}\left(Q^*_{jt}\right).$$

Empirical Implementation

Berry (1994) shows one can replace $\ln \pi_{rjt}$ with the log of observed choice probabilities $\ln \hat{\pi}_{rjt}$ and estimate

$$\ln \hat{\pi}_{rjt} = -\theta_{rt} + \beta_r \ln P_{jt} + \gamma_r s_{jt} + \boldsymbol{\Phi_r' X_{jt}} + u_{rjt} \tag{1.5}$$

using a simple specification that is linear in the log of the shares. The composite error term $u_{rjt} = \alpha_{rj} + \xi_{rjt} + (\ln \hat{\pi}_{rjt} - \ln \pi_{rjt})$ in this equation reflects the fixed and transitory components of race-specific neighborhood unobservables as well as sampling error in measuring the observed choice probabilities. Equation (1.5) is the main estimating equation for the paper.[12]

Careful consideration of the composite error u_{rjt} raises several important identification issues. A naive OLS estimation of equation (1.5) will lead to biased estimates due in part to omitted neighborhood amenities and disamenities α_{rj}, which are likely correlated with the local price. Inclusion of tract fixed effects in a panel setting with at least two observations per neighborhood will eliminate this bias. Even with fixed effects, however, the presence of the transitory taste shocks ξ_{rjt} will lead to a similar positive correlation between the observed neighborhood shares and prices. Morever, since $s_{jt} = \frac{Q_{Bjt}}{Q_{Bjt}+Q_{Wjt}}$, the fraction of black residents in a neighborhood is mechanically correlated with the preference factors of whites and blacks, leading to potential biases.

Assuming a generic set of instruments $\boldsymbol{Z}_{jt}$, one can write a linear first stage system

$$\ln P_{jt} = a_1 + \boldsymbol{b}_1'\boldsymbol{Z}_{jt} + \boldsymbol{c}_1'\boldsymbol{X}_{jt} + e_{1jt} \tag{1.6}$$

$$s_{jt} = a_2 + \boldsymbol{b}_2'\boldsymbol{Z}_{jt} + \boldsymbol{c}_2'\boldsymbol{X}_{jt} + e_{2jt}. \tag{1.7}$$

[12]The existence of s_{jt} in equation (1.5) means that the inversion of the mean utilities into log market shares is not a closed form system of the system of simultaneous equations and thus is not a housing demand equation, per se. However, as pointed out by Bayer et al. (2004), *agents* can take the neighborhood share as exogenous since they have miniscule influence over the racial share of the neighborhood. Still, the *analyst* still must treat s_{jt} as endogenous in the presence of unobserved, aggregate shocks to the neighborhood.

Instruments for prices have seen much attention in empirical industrial organization, but these approaches are infeasible in my setting. The first commonly used instrument uses functions of other product characteristics available in the same market (e.g. Berry, Levinsohn, & Pakes, 1995). However, as Nevo (2001) points out, the inclusion of product (in my case, neighborhood) fixed effects will absorb most of this variation.

He instead uses a second common approach which takes variation in the prices of products in other markets as correlated, but exogenous. When estimating housing demand, BFM take a similar approach utilizing variation in housing quality that comes from faraway housing outside of school district boundary areas. There are two problems with applying this second approach in my setting. First, neighborhoods do not have clear analogs in other cities. Second, the goal of my exercise is to characterize housing demand and segregation in the entire city. Without being able to limit the geographic focus of neighborhoods within a city, equation (1.6) is subject to "endogenous effects" critiques of using leave-out-means from the peer effects literature (Manski, 1993).

Instead, in this paper, I use instrumental variables based on the predicted inflows of whites and blacks who have a particular interest in a given neighborhood driven by their social connections to the pre-existing residents of the neighborhood.[13] The availability of two instruments—one reflecting predicted inflows of whites, the other reflecting inflows of blacks—resolves the need to address the endogeneity of both neighborhood housing prices and the neighborhood share of black residents.

Intuitively, migrant shocks are relevant instruments because they increase the demand of the neighborhood, which should subsequently increase the price. Race-specific demand shocks affect the racial composition of the neighborhood. However, these initial shocks are met with feedback responses as the full system returns to equilibrium.

I show in a forthcoming appendix that migrants' effects on neighborhoods are heterogeneous with respect to the pre-existing black share of the population. In particular, black migrant demand shocks in relatively white neighborhoods lead to *decreases* in population and subsequent decreases in prices if whites have preferences for segregation $\gamma_W < 0$. This arises because the second-order feedback effect of white families leaving in response to a black migrant demand shock trump the first-order increase in population.

By contrast, in relatively black neighborhoods, there are few white families that leave in response to black demand shocks so prices increase. The model's predicted effects of white migrant shocks are symmetric. To capture this heterogeneity, I interact the instruments with the 1930 black share.

Correlated Random Effects and Counterfactual Demand

Subsection 1.2 lays out the requirements to credibly identify β_r and γ_r for both black and white families. One can use fixed effects to absorb static unobserved quality differences

[13] In the context of the model, one can specify ε_{ijt} with two components according to Cardell (1997). One that reflects a draw specific to individuals of race r with the same origin o and a separate component that is individual specific. The shared first component will draw new migrants to particular neighborhoods.

across neighborhoods that reflect in prices and quantities and migration instruments to provide variation independent of time-varying unobservables. If it is a reasonable assumption, the independence of irrelevant alternatives (IIA) assumption of logit models—preference parameters can be determined by how agents interact given the choices they have—means that the existence of a constrained choice set for black residents $J_B \leq J$ poses no direct threat or adjustment to estimating model parameters.[14]

However, in order to quantify supply constraints, I need to be able to compare actual black demand to demand in the absense of those restrictions. An approach that only uses the within-variation of fixed effects models gives little guidance for what black utility might be in neighborhoods where there were no black residents because of supply constraints. In other words, estimating β_r and γ_r does not require parametric assumptions about α_{rj}, but quantifying segregation does.

In this section, I impose a correlated random effects structure to parameterize α_{rj}, which maintains similar identifying assumptions for β_r and γ_r. As I will detail further, the model I specify importantly allows the neighborhood unobservables for black and white residents to be correlated without requiring them to be the same. This structure allows me to use information from observable characteristics and the residuals from the models fit to white shares to extrapolate α_{Bj} and ultimately predict the black choice shares $\hat{\pi}_B^{CF}$ in neighborhoods with no black residents.

α_{rj} as a Correlated Random Effect

To draw the closest parallel to the fixed effects model I present in section 1.5, I impose a linear structure on α_{rj}:

$$\alpha_{rj} = \boldsymbol{A}_r' \bar{\boldsymbol{Z}}_j + \boldsymbol{C}_r' \bar{\boldsymbol{X}}_j + \psi_{rj},$$

where $\boldsymbol{A}_r$ and $\boldsymbol{C}_r$ allow the race-specific static characteristics of the neighborhood α_{rj} to correlate time-varying factors in the model.

I make the standard random effects assumptions:

1. $\boldsymbol{E}\left[\psi_{rj} | \bar{\boldsymbol{Z}}_j, \bar{\boldsymbol{X}}_j\right] = 0$, the static neighborhood unobservabled component of mean utility is conditionally mean zero,

2. $\boldsymbol{E}\left[\xi_{rjt} | \boldsymbol{Z}_{jt}, \boldsymbol{X}_{jt}, \alpha_{rj}\right] = 0$, the time-varying unobservable is conditionally mean zero,

3. ξ_{rjt} is serially uncorrelated, and

4. $\boldsymbol{E}\left[\xi_{rjt}^2\right] = \varsigma_r^2$, the neighborhood unobservables are homoskedastic.

I further assume cross-equation correlation between the black and white unobservables ψ_{Bj} and ψ_{Wj}:

[14] In this formulation, I continue to be agnostic between de jure (e.g. legal) and de facto (e.g. violenct) constraints.

$$\begin{pmatrix} \psi_{Bj} \\ \psi_{Wj} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_B^2 & \sigma_{BW} \\ & \sigma_B^2 \end{pmatrix}. \tag{1.8}$$

Note that Assumption 2 is the same as an exclusion restriction for validity of a 2SLS regression with fixed effects.[15] The additional assumptions 1, 3, and 4 are used to recover A_r, C_r and the covariance structure specified in equation (1.8). Assuming no serial correlation in ξ_{rjt} means that all serial correlation in the composite error u_{rjt} must derive from the presence of a time-invariant unobservable α_{rj}.

Allowing a correlation between ψ_{Bj} and ψ_{Wj} allows me to predict counterfactual unobservables for black residents utilizing residual variation in the choice probabilities of white residents in those neighborhoods. Here, the homoskedasticity and normality assumptions are not crucial but produce a simple regression formula for using the white random effect to predict the black unobserved component, i.e.,

$$\boldsymbol{E}\left[\psi_{Bj}|\psi_{Wj}\right] = \frac{\sigma_{WB}}{\sigma_B^2}\psi_{Wj}. \tag{1.9}$$

It is worth considering what α_{rj} represents. Amenities valued equally among all households should be capitalized into price: in a spatial equilibrium model with homogeneous valuations of neighborhoods, this term should be zero.[16] Recalling that I perform my analysis separately by broad occupation, ψ_{rj} are race-specific deviations from of how individuals of a particular occupation group value neighborhoods. If all heterogeneity in neighborhood taste is driven by occupation, then $A_B = A_W$, $C_B = C_W$, and $\frac{\sigma_{WB}}{\sigma_B^2} = 1$, reflecting characteristics not included in the model such as average commute times. However, this formulation does not impose this restriction.

Extrapolation

I predict the log mean utilities $\widehat{\ln \delta_{rj}}$ for all neighborhoods $j \in \mathcal{J}^*$ by combining:

1. the instrument-identified parameters $\hat{\beta}_B$, $\hat{\gamma}_B$, and observed endogenous characteristics $\ln P_{j,1940}$, $s_{j,1940}$;
2. $\hat{A}_B$, $\hat{C}_B$, and observed exogenous characteristics $\bar{\boldsymbol{X}}_j$ and $\bar{\boldsymbol{Z}}_j$; and
3. the estimated random effect $\hat{\psi}_{Bj}$,
4. setting the time-varying unobserveable $\hat{\xi}_{Bjt} = 0$.

[15] Mundlak (1978) shows that correlated random effects models specified similar to the one I assume give numerically equivalent coefficient estimates for time-varying covariates that are not absorbed by the fixed effect. This numerical equivalence does not hold in my over-identified instrumental variable setting, the results are ultimately similar.

[16] In fact, the household-specific error ε_{ijt} should also be zero. See Kline (2010).

Finally, I renormalize by subtracting $\ln \sum_j \left(\exp \widehat{\ln \delta_{rj}}\right)$ from each predicted log choice probabilities, which becomes the estimate of $\hat{\theta}_{rt}$. This renormalization maintains the linear in log-odds property of the multinomial logit model and also has the probabilities sum to 1.

This approach is not without limitations, that I will discuss in subsection 1.2. However, one of the primary advantages of specifying a random effects variant of the Berry multinomial logit specification is that the choice model can be directly mapped into a model of segregation with terms that can be directly interpreted as contributions from supply side and demand side factors. I turn to this decomposition now.

A Decomposition Framework for Segregation

The estimates from the correlated random effects model allow one to ask quantify latent black demand for neighborhoods at prevailing prices in neighborhoods with no black residents. Intuitively, differences between actual and counterfactual distributions of black demand are driven purely by the constraints. Correspondingly, differences between the latter and actual white demand are driven by preferences.

In this section, I lay out a framework that quantifies this intuition by decomposing the Kullback-Leibler divergence between two distributions. The divergence between black and white families' choices in a given city and time period can be written as

$$KL\left(\pi_B || \pi_W\right) = \sum_j \pi_{Bj} \ln \frac{\pi_{Bj}}{\pi_{Wj}},$$

where as before, π_{rj} is the probability mass function of the race r's multinomial choices. Just like the indices of dissimilarity and isolation, this measure quantifies how two distributions are different; however, its functional form lends itself very naturally to the multinomial logit model.

To incorporate the counterfactual black demand in decomposing actual segregation, I multiply and divide the term within the logarithm by $\hat{\pi}^{CF}$ so

$$\begin{aligned} KL\left(\pi_B || \pi_W\right) &= \sum_j \pi_{Bj} \ln \frac{\pi_{Bj}}{\hat{\pi}^{CF}_{Bj}} \frac{\hat{\pi}^{CF}_{Bj}}{\pi_{Wj}} \\ &= \sum_j \pi_{Bj} \left[\ln\left(\pi_{Bj}/\hat{\pi}^{CF}_{Bj}\right) + \ln\left(\hat{\pi}^{CF}_{Bj}/\pi_{Wj}\right)\right]. \end{aligned} \tag{1.10}$$

One can see immediately that the two terms that emerge reflect exactly the intuition from before: the first term reflects differences driven by an expansion of the choice set and the second reflects differences in preferences for different neighborhoods. Adding and subtracting $\log \hat{\pi}^{CF}$ is similar to the derivation of an Oaxaca decomposition where one adds and subtracts counterfactual predicted in a linear model. Because of the linear-in-log-share specification, a detailed decomposition of components attributable to different covariates immediately follows.

Replacing the actual probabilities with model predicted probabilities in this decomposition gives

$$KL\left(\hat{\pi}_B || \hat{\pi}_W\right) = \sum_j \hat{\pi}_{Bj}\left[\ln\left(\hat{\pi}_{Bj}/\hat{\pi}_{Bj}^{CF}\right) + \ln\left(\hat{\pi}_{Bj}^{CF}/\hat{\pi}_{Wj}\right)\right].$$

Using the predicted probabilities means that one can use the parameter estimates from the structural model in equation (1.5), which yields:

$$\begin{aligned} KL\left(\hat{\pi}_B || \hat{\pi}_W\right) = \sum_j \hat{\pi}_{Bj} \Big\{ & \left(\theta_B^{CF} - \theta_B\right) \\ & + \left[\left(\theta_W - \theta_B^{CF}\right) + \left(\alpha_{Bj} - \alpha_{Wj}\right)\right. \\ & \left. + \left(\beta_{Bj} - \beta_{Bj}\right)\ln P_j + \left(\gamma_{Bj} - \gamma_{Wj}\right) s_j + \left(\boldsymbol{\Phi'}_B - \boldsymbol{\Phi'}_W\right)\boldsymbol{X}_j\right]\Big\}. \end{aligned} \tag{1.11}$$

The supply constraints contribution to segregation on the first line is summarized completely by the difference in the inclusive values $\theta_B^{CF} - \theta_B$ from expansion of the choice set—by construction, these two predicted distributions do not differ in their preference parameters. The second line is a detailed decomposition of the contribution of preferences to segregation. In particular, the $\left(\gamma_{Bj} - \gamma_{Wj}\right) s_j$ term, how segregation is driven by differences in preferences for more or less black neighborhoods, has been the focal point of the past literature.

Limitations of the Model

There are several important limitations to my procedure in capturing supply and demand contributions to segregation. First, I do not simulate a new sorting equilibrium for black and white families. In my approach, the simple model-based counterfactuals constructed from the random effects estimates hold constant $\ln P$ and s. It does not take into account second-order equilibrium adjustments (e.g. neighborhood tipping) that are likely important. But, capturing the first-order effects is the most straightforward way to clearly attribute contributions from supply and demand. Moreover, a full simulation of the new sorting equilibrium in absence of supply constraints will inevitably require knowledge of neighborhood supply curves and even stronger assumptions and parameterizations. While important, I leave these exercises for future research.

Second, my procedure relies on the structure of the model—particularly IIA—to accurately predict the latent black demand for otherwise unavailable neighborhoods $j \in \mathcal{J}^* \backslash \mathcal{J}_B$. Here, the threat to the validity of this speculative extrapolation exercise is that the model is incorrectly specified in those neighborhoods. External validity is always a lurking issue for models estimated via instrumental variables, but in this exercise, interpreting the extrapolated black neighborhood unobservable α_{Bj} merits additional discussion.

When supply constraints come from de jure restrictions, then the counterfactuals are clear: lifting supply constraints means legislating or effecting policy changes that abolish these restrictions. The modeling environment I provide in this paper treats the abolition of de facto restrictions coming from extralegal harassment and terrorism as similarly straightforward in a counterfactual world.

But, if organized violence is simply an extreme form of racism, these counterfactuals are essentially asking what would segregation look like if racism were "bounded above." A more complicated correlated random coefficients model could connect γ_W with the black neighborhood unobservable α_{Bj}, but ultimately counterfactuals generated from such a model will still rely on IIA. I leave such explorations for future work.

1.3 An Instrument for Inmigration

I now turn my attention to constructing an instrument for migration. The 1930's was a nadir for the first wave of the Great Black Migration, marking the tail end of nearly a quarter century of migration ending at the onset of the second World War. Nevertheless, there was substantial black and white migration during the 1930s. Table 1.1 reports the magnitude of gross migrant flows to the 48 major metropolitan areas with census tracts that form my main analysis sample.

The top panel shows that roughly half of black migrant flows were movements between cities, while the other half represented flows from rural counties, defined as those without any census incorporated places between 1910–1940. Most of these black rural migrants came from counties in the southern census region. By comparison, the bottom panel shows white migrant flows in greater absolute magnitude albeit with a smaller fraction represented by rural-to-urban migration.

Nonetheless, much of the migration to major cities for both blacks and whites was intraregional. Prior literature has focused on southern blacks leaving the south, but table 1.1 shows that roughly 60% of the rural-to-urban migration of southern blacks was within the south itself. White migration was also heavily within-region, the notable exception of the westward migration of whites from drought affected "Dust Bowl" regions in Texas, Oklahoma, and Kansas, notwithstanding.

In section 1.2, I motivated using migration shocks as a source of exogenous variation to perturb the sorting equilibrium. Figure 1.1 illustrates some of this variation, plotting black (purple from the left) and white (green from the right) rural county-to-census tract migrant flows from Texas and Oklahoma to Los Angeles County between 1935–1940. Origin counties are shaded in intensity based on the total rural-to-urban migrants to any destination.

The black flows to Los Angeles focus primarily on tracts in Watts and Compton with a high share of black residents in 1930 (red). But, there is dispersion: migrant flows are not perfectly correlated with the 1930 black share with some relatively high black share tracts getting fewer migrants than expected and vice versa.

The flow diagram also suggests that particular origin counties may have ties with particular tracts. Migrants from rural counties outside of Austin seem to be disproportionately directed toward tracts near Glendale and Pasadena. Meanwhile, migrants from rural counties outside Oklahoma City, seem to be particularly drawn toward Carson City and south Compton. I will shortly provide regression evidence that shows that counties do have particular ties to particular tracts.

Correspondingly, Figure 1.1 plots large directed flows of white migrants from rural Oklahoma. These migrant flows also show both dispersion and directedness. Notably, some counties in central Texas have large numbers of migrants, but relatively few are choosing to move to Los Angeles altogether.

The patterns illustrated in Table 1.1 and Figure 1.1 make a case for using granular county-level variation. But, directly applying the 1935–1940 county-level migrant flows measured from the retrospective component of the 1940 census is problematic because it is potentially subject to the same endogeneity that threatens panel OLS estimates of the preference parameters. This motivates using the shift-share instrumental variables approach common in the international migration literature—a neighborhood analog to the occupation partitions in the within-city analysis of Card (2001). This instrument uses past migrant location decisions to proxy for the predictable component of current migrants' decisions reflecting access to similar modes of transportation as well as social networks. However, origin counties of migrants prior to 1930 are conspicously absent in the data.

This section outlines how I utilize surname distributions to build a shift-share instrument. First, I describe the basic intuition of the instrument, which is similar in many respects to those used in the international migration literature. In the second subsection, I outline how I use surname distributions constructed from the 1910–1930 censuses to construct estimates of past migrants' choices. Finally, in the third subsection, I show that surname-constructed choice probabilities from the 1930 census are highly correlated with migration choice probabilities at the census tract level. I also show that these instruments provide independent variation and are not simply proxying for growing cities.

Requirements of an Instrument Using Past Settlement

In section 1.2, I dropped notational dependence on cities to focus purely on neighborhoods. In this section, I reintroduce necessary notational dependence on destination cities d to discuss migration.

Let $q_{j|rod}$ be the probability that a migrant of race r from origin o chooses neighborhood j conditional on being in city d, and let $p_{d|ro}$ be the probability that the migrant chooses city d. Let $inflow_{rdj}$ represent the total inflow of migrants of race group r who move to neighborhood j within city d and let $outflow_{ro}$ be the outflow of migrants of race r from origin o. Observed migrant flows can be written as:

$$inflows_{rdj} = \sum_{o} q_{j|rod} \times p_{d|ro} \times outflow_{ro}. \tag{1.12}$$

Two main sources of endogeneity can potentially arise from this accounting formula. First, even comparing the same cities and neighborhoods over time, migrants might be more likely to move to cities or neighborhoods that have housing supply shocks (e.g., a faster rate of conversion of older single family homes into multi-family apartments). In the context of equation (1.12), this would mean that $q_{j|rod}$ or $p_{d|ro}$ endogenously adapt to housing supply shocks.

One solution to this endogenous adaptation is to replace the q and p in equation (1.12) with probabilities based on the obseved settlement patterns of earlier migrants from thes ame origin and race group, $q^{past}_{j|rod}$ and $p^{past}_{d|ro}$, respectively. Specifically, to predict migrant flows from 1935–1940, I use the settlement patterns for migrants from each origin county who were observed living in larger cities in 1930—so "past" refers to migrants who had settled by 1930 in a particular urban area. As discussed in the next section, the 1930 census does not contain county of birth or residence at an earlier date, so I have to construct $q^{1930}_{j|rod}$ and $p^{1930}_{d|ro}$ based on clustered patterns of surnames.

A second concern is that potential migrants from a given origin may be tightly connected to a specific destination city. The simplest example are migrant flows to cities from rural counties on the outskirts of a city boundary. In this case, $outflow_{ro}$ may be partially endogenous to shocks in city d. To address this concern, one could use the outflows from origin o leaving out migrants who end up moving to city d, $outflow^{-d}_{ro}$. This flow measure is purged of endogeneity arising from demand-pull factors in city d. Combining these substitutions, I construct a predicted inflow of migrants of race group r from all origin counties to neighborhood j in city d between 1935–1940:

$$Z_{rjd} = \sum_o q_{j|rod} p_{d|ro} outflow^{-d}_{ro}. \tag{1.13}$$

where $outflow^{-d}_{ro}$ is based on the questions in the 1940 census that asks each person where they lived 5 years ago.

Past Migrant Flows Using Surname Distributions

An Overview of the Approach

In this section, I describe a series of steps I use to construct estimates of $q^{1930}_{j|rod}$ and $p^{1930}_{d|ro}$ based on counts of migrants who were observed in 1930 living in specific census tracts of larger cities in 1930. The 1910–1930 censuses have information on state of birth and current place of residence, as well as information on fullnames of all respondents. In brief, my procedure uses the fact that surnames were highly clustered in the early twentieth century. Thus, if one knew that a given black person was born in Texas and had a given surname, one could make an informed guess about their likely *county* of birth.

Throughout this section, I also temporarily drop the race group r subscript and the 1930 superscript. All counts and probabilities should be understood as referring to a specific race group observed in the 1930 census.

Let L_{od} represent the number of residents in city d who were born in origin county o and let M_{odj} represent the number of residents in city d and neighborhood j who were born in origin county o. If birth county data were available, one could easily write the city and neighborhood choice probabilities as:

$$p_{d|o} = \frac{L_{od}}{\sum_{d'} L_{od'}} \tag{1.14}$$

$$q_{j|od} = \frac{M_{odj}}{\sum_{j' \in \mathcal{J}_d} M_{odj'}}. \tag{1.15}$$

The key challenge is that none of these objects exist in the data.

Measuring Past Flows

I will show that surname distributions fill this gap. Throughout this section, I focus attention to $p_{d|o}$ because the procedure to construct $q_{j|od}$ is exactly analagous.

To summarize my procedure, I first assign a set of weights to each resident living in cities in 1930. The weights proxy the probability that an individual comes from a particular county. I generate weights by computing the fraction of non-migrant individuals living in the residents' state of birth who share the same last name and 10-year birth cohort, pooling data from the 1910–1930 censuses. Summing the origin-specific probabilities in a city gives an estimate of the origin-specific population.

Formally, let c index cells that identify unique combinations of an individual's last name, state of birth, and 10 year birth cohort, all information readily available in the 1930 census. Let $N_d(c)$ represent the total number of individuals in a cell c. Then a probabilistic measure of the population from origin county o becomes

$$\hat{L}_{od} = \sum_c N_d(c) \times \boldsymbol{Pr}\left(o|\underbrace{\text{surname, cohort, birth state}}_{c}\right), \tag{1.16}$$

which is a count of individuals weighted by the probability that they came from a particular county o.[17]

The next subsection 1.3 gives some background on common surnames and why they might provide information on county of birth. Next, subsection 1.3 describes how I use the population of non-migrants in each state to construct $\boldsymbol{Pr}(o|c)$. Finally, I analyze whether these variables succeed in providing granular information that can be used to construct county-to-county flows.

Signal in Surnames

Researchers in biology and physical anthropology (e.g., Zei et al., 1983; Piazza et al., 1987; Zei et al., 1993) have treated surnames as alleles traditionally transmitted via male lineage to measure patterns of genetic drift—i.e., migration. Several facts about black and white

[17] One can derive the same estimate for $p_{d|ro}$ by using Bayes's rule with an independence assumption $\boldsymbol{Pr}(o, d|c) = \boldsymbol{Pr}(o|c)\,\boldsymbol{Pr}(d|c)$ and assuming that the geographic distribution of non-movers reflects the birth locations of movers.

surnames in the early 20th century are suggestive that these same patterns hold in the mid-20th century United States. The observation that immigrants move to enclaves of past migrants is not a feature unique to recent waves of immigration to the United States, and as a result, native born whites descendents of European migrants from the late 19th and early 20th centuries should be clustered and not totally dispersed (Tabellini, 2018).

For blacks, last names were often imposed by slave masters in the antebellum era, ultimately making it unlikely that surnames carry any signal that can connect former slaves and their descendants to their ethnic origin countries in West Africa. Nonetheless, Cook, Logan, and Parman (2014) find not only evidence of distinctive black first names in the beginning of the 20th century but also find that African Americans are more likely to have the last names of famous figures (e.g. George Washington). African Americans also took surnames reflecting emancipation (e.g. Freeman) or their occupation (e.g. Smith). The empirical question remains of whether common black surnames are clustered and provide signal of county of origin.

Constructing Surname Distributions

To construct the distributions $\boldsymbol{Pr}\left(o|c\right)$ used in equation (1.16), I pool data from the 1910–1940 censuses separately by state. First, I limit the sample to non-migrants by only keeping individuals who are living in their state of birth. Second, I only keep individuals born in the prior 10 years in the 1920–1940 censuses. This way, the distributions for a particular birth cohort only come from a single census.

Most individuals have a common last name, and so this analysis relies on common last names providing a signal of birth location. I do not rely on using uncommon names to find unique match on individual characteristics. I define names as being common if at least one person in each decade between 1900–1940 has the last name. Individuals with uncommon last names are pooled into a single category according to this definition.[18] At this point, I discard individuals born between 1930–1940.

Finally, I define cells analogously to equation (1.16) and construct the fraction of individuals in each state, surname, birth cohort cell living in each county. These fractions are my estimate for $\boldsymbol{Pr}\left(o|c\right)$.

I provide two pieces of evidence that last names provide a strong signal of county of birth, focusing on shares constructed using cells without incorporating birth cohort to analyze the predictive power of last names by themselves. First, figure 1.2 plots the resident shares of three common last names for whites and blacks in Texas: Adams, Carter, and Jones. Whereas black Adamses are more represented in Navarro County, black Carters and Joneses

[18] One question that arises is why limit oneself to surnames and utilize first names, finer age categories, and respondent's gender to in principle generate a better version of $\hat{L}_{od}$. This only makes sense if, for example, a first name contains signal of one's origin of birth, having already conditioned on surname and birth cohort. However, in practice, doing so generates small cells and thus either requires assigning more individuals to the "uncommon" binned category or placing parametric structure on $\boldsymbol{Pr}\left(o|c\right)$.

are overrepresented in Freestone and Walker Counties. The same corresponding surnames are not clustered in exactly the same fashion among whites, but are clustered nonetheless.

Second, I aggregate these case studies to quantify the distinctiveness of the each last name. I systematically compare the county shares $\boldsymbol{\varphi}_o$ of surname ℓ in each state to the county shares leaving out the surname $\tilde{\boldsymbol{\varphi}}_\ell$. I form a Pearson χ^2 test statistic for each last name:

$$\chi^2_\ell = N_\ell \sum_o \frac{(\varphi_{\ell o} - \tilde{\varphi}_{\ell o})^2}{\varphi_{\ell o}}$$

where N_ℓ is the number of individuals with last name ℓ. This test statistic is distributed according to a χ^2 distribution with degrees of freedom equal to the number of counties in the state minus 1, and the null hypothesis of this test is that individuals who have surname ℓ have the same geographic distribution of those who do not. In 1930, 99.4% and 99.6% of black and white individuals have a surname with a p-value that the computer cannot distinguish from zero.[19]

Flow Probabilities

I now substitute the constructed $\boldsymbol{Pr}(o|c)$ into equation (1.16) to generate $\hat{\boldsymbol{L}}_{od}$. I use these destination populations to construct flow probabilities according to equation (1.14), and having already assigned individuals to census tracts in 1930, I do the same for equation (1.15). Next, I provide visual and regression evidence that the surnames have strong predictive power using the actual flows between 1935–1940 constructed from the retrospective question in the census. Figure 1.3 plots the log of the 1935–1940 county-to-tract flow probabilities against the corresponding measure from the 1930 census for both blacks and whites, restricting the sample to the 25% of rural counties with the largest outflows. A clear positive slope emerges.

Table 1.2 quantifies these relationships in regressions of the 1935–1940 flow probability on the surname constructed probability, weighting each observation by the total rural-urban outflows from the origin county. The first column shows that the surname constructed shares are highly predictive of actual migrant choice probabilities.

The auxiliary analysis in column (1) is sufficient to generate an instrument for migration, but in columns (2)–(5), I estimate the same models with increasing number of fixed effects. In column (2), I include destination tract fixed effects. Heterogeneity across the black share of neighborhoods provides an important source of identifying variation in this paper, and the inclusion of tract fixed effects suggests that black migrants are not driven by neighborhood characteristics such as share of black households.

In columns (3)–(4), I include state of origin by destination metro area fixed effects. Using only this within variation, the persistent predictiveness of the surname constructed shares shows that migrants from the same county are not simply chosing the same cities as past migrants, the focal point of the cross-city comparisons in previous work. Migrants choose the same neighborhoods.

[19] I.e., 2.2×10^{-308} for double floating precision on the machines where I perform the analysis.

Finally, in column (5), I include state of origin by destination census tract fixed effects. These highly saturated regressions have two implications. First, neighborhoods have strong connections to origin places *within region*, which are not likely to be purely driven by similar access to transportation networks. Second, the distributions of surnames allows me to exploit migrant variation generated by outflows from rural counties to cities within the same region, and indeed the same state. The surname-constructed flow probabilities' strength is strong evidence of familial and kinship relationships that underly the motivation behind the past-settlement instrument.

1.4 Data and Definitions

This paper uses the full count 1910–1940 decennial censuses digitized by IPUMS and Ancestry.com (Ruggles et al., 2018). In this section, I address definitional questions that arise when implementing the strategy outlined in section 1.2.

Families and Neighborhoods

The first immediate issues is how one defines a neighborhood, which I define as census tracts according to their definition in the 1940 census. Because census tracts were first developed in 1934, they are not readily available in the 1930 census data. Therefore, I assign households in 1930 to 1940 census tracts using a procedure based on street addresses documented in the appendix.

The second immediate issue that arises is how to define the decision making unit in the model. I analyze the collective decisions of family units, defined as households with a male head between the ages of 18–50 with a cohabiting wife and at least one child. Furthermore, I exclude analysis of families living in group housing. Overall, families were less likely to live in group housing (e.g., hotels, the YMCA, institutionalized settings, etc.), where data on housing costs is generally not available.

Even by limiting the analysis to families, there may still be taste heterogeneity among the agents. Papers in empirical industrial organization have found that in models of differentiated products, allowing for taste heterogeneity can have a dramatic effect on the conclusions one draws (see e.g. Berry et al., 1995; Petrin, 2002).[20] To overcome these concerns yet retain parsimony of the choice models, I group families based on broad occupation groupings for the household head. Finally, when constructing the observed choice probabilities, I exclude consideration of tracts where there are fewer than 10 families in the same race and occupation grouping.

[20] Because they only observe a single market share for each product, these papers often have to apply sophisticated method of methods estimators to include taste heterogeneity in their models and overcome ecological fallacies. I do not face this same challenge because I am able to construct shares using the microdata.

Table 1.3 reports details of the 1940 subset of households that I use to construct the shares. The top panel shows that of the 9 million black and white households living in one of the 48 major cities with census tracts in 1930, 3.5 million households are families by my definition. Roughly 6% of families in cities are black families. Black household heads are clustered in three broad occupation groups that constitute 86% of families—laborers, service workers, and operators. Men in these relatively low-skilled occupations include longshoremen, cooks, janitors, cooks, deliverymen, and valets.

By contrast, less than 40% of white household heads are employed in one of these occupations, and a majority of them are operators, a group that also includes apprentices in blue collar professions. A comparable 23% are blue collar craftsmen, and the remaining 40% of white family men are white collar professionals.

Having generated coarse groupings of families, I summarize the characteristics of neighborhoods of the median and white black resident in table 1.4. The top panel reports characteristics of housing, particularly the cost and ownership status of units. A typical white resident lives in a tract with a median price of roughly $3,500 (roughly $63,000 in 2018 dollars), while the neighborhood of the median black resident is $800 less. This masks tremendous differences in the home ownership rate between neighborhoods. The typical white person lives in a neighborhood with double the homeownership rate of the typical black person. Using a composite price index that combines contract rent and self-reported house prices into a single measure (described in detail in the next subsection 1.4), the less than 30% price gap balloons to more than 40%. Focusing on employed white people in black occupations, that gap narrows (~25%), but is still sizeable.

The bottom panel reports characteristics of the people living in these neighborhoods. A stark contrast emerges in terms of the black share of neighborhoods of the median white and black family: roughly 50% of white people living in major cities live in neighborhoods with essentially no black people, but the median black person lives in a predominantly black neighborhood, a fact that does not change when again limiting to whites in low-skill occupations. These differences overshadow smaller differences in neighborhood employment shares.

Defining the Price of the Neighborhood

The 1930 census was the first where measures on housing costs were solicited.[21] Because census tracts have differing shares of renter and owner households, simply using the median rent or the median house price within a neighborhood likely does not consistently capture the cost of the neighborhood. In an extreme case, there are some neighborhoods with no

[21] Non-farm households, "[families] or any other group[s] of persons, whether or not related by blood or marriage, living together with common housekeeping arrangements in the same living quarters" (Ruggles et al.) were surveyed on their monthly contract rent or the estimated value of their home for renter-occupied housing and owner-occupied housing, respectively. For owner occupied housing, the value of the home is self-reported and represents an estimate unless the house was recently purchased. For renter occupied housing that was provided as in-kind compensation of labor, enumerators estimated the rent paid for similar housing.

owner-occupied or renter-occupied units. To avoid issues of composition, I create a create a single cost index that utilizes information on both the neighborhood rents and house prices by converting rents into home price equivalent units.

More formally, I stipulate that each house is defined by a latent cost, which translates proportionally into either a house price or a contract rent. This in turn implies a proportional relationship between a house's inherent price and contract rent, $HomeValue_{it} = \rho MonthlyRent_{it}$.[22]

To estimate ρ, I compare units at the same address that convert from rental to ownership status and vice versa between 1930 and 1940. Operationally, I construct a longitudinal dataset of housing units where the 1930 address could be matched on housing number and street name to a 1940 address documented in a forthcoming appendix. I then construct a single dependent variable that stacks owner-occupied home values and renter-occupied rents into $Y_{it} = (1 - Owner_{it}) \times \log Monthly\_Rent_{it} + Owner_{it} \times \log HomeValue_{it}$ and estimate a regression model:

$$Y_{it} = c_{a(i)} + d_{j(i),t} + (\log \rho)\, Owner_{it} + \varepsilon_{it} \tag{1.17}$$

where $c_{a(i)}$ are address fixed effects and $d_{j(i),t}$ are census tract by year fixed effects, which absorb both changes in the overall price level and tract-level variation in housing cost growth. Thus, $\widehat{\log \rho}$ captures the average change in housing costs from addresses that experience conversions of housing from owner-occupied to renter-occupied. Estimates in the appendix suggest that $\log \rho = 4.8$, which means annual contract rent payments are roughly one tenth of the self-reported value of the home.

1.5 Empirical Implementation and Results

Econometric Issues

Utilizing migrant flows as an exogenous source of variation to identify the model parameters carries with it several econometric issues. First, the static model in section 1.2 suggests a panel IV regression identified using time-varying instruments. However, the instruments laid out in section 1.3 are static. These shift-share instruments in the international migration literature shift changes rather than levels of the endogenous variable.

To obtain parameter estimates of β and γ, I utilize the equivalence of fixed effects and first-differenced regressions with two time periods and adapt a first differenced version of equation (1.5)

$$\Delta \ln \hat{\pi}_{rj} = -\Delta\theta_r + \beta_r \Delta \ln P_j + \gamma_r \Delta s_j + \tilde{\boldsymbol{\Phi}}_{\boldsymbol{r}}' \boldsymbol{X_j} + \Delta u_{rj} \tag{1.18}$$

and instrument for the changes in the endogenous variables $\Delta \ln P_j$ and Δs_j. Thus, identification arguments for these parameters ask whether changes in the residual neighborhood

[22] In a frictionless world, a no-arbitrage condition suggests that rental profits from risk-neutral landlords in perfect competition should be equivalent to interest income. Simply using the prevailing bank lending rates during the 1930s to scale home values (roughly 4–7% according to Basile, Landon-Lane, and Rockoff, 2010) is problematic because I do not observe costs facing landlords and thus would tend to overstate ρ.

choice probability Δu_{rj} are correlated with the set of instruments $\boldsymbol{Z}_j$, conditional on controls $\boldsymbol{X}_j$.[23]

An immediate question relates to what constitute valid controls $\boldsymbol{X}_j$. Allowing neighborhood choices to agnostically evolve according to controls is not well-motivated by the model, but in many empirical settings as well as in mine, controls improve the empirical performance of the estimators. In particular, the procedure of matching addresses to census tracts in 1930 is subject to measurement error. I will outline the controls I use and argue their necessity in my empirical specification.

Controlling for the Sum of Shares

The first set of controls I include relate generally to shift-share instruments. In a forthcoming appendix, I show that in a linear representation of rural outflows, shift-share migration instruments in the spirit of Altonji and Card (1991), Card (2001) have two components: the share-weighted sum of origin push factors and the share-weighted sum of a constant, or the sum of shares. Rhetorical justification for shift-share instruments often focuses on the former. Endogeneity concerns focus on the latter. Critics worry that the shares are potentially related to unobserved omitted variables that determine the outcome of interest.[24,25] I include the sum of shares as a control variable to partially alleviate this concern.

Controlling for 1930 Characteristics

The second set of controls relates to limitations of using migrant demand shocks as instruments for a neighborhood choice model. In the instrument set, I include both black and white shift-share demand shocks as well as the demand shocks interacted with the 1930 black share as instruments. These interactions are necessary: as I will show, the model's predicted heterogeneous effects on the neighborhood black share materializes empirically. As such, I include a main effect for the 1930 tract black share as a control variable. I also include the 1930 tract population and the 1930 median log housing cost as controls, which improve the power to detect the migrants' effects on prices in the first stage regressions.

From the perspective of the model specified in changes in equation (1.18), one can include any pre-determined lagged characteristic, including both the sum of shares and the lags of the endogenous variable, under the assumption of sequential exogeneity. That is, the unobserved determinants of 1940 choice probabilities are unrelated to the 1930 characteris-

[23] Converting the first differenced model to a fixed effects model is tantamount to utilizing as time-varying instruments and controls $\boldsymbol{Z}_j$ and $\boldsymbol{X}_j$ interacted with a dummy variables for the observation being in 1940.

[24] In my setting, these concerns are not without merit. In the aforementioned appendix, I show that both the black and white sums of shares are unconditionally strong predictors of population growth of both black and white residents across cities, suggesting that the sum of shares can easily proxy for growing cities generally.

[25] Identification issues related to the shares is a matter of continued debate for Bartik style instruments. See e.g. Goldsmith-Pinkham, Sorkin, and Swift (2018).

tics, conditional on the tract fixed effect.[26] More directly, identification of β and γ relies on the unobserved *changes* in neighborhood choice being uncorrelated with the instrument set conditional on 1930 characteristics.

In fact, Wooldridge (2010) suggests that these characteristics can even be included as instruments in the context of the static model. However, the impetus for including lagged endogenous variables as controls is that 1930 prices and black share are measured with error, driven in part to the procedure I implemented to reconstruct the 1940 census tracts in 1930. In turn, measurement error in the change in the endogenous variable is mechanically negatively correlated with measurement error in the control.

Identification Amid Measurement Error

Both the main effects of the demand shocks and the interactions with the 1930 black share are likely to have a mechanical unconditional correlation with the endogenous variables. The latter is clear because of the lag of the endogenous variable is in the term. But, even the former is subject to this problem. The origin shares used in equation (1.13) are constructed with the same census tract construction procedure and may in turn produce similar correlated measurement error. In sum, estimation of the model parameters relies in part on the controls' ability to absorb correlated measurement error between the endogenous variable and the instruments.

Reduced Form and First Stage Effects of Migration

Population Effects

In this section, I present reduced-form, qualitative effects of black and white migration on neighborhoods. To do so, I estimate basic ordinary least squares (OLS) regressions of the form

$$\Delta Y_j = a_{m(i)} + \sum_r \boldsymbol{b}_r' \boldsymbol{Z}_j + \boldsymbol{c}_r' \boldsymbol{Z}_j \times s_j + \boldsymbol{G}' \boldsymbol{X}_j + \varepsilon_j \tag{1.19}$$

where $\alpha_{m(i)}$ is a metropolitan area fixed effect and $\boldsymbol{X}_j$ is a vector that includes the black and white sum of shares, the tract 1930 population, and the 1930 median housing cost as controls. The model suggests that a black migrant is likely to have different effects in relatively black and relatively white neighborhoods. Thus, I report results both with and without interacting the instruments with s_j. As I will show, there is important heterogeneity along the gradient of 1930 black share.

Table 1.5 presents the relationships between changes in black, white, and total population and the instruments. The coefficients are not directly interpretable since they reflect not only the effect of migrant inflows but also measurement error that arises from the last name procedure, mismeasurement in county outflows, and the leave-out procedure mechanically scaling down the magnitude of the instrument. Nonetheless, qualitatively, columns 1 and 2

[26] I show this formally in a forthcoming appendix.

show that black and white migrant shocks are associated with black and white populations for the average tract and smaller declines in white and black populations.

However, these effects by themselves mask tremendous heterogeneity. Columns 4–6 report models where the instruments have been interacted with the 1930 black share. The second panel of the table reports linear combinations of the coefficients to produce implied effects in relatively white and relatively black neighborhoods. Whereas column 2 reports estimates for a typical tract, the estimates in column 4–5 suggest that white residents leave more rapidly than black residents enter, in these neighborhoods. The bottom panel of the table reports joint tests of significance among all the coefficients. The first row labeled "All Instruments" tests the joint significance of all coefficients in the top panel of the table, while the second two rows test the joint significance of Z_B and Z_W and their interactions, respectively. Consistent with strong heterogeneity, the F-statistic from the Wald joint test of significance on the black instrument increases from 0 to 27 between columns 2 and 5.

Relative to black demand shocks, white demand shocks appear to have less predictive power in general reflecting white rural migration being less directed than black migration during this period. Nonetheless, even in a relative sense, white demand shocks do not seem to have a qualitatively important impact on tract black populations operating only to change white populations.

First Stage Regressions

Table 1.6 mirrors Table 1.5 except that it reports coefficients of regressions for housing cost and neighborhood black share. Similar important heterogeneity can be seen across the gradient of black share of neighborhoods. Because of scale, regressors have been divided by 1,000 before estimation.

Consistent with the model, costs fall in response to black migrant shocks in relatively white neighborhoods and increase in relatively black neighborhoods. Additionally, neighborhood minority share evolves in exactly the way one would expect: black migrants increase the black share in relatively white neighborhoods but have no effect in relatively black neighborhoods, and vice versa for white migrants. Interestingly, though white migrants had a consistently positive effect on the neighborhood total population, white migrants seem to have a negative effect on housing costs. Though they do not produce nearly as much variation as the black demand, white demand shocks are nonetheless always statistically significant at conventional levels.

Estimates of Preference Parameters

Estimation Procedure

I estimate equation (1.5) in first differences via two-sample two-stage least squares (2SLS) in order to keep the first stage estimates the same across all samples. Because there are many neighborhoods with few or no black residents, the sample of tracts used for the second stage can be considerably smaller when estimating the black preference parameters versus when estimating white preference parameters. These smaller samples can be particularly meaningful when estimating the interaction terms in the first stages.

The two-step procedure, however, poses no particular threat to the second stage estimation under the I.I.A assumption of the multinomial logit model. I report robust standard errors derived for two sample 2SLS according to the procedure suggested by Pacini and Windmeijer (2016), which also incorporates the potential covariance in first stage and reduced form parameters induced by the partially or completely overlapping samples.

2SLS Estimates of Parameters Governing Prices and Racial Composition

Table 1.7 reports estimated preference parameters over housing costs for black households in panel A and white households in panel B. Column (1) estimates the model on shares pooling the location choices of households in all occupation, column (2) focuses on heads of household who are in low-skill occupations typical of black workers, and column (3) estimates parameters pooling blue- and white-collar workers.

The estimates of the model are consistent with households having downward sloping demand. Interestingly, white household utility appears to be somewhat more sensitive than black households to price.

Starker contrasts emerge in the preferences over neighborhood black share. My preferred estimates for black residents are in column (2) where I construct location shares using households in the three broad occupation groups that represent most black households. These estimates suggest that black households have no particular affinity for more or less black neighborhoods. The pooled estimates that include families of higher-skilled black men in column (1) are potentially confounded by taste heterogeneity for certain neighborhoods, but if anything point to black households having an affinity for black neighbors.

Corresponding estimates for white households are consistent with a prior literature that find that white households have a high willingness to pay to avoid growing minority shares. Interpreting these results through the model, in the bottom panel, I report elasticities that reflect how households would need to be compensated for a 1 p.p. increase in the neighborhood black share to keep utility constant. For white households, these estimates sugges that a 1 p.p. increase in the black share would need to be offset by an approximately 0.7–1% decline in house prices to keep white households indifferent.

Estimates of the Correlated Random Effects

Table 1.7 speaks to the prior literature and by using migrant shocks, estimates preferences that do not confound potential supply and demand factors. However, these analyses by themselves are not able to quantify segregation. As outlined in section 1.2, I first present parameter estimates from a correlated random effects model. I then use these counterfactual distributions directly in a decomposition of the KL divergence, which I use as a measure of segregation.

Estimation Procedure

Direct estimation of a CRE version of the model in equation (1.5) is infeasible in my setting. The demand shocks are suitable instruments for the model in changes in equation (1.18), but the a CRE model requires instruments for the 1930 levels of $\ln P$ and s.

However, under the identification assumptions laid out before, the 2SLS regressions in first differences give consistent estimates of β and γ. Thus, a consistent estimator for u_{rjt} is

$$\hat{u}_{rjt} = \ln \hat{\pi}_{rjt} - \hat{\beta}_r \ln P_{jt} - \hat{\gamma}_r s_{jt}. \tag{1.20}$$

To obtain the parameters of the CRE model, I can then estimate a cross-sectional OLS regression pooling the two periods

$$\hat{u}_{rjt} = \vartheta_{rdt} + \underbrace{\boldsymbol{\Omega_r'X_j} + \boldsymbol{\Upsilon_r'Z_j} + \psi_{rj}}_{\alpha_{rj}} + \xi_{rjt} \tag{1.21}$$

where $\boldsymbol{X}_j$ includes the sum of shares and the 1930 total population.

Identification of the variance components in equation (1.8) comes from the residuals

$$\hat{e}_{rjt} = \psi_{rj} + \xi_{rjt}.$$

This residual has two components. Using the three assumptions in equation (1.8) in section 1.2, one can write the full variance-covariance matrix of the errors:

$$\boldsymbol{Var}\begin{pmatrix} e_{Bj,1930} \\ e_{Bj,1940} \\ e_{Wj,1930} \\ e_{Wj,1940} \end{pmatrix} = \begin{pmatrix} \sigma_B^2 + \varsigma_B^2 & & & \\ \sigma_B^2 & \sigma_B^2 + \varsigma_B^2 & & \\ \sigma_{BW} + \varsigma_{BW} & \sigma_{BW} & \sigma_W^2 + \varsigma_W^2 & \\ \sigma_{BW} & \sigma_{BW} + \varsigma_{BW} & \sigma_W^2 & \sigma_W^2 + \varsigma_W^2 \end{pmatrix}$$

where $\boldsymbol{Cov}\{\xi_{Bjt}, \xi_{Wjt}\} = \varsigma_{BW}$. The coefficients I need for the counterfactuals are σ_B^2, σ_W^2, and σ_{BW} to plug into equation (1.9). The population variance-covariance matrix directly maps into sample covariances of the regression residual. That is

$$\hat{\sigma}_r = \frac{1}{N-K} \sum_j \hat{e}_{rj1930} \cdot \hat{e}_{rj1940}, \tag{1.22}$$

$$\hat{\sigma}_{BW} = \frac{1}{2N-K} \sum_j \left(\hat{e}_{Bj1930} \cdot \hat{e}_{Wj1940} + \hat{e}_{Bj1940} \cdot \hat{e}_{Wj1930} \right). \tag{1.23}$$

I compute the variance and covariance components using the sample of tracts where there are both white and black residents for a given occupation group.[27]

[27]Utilizing the full sample of neighborhoods for white residents has no substantial effect on the variance estimates. Results available upon request.

Estimates of Random Effects and Variance Components

Table 1.8 summarizes the estimated correlated random effects parameters. The top panel reports estimates of the random effects variance among different occupation groups of households. The first line in each row reports the raw variance estimate, and the second line in each row reports the correlation of the composite residuals used to estimate the covariance. For example, the first row in column (2) reports a random effect variance $\hat{\sigma}_B^2 = 0.331$ and a sample correlation $\boldsymbol{Corr}\left[\hat{u}_{Bj1930}, \hat{u}_{Bj1940}\right]$ of 0.735. The first two rows across the sample suggest that neighborhood unobservables are highly correlated across periods, which implies that static characteristics of the neighborhood explain a large portion of the variance in unobservable neighborhood choice.

Interestingly, the white and black unobservables σ_{BW} are strongly *positively* correlated. Ex ante, the opposite result is equally plausible. For instance, one can imagine that neighborhoods that have better schools for whites would have worse schools for blacks, reflecting educational resource inequality growing with the white unobservable. This does not seem to be the case. In the sample of tracts where there are both black and white households, the probability that a black family choose a neighborhood grows with the probability that a white family chooses the neighborhood, conditional on the neighborhood's black share and price.

Extending this analysis to include observables, the bottom panel reports the model implied covariances between α_{Bj} and α_{Wj}. The first line in each row reports the raw covariance, and the second line divides this covariance by the variance of α_{Wj}. This gives the interpretation of a regression coefficient where the dependent variable is the component indicated by the label in the row, and the the independent variable is α_{Wj}.

Focusing on the group of households in low-skilled occupations in column (2), one can see that α_{Bj} has a strong positive correlation with α_{Wj}. The implied regression coefficient suggests that a one unit increase in the log odds for white households driven by α_{Wj} corresponds to a 0.8 increase in the log odds for black households.

The second row focuses only on the components that are driven by the observable covariates. The implied regression coefficient suggests that almost 90% of the variation in the implied 0.8 coefficient is driven by observable characteristics.

Altogether, this analysis could have very easily shown that among neighborhoods with both white and black families, unobservables drive segregation by being negatively correlated. This does not seem to be the case. The between-race correlation of the random effects seems to suggest that the random effect can be most accurately thought of a shared amenity rather than amenities and disamenities that are race-specific.

Overall Decomposition

Using the CRE model in conjunction with the underlying multinomial logit I generate model predicted mean utilities $\widehat{\log \delta}$ and the implied inclusive value to make the choice probabilities sum to 1 according to the procedure described at the end of section 1.2. I then replace the terms in equation (1.2) to decompose the KL divergence into components interpretable as driven by supply and demand factors.

Table 1.9 reports the results of the decomposition using these predicted choice probabilities. The top panel of the table presents the overall divergence, and the bottom panel presents each component. In each row, the second line is the fraction of the total divergence explained by the component.

The first line in the bottom panel compares the choice probabilities of black households to those of counterfactual black households. The latter group, by definition has the same exact preference parameters, which I interpret as driven by constrained supply. The second line is the remainder and compares the counterfactual demand to white demand. Here, because the counterfactual is extrapolated to include all neighborhoods, there are no supply constraints.

The results in this table consistently show that roughly one of observed segregation within an occupation group is driven by supply. The accumulated results to this point are consistent with a supply based explanation. If segregation were driven purely by preferences, the willingness of white households to pay to avoid black families would be capitalized into price—one should expect that white families are concentrated in expensive all white neighborhoods. This is partly true. Recall from table 1.4 that white neighborhoods tend to be more expensive than black neighborhoods.

However, the observed price gap is not large enough to fully deter black families from moving into those neighborhoods. Lifting the constraints through the counterfactual exercise shows that black families would move to those neighborhoods without those prohibitions.

1.6 Conclusion

The analysis of this paper suggests that roughly one half of observed segregation in 1940 can be explained by divergent preferences over neighborhoods. I estimate parameters of a multinomial logit demand model identified using information from migrant demand shocks. To avoid migrant demand shocks being correlated with neighborhood unobservables, I utilize the predictable component of migration by connecting residents in 1930 to rural origin counties using surname distributions.

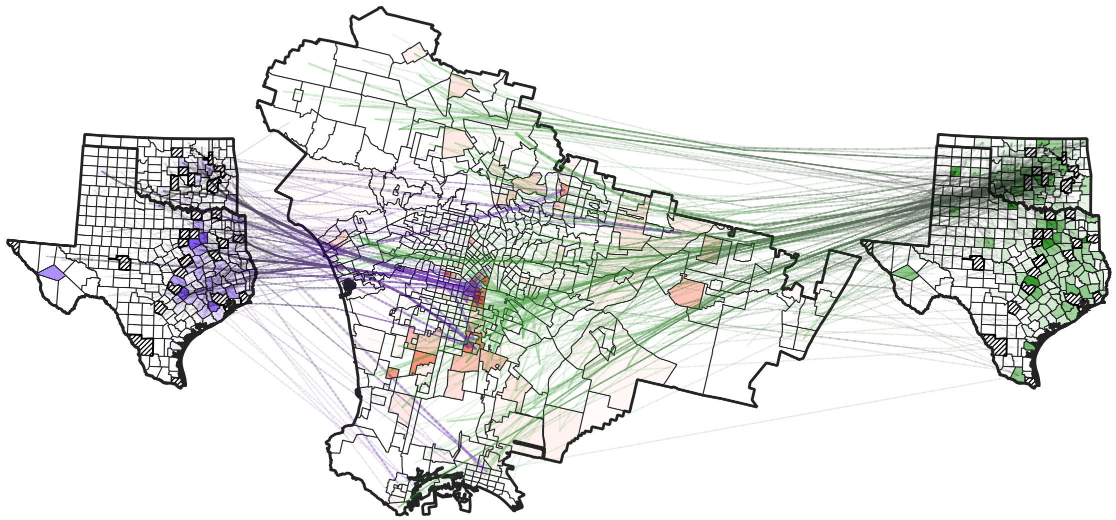
Structural estimates suggest that white households willingness-to-pay to avoid a 1 p.p. increase in black neighbors is roughly 1% of the house value. While preferences over the black share of the neighborhood diverge between races, they seem to converge on everything else: between races, the implied price elasticities of demand with respect to housing are roughly the same, and unobservable neighborhood characteristics are highly correlated.

The measurement of these preference parameters in conjunction with neighborhoods with essentially no black residents are highly suggestive of supply constraints driving a large share of segregation. Constructing counterfactual demands from the parameter estimates and incorporating them in the decomposition of the KL divergence suggests the share of segregation explained by formal and informal constraints is comparable to the preference based explanations.

Was the persistent concentration of high poverty black neighborhoods in inner cities the inevitable consequence of decentralized, individualistic decisions of households? Boustan's (2010) research suggests that suburbanization in the latter half of the 20th century was driven in large part to white flight. While my partial equilibrium model does not account for dynamics, it does suggest state dependence on the neighborhood black share, implying that at least some of post-war white flight has origins in supply-driven segregation prior to the second World-War.

1.7 Figures

Figure 1.1: Rural-to-Urban Migrant Flows from Texas and Oklahoma to Los Angeles, 1935–1940



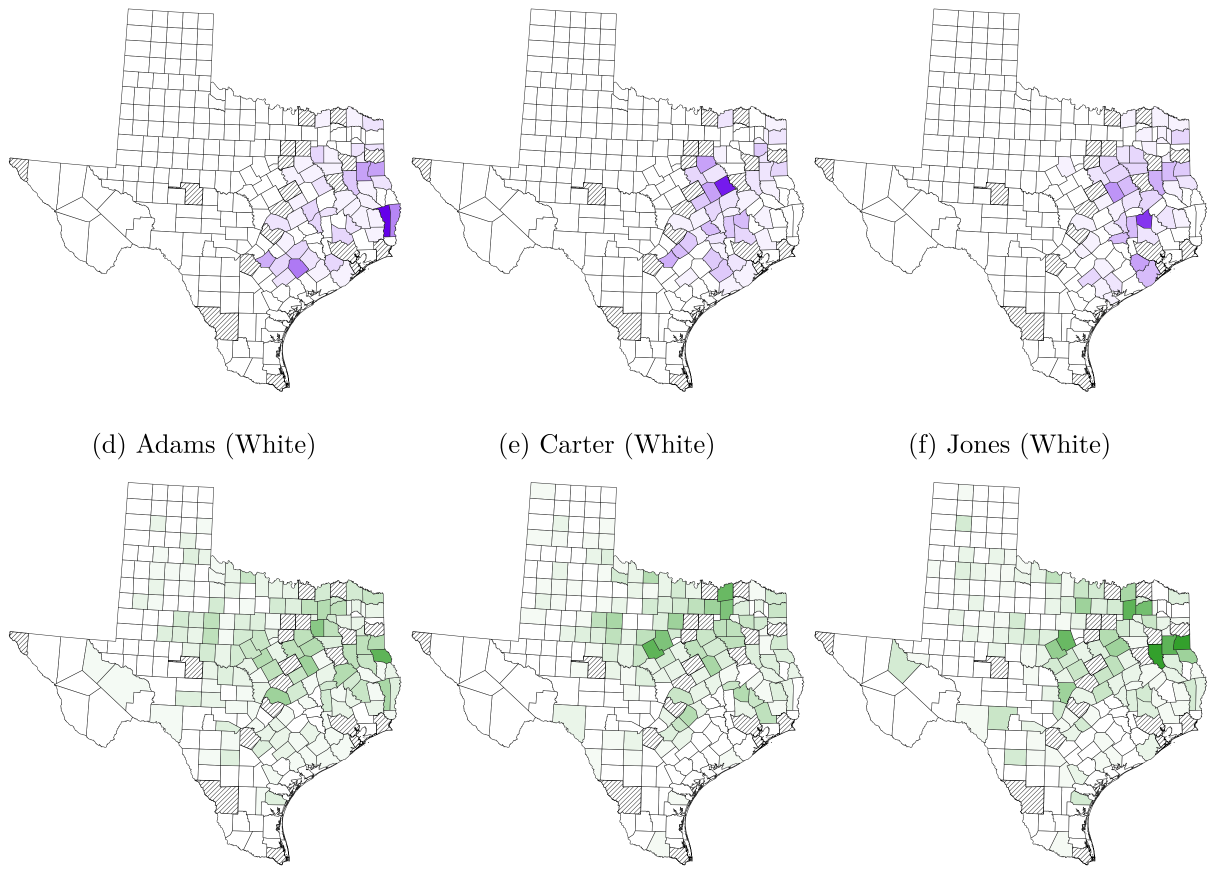
[a] Black flows plotted in purple, and white flows plotted in green.

[b] Origin counties shaded in purple and green with the intensity corresponding to the total county outflows of blacks and whites to major cities with census tracts, respectively. Cross-hatched counties are urban counties.

[c] Census tracts in Los Angeles shaded in red according to the tract share of black residents in 1930.

[d] Flows bundled via algorithm documented in Graser, Schmidt, Roth, and Brändle (2017) using software from https://github.com/dts-ait/qgis-edge-bundling.

Figure 1.2: Comparison of Three Common Last Names in Texas



(a) Adams (Black) (b) Carter (Black) (c) Jones (Black)

(d) Adams (White) (e) Carter (White) (f) Jones (White)

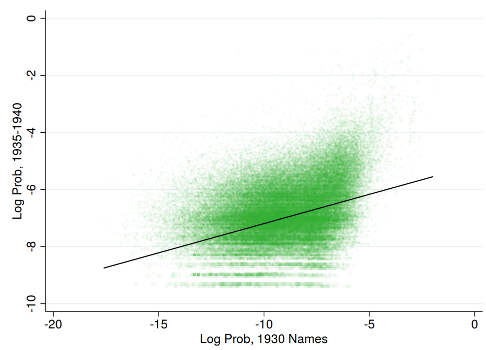
[a] The first row of figures plots the residential share of non-migrant blacks. The second row of figures plots the distribution for whites.

[b] Cross-hatched counties are urban counties.

Figure 1.3: Current and Surname-Constructed Flows Probabilities for Counties in the Top Quartile of Outflows

(a) Black

(b) White



[a] Each point is an origin county-destination tract pair.

[b] The vertical axis measures the log of the 1935–1940 flow probabilities from the 1940 census, and horizontal axis represents the log of the probabilities estimated from the surname procedure in the 1930 census. See text for details.

[c] The size of each point is weighted by the total race-specific migrant outflows from the origin county. The sample of points come from rural counties in the top quartile of race-specific outflows for display purposes.

1.8 Tables

Table 1.1: Migrant Inflows to 48 Major Cities with Census Tracts by Census Region, 1935–1940 (Thousands)

(a) Black

	All	South	West	Midwest	Northeast
Total	250	105	18	55	72
From rural counties	116	69	4	22	22
in the South	110	68	3	19	20
in the West	1	0	1	0	0
in the Midwest	4	0	0	3	0
in the Northeast	2	0	0	0	2

(b) White

	All	South	West	Midwest	Northeast
Total	3,960	767	929	952	1,311
From rural counties	1,087	344	308	297	138
in the South	440	315	49	49	28
in the West	180	5	165	6	4
in the Midwest	355	19	90	235	11
in the Northeast	112	6	4	6	96

[a] Rural counties are those without a census incorporated place. See text for details.
[b] Flows are from a retrospective question in the 1940 census asking about respondents' location five years prior.
[c] All flows including the total only include individuals for whom an origin county could be ascertained.

Table 1.2: Regressions of 1935–1940 Flow Probabilities on Surname-Constructed Probabilities

(a) Black

	(1)	(2)	(3)	(4)	(5)
Past Flowrate	1.979	2.001	2.035	1.952	1.999
	(0.145)	(0.144)	(0.173)	(0.176)	(0.422)
Tract(Dest) FE		✓	✓	✓	
State(Origin) × Metro(Dest) FE			✓		✓
County(Origin) × Metro(Dest) FE				✓	
State(Origin) × Tract(Dest) FE					✓
R^2	0.0996	0.121	0.126	0.166	0.271
Obs	19,185,552	19,185,552	19,185,552	19,185,552	19,166,328

(b) White

	(1)	(2)	(3)	(4)	(5)
Past Flowrate	1.601	1.581	1.680	1.625	6.030
	(0.133)	(0.137)	(0.184)	(0.185)	(0.809)
Tract(Dest) FE		✓	✓	✓	
State(Origin) × Metro(Dest) FE			✓		
County(Origin) × Metro(Dest) FE				✓	✓
State(Origin) × Tract(Dest) FE					✓
R^2	0.187	0.200	0.211	0.270	0.444
Obs	23,126,472	23,126,472	23,126,472	23,126,472	23,116,860

[a] The unit of observation is an origin county-destination census tract pair.
[b] Regressions weighted by total rural-to-urban origin county migrant outflows.
[c] Robust standard errors clustered by origin county reported in parenthesis.

Table 1.3: Occupation Distribution by Race, 1940

	Black	White
All Households	789,140	8,358,405
...with employed male head of household, age 18–55,	488,045	5,350,075
... with wife and at least one child	212,624	3,342,510
... in tracts with at least 10 with same occ. $\times$ race	207,701	3,340,763
Broad Occupation Shares		
"Black" Occupations		
Laborers	46.4	9.4
Services	21.6	6.7
Operators	18.4	22.6
Other Occupations		
Craftsmen	7.8	23.0
Clerical	2.6	7.9
Professional	1.5	6.1
Sales	1.1	12.2
Managers	0.7	12.2

[a] The top panel reports counts of households living in one of 48 tracted metropolitan areas in 1940.

[b] The bottom panel reports the shares of families (a cohabiting husband, wife, and child) living in tracts with at least 10 other families of the same broad occupation and race in both 1930 and 1940.

Table 1.4: Neighborhood Characteristics of Median Black and White Families, 1940

	All	White	White (Low Skilled Occs)	Black
Characteristics of Neighborhood Housing				
Median Housing Cost	3,402.3	3,500	3,037.8	2,430.2
Median Home Values (Owners)	3,500	3,500	3,000	2,700
Median Rent (Renters)	27	28	25	20
Home Ownership Rate	0.283	0.304	0.286	0.141
Characteristics of Neighbors				
Black Share	0.00289	0.00195	0.00180	0.730
Mean Household Size	3.607	3.590	3.677	3.763
Median Household Income	1,500	1,551	1,500	979
Share Employed (Head)	0.704	0.711	0.690	0.636
Share Employed in Low-skilled Occs. (Head)	0.280	0.269	0.305	0.401
Average Years of Education (Head)	8.357	8.506	7.949	7.050
Sum of Weights	33,283,800	29,920,195	9,771,394	3,237,710

[a] Each column reports a weighted median of tract characteristics where weights are the number of families described by the column labels.

[b] Low-skilled occupations include laborers, service workers, and operators.

[c] Housing cost combines both home values and rents into a single measure. See text for details.

Table 1.5: Reduced Form Effects of Migrants on Population

	(1) Black	(2) White	(3) Total	(4) Black	(5) White	(6) Total
Coefficient Estimates						
Z_B	8.856	-0.813	7.958	21.06	-31.24	-10.46
	(2.284)	(2.152)	(2.186)	(5.934)	(5.080)	(5.395)
Z_W	-2.093	4.436	1.976	-1.866	3.461	1.218
	(0.431)	(1.804)	(1.855)	(0.452)	(1.785)	(1.845)
$Z_B \times s$				-18.06	36.92	19.03
				(6.453)	(5.097)	(5.796)
$Z_W \times s$				5.897	9.690	15.87
				(3.092)	(3.264)	(4.032)
Implied Effects @ $s = 0.2$						
Z_B				17.45	-23.85	-6.655
				(4.783)	(4.164)	(4.343)
Z_W				-0.687	5.399	4.391
				(0.725)	(2.035)	(2.180)
Implied Effects @ $s = 0.8$						
Z_B				6.608	-1.698	4.762
				(2.423)	(2.096)	(2.061)
Z_W				2.851	11.21	13.91
				(2.466)	(3.482)	(4.074)
Tracts	6132	6132	6132	6132	6132	6132
Wald F-statistics and p-values						
All Instruments	21.39	3.372	6.678	24.73	15.33	7.683
	⟨0.000⟩	⟨0.034⟩	⟨0.001⟩	⟨0.000⟩	⟨0.000⟩	⟨0.000⟩
Black Effects	15.03	0.143	13.25	6.895	27.05	9.693
	⟨0.000⟩	⟨0.705⟩	⟨0.000⟩	⟨0.001⟩	⟨0.000⟩	⟨0.000⟩
White Effects	23.62	6.046	1.135	9.602	5.250	7.780
	⟨0.000⟩	⟨0.014⟩	⟨0.287⟩	⟨0.000⟩	⟨0.005⟩	⟨0.000⟩

[a] Robust standard errors reported in parentheses, p-values reported in angular brackets.
[b] The dependent variables are tract decadal changes in black, white and total population, respectively.
[c] Z_B and Z_W are black and white demand shocks. See text for details.
[d] All equations include metropolitan area fixed effects and controls for the 1930 population, black share, the black and white sum of shares, and median log housing cost.
[e] The Wald test for "All Instruments" tests the joint significance of all coefficients in the top panel. "Black Effects" and "White Effects" test Z_B and Z_W and their interactions (if applicable), respectively.

Table 1.6: First Stage Regressions

	(1) Cost	(2) Share	(3) Cost	(4) Share
Coefficient Estimates				
$Z_B/1,000$	0.188	0.102	-1.486	2.899
	(0.275)	(0.141)	(0.437)	(0.365)
$Z_W/1,000$	0.0789	-0.147	0.0518	-0.0679
	(0.168)	(0.0426)	(0.170)	(0.0370)
$Z_B/1,000\times s$			2.557	-3.598
			(0.473)	(0.388)
$Z_W/1,000\times s$			-1.047	-0.276
			(0.344)	(0.186)
Implied Effects @ $s=0.2$				
$Z_B/1000$			-0.974	2.179
			(0.368)	(0.294)
$Z_W/1000$			-0.158	-0.123
			(0.168)	(0.0475)
Implied Effects @ $s=0.8$				
$Z_B/1000$			0.560	0.0201
			(0.275)	(0.140)
$Z_W/1000$			-0.786	-0.288
			(0.288)	(0.147)
Tracts	6132	6132	6132	6132
Wald F-statistics and p-values				
All Instruments	0.329	6.427	19.76	35.51
	$\langle 0.719\rangle$	$\langle 0.002\rangle$	$\langle 0.000\rangle$	$\langle 0.000\rangle$
Black Effects	0.470	0.525	15.51	44.60
	$\langle 0.493\rangle$	$\langle 0.469\rangle$	$\langle 0.000\rangle$	$\langle 0.000\rangle$
White Effects	0.222	11.97	4.698	3.392
	$\langle 0.638\rangle$	$\langle 0.001\rangle$	$\langle 0.009\rangle$	$\langle 0.034\rangle$

[a] See the table notes a, c, d, and e from table 1.5.

[b] The dependent variables are tract decadal changes in median log housing cost and neighborhood black share, respectively. See section 1.4 for details on construction of housing costs.

[c] Z_B and Z_W are divided by 1,000 before estimation for reporting.

Table 1.7: Preference Parameters for Broad Occupation Groups

(a) Black

	(1) Pooled	(2) Low Skilled Occ	(3) Other Occ
Log Housing Costs	-1.228	-1.906	-0.284
	(0.575)	(0.553)	(0.452)
Black Share	0.887	-0.0113	0.350
	(0.701)	(0.704)	(0.639)
Tracts	1196	1087	490
Elasticity	0.722	-0.00593	1.230
	(0.879)	(0.368)	(4.092)

(b) White

	(1) Pooled	(2) Low Skilled Occ	(3) Other Occ
Log Housing Costs	-3.542	-4.109	-2.743
	(0.950)	(1.026)	(0.828)
Black Share	-2.473	-3.982	-2.134
	(1.011)	(1.109)	(0.928)
Tracts	6049	5750	6015
Elasticity	-0.698	-0.969	-0.778
	(0.165)	(0.143)	(0.187)

[a] Robust standard errors adjusted for two-step procedure according to Pacini and Windmeijer (2016) reported in parentheses. First stage coefficients are reported in table 3.5.

[b] Reported elasticities are the percentage change in housing costs needed to offset a 1 p.p. increase in the black share and keep an average household indifferent to the neighborhood. Standard errors are computed using the delta method.

[c] Low skilled occupations are laborers, service workers, and operators.

Table 1.8: Scaled Covariances of Correlated Random Effects

	(1) Pooled	(2) Low Skilled Occ.	(3) Other Occ.
Estimated Covariances and {Correlations}			
σ_B^2	0.416	0.494	0.235
	{0.742}	{0.776}	{0.703}
σ_W^2	1.153	0.855	0.855
	{0.724}	{0.775}	{0.820}
σ_{BW}	0.367	0.449	0.153
	{0.390}	{0.538}	{0.262}
Covariances, Raw and [Implied Regression Coefficents]			
α_{Bj}	0.315	0.650	0.145
	[0.198]	[0.531]	[0.128]
$\left(A_B\bar{\boldsymbol{Z}}_j + C_B\bar{\boldsymbol{X}}_j\right)$	-0.0523	0.201	-0.00774
	[-0.0328]	[0.164]	[-0.00683]
σ_{BW}	0.367	0.449	0.153
	[0.231]	[0.367]	[0.135]
Tracts	1118	915	396

[a] The top panel reports estimates of the variance components of the permanent neighborhood unobservable from a correlated random effects model. The first line in each row is the point estimate of the variance, and the second line in each row is the correlation between the two residuals from which the variance is estimated. For example, the first line in the first row is the covariance between the 1930 and 1940 black residuals, and the second line is the correlation coefficient. See text for details.

[b] σ_B and σ_W are estimated from the covariance between the 1930 and 1940 black and white residual, respectively. σ_{BW} is a pooled covariance of the 1930 black residual and 1940 white residual and 1930 white residual and 1940 black residual. The correlation coefficient for the latter is an average of the two estimates.

[c] All covariances are estimated on the subset of tracts for which there are both black and white residuals.

[d] The bottom panel reports the covariances between α_{Bj} and α_{Wj}, both raw, and scaled by the variance of α_{Wj}. The scaled covariances have the interpretation of a regression coefficient.

Table 1.9: Decomposition of KL Divergence

	(1) Pooled	(2) Low-Skilled Occ.	(3) Other Occ.
Divergence	116.7	105.1	129.5
	[1]	[1]	[1]
$\pi_{Bj} \ln \left(\pi_{Bj}/\hat{\pi}_{Bj}^{CF}\right)$	55.60	49.94	76.02
	[0.476]	[0.475]	[0.587]
$\pi_{Bj} \ln \left(\hat{\pi}_{Bj}^{CF}/\pi_{Wj}\right)$	61.14	55.16	53.53
	[0.524]	[0.525]	[0.413]

[a] The top line reports the actual divergence from the black multinomial distribution to the white distribution. The second and third line reports the divergence between the actual black distribution and a counter-factual distribution computed via the random effects model and the divergence between the counter-factual distribution and the white distribution.

[b] The first row in each line reports the component of the Kullback-Leibler divergence. The second row in each line reports the share of the actual divergence explained by that component.

[c] The sample of tracts over which the divergence is calculated are those with some white residents. Probabilities are renormalized to sum to 1 according to a procedure outlined in the text.

Chapter 2

Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata

2.1 Introduction

The shift out of agriculture and into other more "modern" sectors (e.g., manufacturing) has long been viewed as central to economic development. This structural transformation was a focus of influential early scholarship (Rosenstein-Rodan 1943; Lewis 1955; Rostow 1960; Pack 1972; Kuznets 1973; Johnston and Kilby 1978; Schultz 1988) with the issue even stretching back to Soviet debates over whether to "squeeze" farmer surplus to hasten industrialization (Preobrazhensky 1921). A more recent macroeconomic empirical literature has documented that the share of labor in the agricultural sector correlates strongly with levels of per capita income, and yet agricultural workers are many times more productive in rich countries. This creates a double disadvantage: agricultural work tends to be far less productive in poor countries, yet their workforce is concentrated in this sector.[1]

Whether or not these gaps reflect labor market frictions that can be remedied by policy depends on the extent to which these productivity gaps across sectors can reasonably be viewed as *causal impacts* rather than mainly reflecting *worker selection*. By a causal impact of sector, we mean that a given worker employed in the non-agricultural (or urban) sector is more productive than the same worker employed in the agricultural (rural) sector. In contrast, selection implies that observed differences in productivity are driven by the fact that workers of varying ability and skill are concentrated in certain sectors and not by inherent differences in the sectors themselves.

Gollin, Lagakos, and Waugh (2014, henceforth GLW) and Young (2013) explore this identification issue by applying different methods to adjust for differences in observable

[1]See Appendix Figures A.1 and A.2, respectively, for an illustration of these differences using cross-country data.

characteristics, particularly education. In their main contribution, GLW show that accounting for differences in hours worked and average worker schooling attainment across sectors reduces the average agricultural productivity gap by a third, from roughly 3 to 2. GLW remain agnostic regarding the causal interpretation of the large agricultural productivity gaps that they estimate. Nonetheless, if individual schooling largely addresses selection biases by capturing the most important dimensions of worker skill, GLW's estimates would imply that the causal impact of moving workers from agriculture to the non-agricultural sector in low-income countries would be to roughly double productivity, a large effect.

Young (2013) similarly examines the related question of urban-rural differences in consumption (as proxied by measures of household asset ownership, education, and child health), rather than productivity, and similarly finds large cross-sectional gaps.[2] Using Demographic and Health Surveys that have retrospective information on individual birth district, Young shows that rural-born individuals with more years of schooling than average in their sector are more likely to move to urban areas, while urban-born individuals with less schooling tend to move to rural areas. Diverging from GLW, Young argues that a simple model of sorting on education and relatively high demand for skill in urban areas can fully explain urban-rural consumption gaps.[3]

This paper examines whether measured productivity gaps are causal or mainly driven by selection using long-term individual-level longitudinal (panel) data on worker productivity. We focus on two country cases—Indonesia and Kenya—that have long-term panel micro data sets with relatively large sample sizes, rich measures of earnings in both the formal and informal sectors, and high rates of respondent tracking over time.[4] We examine whether workers who changed sector correspondingly experienced the large increase in wages suggested by existing estimates. We find that the raw mean differences in productivity can be reduced by roughly 80% with the inclusion of individual fixed effects, which capture all time-invariant dimensions of worker heterogeneity.

For both countries, we start by characterizing the nature of selective migration between non-agricultural versus agricultural economic sectors, and between urban versus rural residence. Like Young (2013), we show that individuals born in rural areas who attain more

[2]While Young (2013) focuses on urban-rural gaps, he sometimes uses data on non-agricultural vs. agricultural differences when urban-rural data is missing; GLW similarly use urban-rural data when they lack data on agriculture.

[3]Porzio (2017) argues that a model of worker sorting can explain a large share (roughly 40%) of intersectoral productivity gaps, considering agriculture as well as a range of non-agricultural sectors. Lagakos and Waugh (2013) similarly model how worker sorting across sectors could generate sectoral productivity differences in equilibrium.

[4]The datasets, the Indonesia Family Life Survey and Kenya Life Panel Survey (henceforth "IFLS" and "KLPS"), are described in greater detail below. There are other panel data sets where similar approaches could be employed, for instance, the Mexican Family Life Survey; we leave this for future work. It is worth noting that Mexico is a member of the OECD and is considerably richer in per capita terms than Indonesia or Kenya. In related work, Alvarez (2018) finds substantial narrowing of productivity gaps in Brazil with the inclusion of individual fixed effects, albeit only using formal sector wage data, and Herrendorf and Schoellman (2018) employ cross-sectional microdata to assess sectoral differences in human capital.

schooling are significantly more likely to migrate to urban areas and are also more likely to hold non-agricultural employment, while those born in urban areas with less schooling are more likely to move to rural areas and into agriculture. We exploit the unusual richness of our data, in particular, the existence of measures of cognitive ability (a Raven's Progressive Matrices score), to show that those of higher ability in both Indonesia and Kenya are far more likely to move into urban and non-agricultural sectors, even conditional on educational attainment. This is a strong indication that conditioning on completed schooling is insufficient to fully capture differences in average worker skill levels across sectors.

Next, we estimate sectoral productivity differences. Treating our data as a series of repeated cross-sections we find substantial productivity gaps of 70 to 80 log points, echoing previous work. Though these are somewhat smaller than GLW's main estimates of roughly 130 log points (reproduced in Figure 2.1), recall that GLW's estimates using household survey data (like ours) also tend to be smaller. Conditioning on individual demographic characteristics (age and gender) as well as hours worked and educational attainment narrows the gap, but it remains large at 35 to 55 log points. Finally, including individual fixed effects reduces the agricultural productivity gap in wages to 7.8 log points in Indonesia and to 6.1 log points in Kenya. Analogous estimates show that urban productivity gaps are also reduced substantially, to zero in Indonesia and 16.5 log points in Kenya. The estimated gaps in GLW are an order of magnitude larger than our estimates.

Mirroring our main results, Hendricks and Schoellman (2017) find that estimated returns to international migration are greatly mediated by the inclusion of individual fixed effects in panel fixed effects regressions (by roughly 60%). In line with our results, Alvarez (2018) finds that wage gaps between agricultural and non-agricultural formal sector Brazilian workers are primarily due to differences in worker characteristics, rather than differential pay for similar workers across sectors. Similarly, McKenzie, Gibson and Stillman (2010) show that cross-sectional estimates of the returns to international immigration (to New Zealand) exceed those using individual panel data or those derived from a randomized lottery.

Bryan, Chowdhury and Mobarak (2014) estimate positive gains in consumption (of roughly 30%) in the sending households of individuals randomly induced to migrate to cities within Bangladesh. Using data from Tanzania and observing individuals at two points in time, Beegle, de Weerdt and Dercon (2011) estimate consumption gains of 36% among those who moved away from their origin area.[5] We improve on the latter study by observing the same individuals at many points in time, allowing us to include time fixed effects to absorb covariate shocks, and by using richer individual-level information on productivity, sector of employment, and cognitive ability. Bazzi, Gaduh, Rothenberg and Wong (2016) argue that cross-sectional estimates of productivity differences across rural areas within Indonesia are likely to overstate estimates derived from panel data using movers. While Bryan and Morten (2017) suggest there are substantial gains to removing migration frictions in Indonesia, gains are associated with moves across region rather than urban residence per se. Related studies

[5]Estimating urban-rural wage gaps is challenging in their setting since only 138 individuals are observed in urban areas.

on selective migration include Chiquiar and Hanson (2005), Yang (2006), Kleemans (2016), and Rubalcava, Teruel, Thomas and Goldman (2008).

A limitation of the current study is its focus on two countries. This is due to the relative scarcity of long-run individual panel data sets in low-income countries that contain the rich measures necessary for our analysis. That said, the finding of similar patterns in two countries with large populations (250 million in Indonesia, 45 million in Kenya) – together with the similar results for formal sector wage gaps in Brazil (Alvarez 2018) – suggests some generalizability across different world regions.

Another important issue relates to the local nature of our estimates, namely, the fact that the fixed effects estimates are derived from movers, those with productivity observations in both the non-agricultural and agricultural (or urban and rural) sectors. The inclusion of individual fixed effects addresses a large class of omitted variable biases that would cause one to incorrectly attribute the inherent productivity of certain individuals—their absolute advantage—to the sector itself. However, because sector itself is a choice, it is possible that productivity gains could be different among non-movers, an issue we discuss in Section 2.2 below. Within, a typical Roy (1951) model of self-selection would suggest that the movers are also selected with respect to their individual-specific benefit—their comparative advantage—and thus our estimates identified from the population of movers are larger than those for the population on average. Nonetheless, we acknowledge that our estimates would be downwardly biased if individuals who are particularly well suited to urban, non-agricultural work face large barriers (e.g. credit constraints, information barriers, etc.) that are not compensated by the increase in expected wages. Similarly, it is possible that very long-run and even inter-generational "exposure" to a sector could improve individual productivity due to skill acquisition. Indeed, we find that children born in urban areas score higher on cognitive tests than those born in rural areas, even after controlling for parental education and cognitive performance. We return to issues of interpretation in the conclusion, including ways to reconcile our estimates with existing findings.

The paper is organized as follows. Section 2.2 presents a conceptual framework for estimating sectoral productivity gaps, and relates it to the core econometric issue of disentangling causal impacts from worker selection. Section 2.3 describes the two datasets (IFLS and KLPS); characterizes the distinctions between the non-agricultural and agricultural sectors, and urban vs. rural areas; and presents evidence on individual selection between sectors. Section 2.4 contains the main empirical results on productivity gaps, as well as the dispersion of labor productivity across individuals by sector. Section 2.4 presents medium- and long-term dynamic effects. The final section presents alternative interpretations of the results, and concludes.

2.2 Theoretical Framework

We present a simple framework to disentangle explanations for the aggregate productivity gap across sectors. We consider both observable and unobservable components of human capital, and whether intrinsic worker preferences for sector may bias direct measurement of the productivity gap. A standard model suggests that worker selection is most likely to bias sectoral productivity gaps upward when estimated among those moving into non-agriculture (urban areas) but lead to a downward bias when estimated among those moving into agriculture (rural areas).

The Agricultural Productivity Gap through the Lens of an Aggregate Production Function

Following Hendricks and Schoellman (2017), we denote revenue in sector s as $Q_s = K_s^{\alpha}\left(A_s H_s L_s\right)^{1-\alpha}$. Dropping subscripts for convenience, a representative firm in sector s solves:

$$\max_{K,HL} K^{\alpha}\left(\mathrm{AHL}\right)^{1-\alpha} - RK - ZHL$$

where R and Z represent returns per unit of physical capital K and a wage per efficiency unit (comprised of the product of human capital per unit of labor, H, and quantity of labor, L), respectively.

Solving the first order condition with respect to the quantity of the labor aggregate yields:

$$Z = (1-\alpha)\left(\frac{K}{Q}\right)^{\alpha/1-\alpha} A$$

Though this wage equation does not represent a closed form inverse labor demand function per se, it captures how sector-specific wages can emerge depending on the inherent productivity of the sector and how intensely it uses capital. We follow Hendriks and Schoellman (2017) and treat the capital-output ratio as exogenously determined from the perspective of the worker. Finally, we assume that labor is perfectly substitutable, but workers supply a different amount of human capital.

An individual's income in sector s is given by $Y_{is} = Z_s H_i L_{is}$. Denoting logs in lower case, the average log-income gap across the non-agricultural (n) and agricultural (a) sectors is:

$$\bar{y}_n - \bar{y}_a = \underbrace{(z_n - z_a)}_{\text{residual income gap}=\beta} + \underbrace{\left(\bar{l}_n - \bar{l}_a\right)}_{\text{labor supply gap}} + \underbrace{\left(\bar{h}_n - \bar{h}_a\right)}_{\text{human capital gap}} \tag{2.1}$$

The agricultural productivity gap is comprised of a labor supply gap, a human capital gap, and a productivity residual, β, the key parameter of interest. To the extent that there are no systematic compensating differentials, non-zero β suggests that there are labor market frictions that prevent workers switching sector driving marginal revenue products to be equalized.

We assume that individual human capital is not sector-specific and takes a Mincerian form, $H_i = \exp\left[\boldsymbol{X}_i'\boldsymbol{b} + \eta_i\right]$ where $\boldsymbol{X}_i'$ is a vector of observed characteristics (e.g., years of schooling) with corresponding returns $\boldsymbol{b}$, and η_i represents unobserved skill. Substituting into the wage equation, yields a simple regression formulation of log income in sector s:

$$y_i = z_a + \beta D_i + l_i + \boldsymbol{X}_i'\boldsymbol{b} + \eta_i \tag{2.2}$$

where D_i is an indicator for working in non-agriculture. The agricultural productivity gap becomes:

$$\bar{y}_n - \bar{y}_a = \beta + \left(\bar{l}_n - \bar{l}_a\right) + \left(\bar{\boldsymbol{X}}_n - \bar{\boldsymbol{X}}_a\right)'\boldsymbol{b} + \left(\bar{\eta}_n - \bar{\eta}_a\right) \tag{2.3}$$

It is evident that any differences in unobserved components of worker human capital will be absorbed into the residual wage gap here, and an OLS estimate of β will be biased.

There are two immediate approaches for obtaining better estimates of β. First, one can obtain a richer set of observable characteristics $\boldsymbol{X}_i$, reducing the scope for unobserved (to the econometrician) ability to determine income. Second, one can utilize panel data and estimate *within person* wage differences over time to purge the estimation of the time-invariant components of unobserved individual characteristics. While our estimation explores both avenues, our preferred estimates use the second approach, using fixed effects panel data estimation.

In a dynamic setting, the Mincerian human capital equation changes slightly to become: $H_{it} = \exp\left[\boldsymbol{X}_i'\boldsymbol{b} + \eta_i + \omega_{it}\right]$. Here, η_i is again unobserved individual skill, and ω_{it} is a mean zero, individual, time-varying shock. An individual's time-invariant human capital (which we estimate below as an individual fixed effect) is thus $\theta_i = \boldsymbol{X}_i'\boldsymbol{b} + \eta_i$. Equation 2.2 becomes:

$$y_{it} = z_a + \beta D_{it} + l_{it} + \theta_i + \omega_{it} \tag{2.4}$$

and the analogue of equation 2.3 is:

$$\bar{y}_{nt} - \bar{y}_{at} = \beta + \left(\bar{l}_{nt} - \bar{l}_{at}\right) + \left(\bar{\theta}_n - \bar{\theta}_a\right) + \left(\bar{\omega}_{nt} - \bar{\omega}_{at}\right) \tag{2.5}$$

Here, the time-varying components of human capital ω_{it} are potential sources of omitted variable bias. Equation 2.4 is the key estimation equation; we explore potential limitations and pitfalls to this approach in what follows.

Econometric Issues Related to Worker Selection and Heterogeneous Sectoral Effects

In the previous section, we assumed that sectors had a constant effect regardless of individual. In the framework, the dominant issue facing interpreting cross-sectional regression adjusted comparisons as causal was absolute advantage: omitted unobservable characteristics of human capital that confer higher income and productivity to individuals regardless of their sectoral choice. Utilizing fixed effects regressions and focusing on the experiences of switchers addresses many of those concerns.

However, in the presence of effect heterogeneity stemming from individual comparative advantages, using the sample of switchers to infer population average effects can yield biased estimates. In this section, we relax the constant effects assumption in a Roy (1951) model of self-selection and explore issues related to inferring a population average effect of sector from the subset of individuals who actually switch.

We first specify human capital to be sector-specific and modify our Mincerian human capital equation to be $H_{ist} = \exp\left[\theta_{is} + \omega_{ist}\right]$. This human capital formulation generates individual and sector-specific wages—and correspondingly, incentives—to choose sectors. Modeling sectoral choice formally, we specify workers' utility for sector s is given by $v_{ist}\left(y_{ist}\right)$, which is a function of the income that they would enjoy were they to work in that sector y_{ist}. Workers choose to work in non-agriculture if their utility for working in the sector is higher $v_{int} - v_{iat} \geq 0$.

To explore self-selection bias, we further assume that net utility can be written non-parametrically as $v_{int}\left(y_{int}\right) - v_{iat}\left(y_{iat}\right) = \left(y_{int} - y_{iat}\right) + \xi_{\text{it}}$, where ξ_{it} is an idiosyncratic preference for working in non-agriculture meant to capture other benefits (e.g. taste for urban non-agricultural work) and costs (e.g. pecuniary migration costs, non-pecuniary influences of past migration patterns, etc.) that are not capitalized into productivity. We do not impose further restrictions on ξ_{it}. For instance, if there are substantial switching costs or experience effects, then ξ_{it} is serially correlated. If there are compensating differentials, then ξ_{it} is not mean zero. ξ_{it} may also be arbitrarily correlated with worker characteristics.

Substituting in equation 2.4 for the individual specific productivity term, the probability of an individual choosing a particular sector (abstracting away from labor supply differences across sectors for parsimony) is given by:

$$\Pr\left\{v_{int} - v_{iat} > 0\right\} = \Pr\left\{\beta + \left(\theta_{in} - \theta_{ia}\right) + \left(\omega_{int} - \omega_{iat}\right) + \xi_{\text{it}} > 0\right\} \tag{2.6}$$

The possible bias here is classic simultaneity: we observe workers in the sector that benefits them the most. Unobserved wage innovations driven by H_{ist} simultaneously affect the worker's choice and her wage.

In a richer formulation of human capital with comparative advantage, the modified aggregate productivity gap in equation 2.5 (again abstracting away from labor supply differences) is:

$$\begin{aligned}\bar{y}_{nt} - \bar{y}_{at} &= \beta + \boldsymbol{E}\left\{H_{int} | v_{int} > v_{iat}\right\} - \boldsymbol{E}\left\{H_{iat} | v_{iat} - v_{int}\right\} \\ &= \beta + \boldsymbol{E}\left\{H_{int} - H_{iat} | v_{int} > v_{iat}\right\} + \left(\boldsymbol{E}\left\{H_{iat} | v_{int} > v_{iat}\right\} - \boldsymbol{E}\left\{H_{iat} | v_{iat} > v_{int}\right\}\right)\end{aligned} \tag{2.7}$$

The first bias term captures OLS's bias due to heterogeneous effects. The second bias term captures differences in agricultural productivity driven by absolute advantage, the focal point of the previous section.

Consider workers who are born in rural areas. Even if migrants and non-migrants have the same productivity in agricultural work, if workers with a comparative advantage (i.e. wage benefits) are the ones who tend to migrate—i.e. $\boldsymbol{Cov}\left\{v_{int} - v_{iat}, y_{int} - y_{iat}\right\} > 0$—then OLS will continue to be biased upward because we observe urban wages only for those who benefit. This is analogous to generalizing an estimate of treatment effect on the treated to an average treatment effect in the program evaluation literature.

This argument raises several considerations. First, for OLS to instead generate a negative bias, a necessary condition is that $\boldsymbol{Cov}\left\{v_{int} - v_{iat}, y_{int} - y_{iat}\right\} < 0$, which is permitted by virtue of our unrestrictive formulation of non-wage utility ξ_{it}. However, this immediately implies that $\boldsymbol{Cov}\left\{\xi_{it}, y_{int} - y_{iat}\right\} \leq -1$. Note that this does not just imply that those who have the most to gain from sectoral switches face the largest costs (i.e. $\boldsymbol{Cov}\left\{\xi_{it}, y_{int} - y_{iat}\right\} \leq 0$). This suggests that costs associated with switching sector—whether pecuniary or psychological—grow on average *faster* as benefits increase. This also precludes consumption complementarities considered in the literature where for example college educated workers are both relatively more productive in cities and have greater tastes for the amenities (i.e. $\boldsymbol{Cov}\left\{\xi_{it}, y_{int} - y_{iat}\right\} > 0$). Our results show that migrants have more human capital along all dimensions observable to us, but we nonetheless acknowledge the possibility however remote that non-migrants may have untapped potential for productivity in urban environments.

This is certainly not outside the realm of possibility, though it does beg the question of how extreme policies need to be to achieve reallocation and how welfare improving such policies can be. However, if non-wage preferences for work are positively correlated or even mean independent of benefits, then OLS estimates will give upward biased estimates.

Similar arguments carry over to the fixed effects estimates except for the fact that they are identified on the subset of workers for whom we observe a wage in both sectors. Owing to the long-term nature of our datasets, we are able to observe many individuals over the course of their entire work histories, but our estimates would preclude for example individuals who migrated prior to entering the labor market. Continuing to acknowledge that under certain cost structures the biases may be reversed, this suggests that fixed effects estimates would be downwardly biased estimates for those early movers but continue to be upwardly biased for the non-mover majority and thus the population as a whole.[6] We ask whether this latter population has trapped potential for large productivity gains that are encumbered by frictions.

Finally, much of the previous discussion has focused on workers born in rural areas, which is the case in our Kenya data where the entire population lived in rural areas at baseline. However, the IFLS also features individuals born in urban areas and sorting in both directions. By a parallel logic to above, it is conceptually possible to observe a non-agricultural (or urban) premium every time an individual selects into non-agriculture (urban areas) and an agricultural (rural) premium every time an individual selects into agriculture (rural areas).[7] The resulting estimates would then serve as *lower bounds* on the true average productivity gain.

The IFLS provides an unusually rich testbed to understand the role of these biases, especially in terms of estimating the urban-rural gap. In the spirit of Young's (2013) observation that migrants flow in both directions, the data allow us to condition on individual birth

[6] Hendricks and Schoellman (2017) make the related point that their estimates of the returns to international migration are likely to be upper bounds.

[7] Formally, individuals born in rural areas have a smaller ξ_{it} than those born in urban areas.

location and measure the dynamic impacts on wages after migration. The bounding argument above predicts that the estimated urban-rural productivity gap would be larger when estimated for movers from rural to urban areas than it is when estimated for movers from urban to rural areas. We take this prediction to the data and find evidence for it below. This model of selection implies that the true sectoral productivity gap in Indonesia is bounded by these two estimates, generated by movers in each direction.

2.3 Data

This paper uses data from Indonesia (IFLS) and Kenya (KLPS). At 250 million, the Southeast Asian country of Indonesia is the world's fourth most populous, and Kenya is among the most populous Sub-Saharan African countries with 45 million inhabitants. They are fairly typical of other low income countries with respect to their labor shares in agriculture, estimated agricultural productivity gaps using national accounts data, and the relationships between these variables and national income levels.[8]

The high tracking rates of the datasets we employ allow us to construct multiyear panels of individuals' location decisions. Moreover, both datasets include information on both formal and informal sector employment. The latter is difficult to capture in standard administrative data sources yet often accounts for a large share of the labor force in low-income countries.

Indonesia

Data were collected in five rounds of the Indonesia Family Life Survey between 1993 and 2015 (Strauss, Witoelar and Sikoki, 2016). The survey is representative of 83% of the country's population who lived in 13 of the 27 provinces that existed in 1993. While the original sample consisted of 22,347 individuals, efforts to track them even when they had moved outside of the original study area, as well as the inclusion of members from split-off households during subsequent rounds (1997-98, 2000, 2007-08, and 2014-2015), ultimately results in a sample of 58,337 individuals. Attrition is often high in panel data; however, with an intensive focus on respondent tracking, the IFLS is unusually well-suited to study migration. In particular, re-contact rates between any two rounds are above 90%, and 87% of the original households were contacted in all five rounds (Strauss et al. 2016).[9]

Detailed employment data were collected during each survey round. In addition to current employment, the survey included questions on previous employment, allowing us to create up to a 28-year annual individual employment panel from 1988 to 2015. Employment status and sector of employment are available for each year, but in the fourth and fifth IFLS round, earnings were collected only for the current job. The IFLS includes information on the respondent's principal as well as secondary employment. Respondents are asked to include

[8]See Appendix Figures A.1 and A.2 based on data from GLW. In both figures, the values for both Kenya (KEN) and Indonesia (IDN) are close to the best fitting regression line.

[9]Thomas et al. (2012) contains a detailed discussion of tracking and attrition in the IFLS.

any type of employment, including wage employment, self-employment, temporary work, work on a family-owned farm or non-farm business, and unpaid family work. In addition to wages and profits, individuals are asked to estimate the value of compensation in terms of share of harvest, meals provided, transportation allowance, housing and medical benefits, and credit; our main earnings measure is the sum of all wages, profits, and benefits. [10]

Individuals are asked to describe the sector of employment for each job. The single largest sector is "agriculture, forestry, fishing, and hunting": 31% of individuals report it as their primary employment sector, and 50% have secondary jobs in this sector. Agricultural employment is primarily rural: 43% versus 9% of rural and urban individuals, respectively, report working primarily in agriculture (Table 2.1a). Other common sectors are wholesale, retail, restaurants, and hotels (22% of main employment); social services (22%); manufacturing (14%); and construction (5%). These non-agricultural sectors are all more common in urban areas. Men are more likely than women to work in agriculture (35 vs. 23%) and less likely to work in wholesale, retail, restaurants, and hotels, in and social services. Smaller male-dominated sectors include construction (7% of male employment vs. 0.7% for females) and transportation, storage, and communications (6% vs. 0.4%).

In the analysis, we employ an indicator variable for non-agricultural employment, which equals 1 if a respondent's main employment is not in agriculture and 0 if main employment is in agriculture. The main analysis sample includes all individuals who are employed and have positive earnings and positive hours worked to ensure that the main variable of interest, the log wage, is defined. The sample includes 31,843 individuals and 275,600 individual-year observations.[11]

In addition to studying earnings, we explore consumption to get a broader sense of overall welfare and total income. IFLS consumption data were collected by directly asking households the value in Indonesian Rupiah of all food and non-food purchases and consumption in the last month, similar to consumption data collection in the World Bank's Living Standards Measurement Surveys.[12] In contrast to the retrospective earnings data in the IFLS, the consumption data are all contemporaneous to the survey. Consumption data were collected at the household level, which we divide by the number of household members to obtain a per capita measure, and are presented in real terms, taking into account prices in rural and urban areas. The consumption sample includes 82,272 individual-year observations from 34,820 individuals in IFLS rounds 1–5. In the consumption analysis, we expand the sample to also include individuals without current earnings data; we also perform a robustness check using the main productivity sample.

Data were collected on the respondent's location at the time of the survey, and all rounds of the IFLS also collected a full history of migration within Indonesia. All residential moves

[10]De Mel, McKenzie, and Woodruff (2009) argue that self-reported profits give a more accurate depiction of firm profits in microenterprises than reconstructed measures.

[11]The panel is unbalanced due to attrition, death, and to limiting observations to respondents at least 16 years old.

[12]Note that for a small number of frequently-consumed items, information was collected for the last week, and for a few low-frequency items, data was collected for the last year.

across sub-districts (“kecamatan”) that lasted at least six months are included, i.e., seasonal migration is excluded. Figure 2.2a presents a map of Indonesia with each dot representing an IFLS respondent’s residential location. While many respondents live on Java, we observe considerable geographic coverage throughout the country. The IFLS also asked respondents for the main motivation of each move. Family-related reasons are most common at 50%, especially for women (53%), who are more likely than men to state they migrated for marriage. The second most common reason to migrate is for work (32%), with little difference by gender, while migrating for education is less common. We combine data across IFLS rounds to construct a 28-year panel, from 1988 to 2015 with annual information on the person’s location, in line with the employment panel; refer to Kleemans (2016) and Kleemans and Magruder (2017) for more information on the IFLS employment and migration panel.

We utilize a survey-based measure of urban residence: if the respondent reports living in a “village”, we define the area to be rural, while they are considered urban if they answer “town” or “city.” We present the correspondence between urban residence and employment in the non-agricultural sector in Table 2.1a. In 69% of individual-year observations, people are employed in the non-agricultural sector, and in 35% of the observations, they live in urban areas. One can see that a substantial portion of rural employment is in both agriculture and non-agricultural work, while urban employment is almost exclusively non-agricultural, as expected.

Given the migration focus of the analysis, it is useful to report descriptive statistics both for the main analysis sample, as well as separately for individuals in four mutually exclusive categories (Table 2.2a): those who always reside in rural areas throughout the IFLS sample period (“Always Rural”), those who were born in a rural area but move to an urban area at some point (“Rural-to-Urban Migrants”), those who are “Always Urban,” and finally, the “Urban-to-Rural Migrants” (born urban but move to a rural area at some point). As discussed above, the fixed effects analysis is driven by individuals who move between sectors during the sample period.

In the main IFLS analysis sample, 87% of adults had completed at least primary education, and more than a third had completed secondary education, while tertiary education remain quite limited, at 11%. Among those who are born in rural areas in columns 2 and 3 (of Table 2.2a), we see that migrants to urban areas are highly positively selected in terms of both educational attainment, and in terms of cognitive ability, with Raven’s Progressive Matrices exam scores roughly 0.2 standard deviation units higher among those who migrate to urban areas, a meaningful effect.[13] Migration rates do not differ substantially by gender.

These relationships are presented in a regression framework in Table 2.3a (columns 1 to 5), and the analogous relationships for moves into non-agricultural employment are also evident (Table 2.4a). Importantly, the relationship between higher cognitive ability and likelihood of migrating to urban areas holds even conditional on schooling attainment and demographic characteristics (column 6 of both tables), at 99% confidence. This indicates

[13] Raven’s Matrices were administered to a subset of individuals in IFLS 3, 4 and 5, namely those 7 to 24 years old. The Raven’s Matrices test is designed to capture fluid intelligence.

that sorting on difficult-to-observe characteristics is relevant in understanding sectoral productivity differences.

It is worth noting that if we ignore migrants, individuals who are born and remain in urban areas are far more skilled than those who stay in rural areas. “Always Urban” individuals score over 0.3 standard deviation units higher on Raven’s matrices and have almost triple the rate of secondary schooling and four times the rate of tertiary education relative to “Always Rural” individuals. The urban-to-rural migrants in Indonesia are also negatively selected relative to those who remain urban residents, consistent with Young’s (2013) claim. These patterns emerge in Table 2.2a, where the urban-to-rural migrants score lower on all skill dimensions relative to those who remain urban; Appendix Tables A.1 and A.2 report analogous results among those individuals born in urban areas.

Kenya

The Kenya Life Panel Survey (KLPS) includes information on 8,999 individuals who attended primary school in western Kenya in the late 1990s and early 2000s, following them through adolescence and into adulthood. These individuals are a representative subset of participants in two school-based randomized interventions: a scholarship program for upper primary school girls that took place in 2001 and 2002 (Kremer, Miguel, and Thornton 2009) and a deworming treatment program for primary school students during 1998–2002 (Miguel and Kremer 2004). In particular, the KLPS sample contains information on individuals enrolled in over 200 rural primary schools in Busia district at the time of these programs’ launch. According to the 1998 Kenya Demographic and Health Survey, 85% of children in Western Province aged 6–15 were enrolled in school at that time, and Lee et al. (2015) show that this area is quite representative of rural Kenya as a whole in terms of socioeconomic characteristics. To date, three rounds of the KLPS have been collected (2003–05, 2007–09, 2011–14).

KLPS data collection was designed with attention to minimizing bias related to survey attrition. Sample individuals who had left the original study area were tracked throughout Kenya (as well as into neighboring Uganda and beyond, although we exclude international migrants from the present analysis).[14] Respondents were sought in two separate “phases” of data collection: the “regular tracking phase” proceeded until over 60% of respondents had been surveyed, at which point a representative subset of approximately 25% of the remaining sample was chosen for the “intensive tracking phase” (and remaining unfound individuals no longer sought). These “intensive” individuals receive roughly four times as much weight in the analysis, to maintain representativeness with the original sample. The effective tracking rate for each KLPS round is roughly 85%.[15]

Similar to the IFLS, the KLPS includes information on educational attainment, labor market participation, and migration choices. Employment data was collected in wage em-

[14]The results presented below are robust to the inclusion of international migrants (not shown).

[15]Baird, Hamory and Miguel (2008) describe the motivation behind this methodology and calculate effective tracking rates.

ployment and self-employment modules, designed to capture both formal and informal employment. Most individuals were quite young (typically teenagers) during data collection for KLPS Round 1, and few had wage employment or self-employment to report. Full employment histories, including more detailed questions, were collected during Rounds 2 and 3, and it is from these rounds that we draw the data on individual earnings, hours worked, and wages used in the present analysis.

The Kenya agricultural productivity data deserves detailed discussion. Whenever total household annual agricultural sales were at least moderate, exceeding 40,000 Kenyan Shillings (approximately 400-500 US dollars), full agricultural production and profit information was collected in the self-employment module and included in the present analysis. Agricultural wage employment is also common, and these data are always included. Limited questions on subsistence agricultural production were collected in KLPS rounds 1 and 2, but these are insufficient to create an individual productivity measure; more detailed information on agricultural productivity (in the previous 12 months) is contained in round 3, and this is included in the present analysis. To create a measure of individual productivity comparable with other sectors, we focus on agricultural activities (e.g., growing a particular crop) in which the respondent provided all reported labor hours; we also restrict attention to activities in which the respondent reports being the main decision-maker, since it seems likely that they are most knowledgeable about such activities (although results are not sensitive to this restriction). The profit in an agricultural activity is the sum of all crop-specific production – valued either through actual sales or at the relevant crop price (collected in regular local market price surveys) if consumed directly – minus all input costs and hired labor costs. The individual wage divides this net profit by the labor hours the respondent supplied to the activity. Given possible measurement concerns, we show below that estimates are robust to alternative approaches to constructing individual agricultural productivity, including the exclusion of subsistence agriculture data.

KLPS respondents reported industry for all wage and self-employment. Most individuals are engaged in relatively low-skilled work. The most common industry for wage employment is services, at 57% overall and 74% for females (with many women employed in domestic services). In rural areas, the most common industries for wage employment are services and agriculture (50 and 21%, respectively), while in urban areas they are services, and manufacturing and construction (62 and 11%). The largest self-employment industries are retail and services (41 and 25%).[16]

KLPS round 3 collected detailed consumption expenditure data for a subset of individuals. However, because it was only collected for this round, we are unable to utilize it in panel estimation. Instead, in the panel analysis we utilize a proxy for consumption, the number

[16]For wage employment, respondents also report occupation, and these tell a similar story. The most common occupations fall in the "unskilled trades" category (32%), followed by "skilled & semi-skilled trades" (19%), "retail and commercial" (18%), "professionals" (16%), and "agriculture" (15%). Agricultural wage employment is more common for men (20%) than women (6%), and as expected, agricultural employment is far higher in rural than urban areas (29% vs. 5%). Common urban occupations are "unskilled trades" (37%), "skilled & semi-skilled trades" (22%) and "retail and commercial" (20%).

of meals eaten in the previous day, which is available in both KLPS rounds 2 and 3. Reassuringly, meals eaten is strongly correlated with our primary measures of labor productivity as well as consumption expenditures per capita (in KLPS-3); see Appendix Table A.3. As with Indonesia, in the meal consumption analysis, we are able to expand the sample to also include individuals without current earnings data.

KLPS respondents provide a history of residential locations since their last interview, and this data includes residential district, town, and village, allowing us to classify individuals who lived in towns and cities as urban residents. The KLPS includes information on all residential moves that lasted at least four months in duration, a slightly more permissive definition than in the IFLS, and we are able to construct a monthly residential panel from March 1998 to October 2014.[17] Combined with the retrospective labor productivity data, the main analysis sample is a monthly panel with 134,221 individual-month observations for 4,791 individuals.

Figure 2.2b presents a map of Kenya, with each dot representing a respondent residential location during 1998–2014. Most residences in western Kenya are in Busia district (where the sample respondents originally resided), with substantial migration to neighboring areas as well as to cities. Appendix Table A.4 presents the list of main towns and cities, and shows that 70% of urban residential moves are to Kenya's five largest cities, namely, Nairobi, Mombasa, Kisumu, Nakuru, and Eldoret. Men are slightly more likely than women to report migrating for employment reasons (60% of moves compared to 55% for females) while women are more likely to migrate for family reasons, including marriage (13% vs. 1% for men). A smaller share of moves (approximately 6%) are for education.

Summary statistics on employment sector and urban residence for KLPS respondents are presented in Tables 2.1b and 2.1c. Table 2.1b presents data for the main analysis sample; as described above, this contains subsistence agricultural information where available (from KLPS-3). The employment share in agriculture is much higher in rural areas (26.1%) than urban (5.4%), as expected, but the share in rural areas is somewhat lower than expected, likely because subsistence agricultural activities were not captured in earlier KLPS rounds. For a more complete portrait, Table 2.1c focuses on data from the 12 months prior to the KLPS-3 survey, which contains detailed information on subsistence agriculture, and here the agricultural employment share in rural areas is much higher.

Recall that the Kenya sample is all rural at baseline (they were originally attending rural schools). Similar patterns emerge regarding positive selection into urban migration, with educational attainment and normalized Raven's matrix scores both far higher among those who migrate to cities (Table 2.2b). In particular, there is a raw gap of nearly 0.3 standard deviation units in Raven's matrix scores between urban migrants and those who remain rural. Overall migration rates in Kenya are similar for females and males. Tables 2.3b and 2.4b report these patterns in terms of regression estimates, for urban migration and employment in non-agricultural work, respectively. As with Indonesia, controlling for

[17]Similar to the IFLS, the panel is unbalanced due to attrition, death, and inclusion of individuals 16 and older.

educational attainment and gender, the Raven's score is strongly positively correlated with urban migration (at 99% confidence).

2.4 Results

Main Agricultural and Urban Productivity Gap Estimates

GLW estimate raw and adjusted agricultural productivity gaps of 138 and 108 log points in Indonesia, respectively (Figure 2.1, Panel A). The estimate of this raw gap from the IFLS is somewhat smaller at 70 log points (Table 2.5a). The most straightforward explanation for this discrepancy is an issue of measurement. GLW observe that, in an analysis of 10 countries, the average agricultural productivity gap was 17 log points smaller when estimated in Living Standards Measurement Study (LSMS) data that is similar to the IFLS, and which is more likely to capture earnings in informal employment.[18] That said, the raw gap we estimate in the IFLS remains substantial.

Inclusion of control variables similar to those used by GLW to adjust macro data gaps reduces the estimated agricultural productivity gap in the IFLS to 57 and 35 log points (in Table 2.5a, columns 2 and 3). Estimating on the subsample for which we have scores from Raven's matrix tests, the gap is reduced slightly, although note the smaller sample size in this case.

Limiting the analysis to those who have productivity measurements at some point in both agricultural and non-agricultural employment, the productivity gap drops to 25 log points (col. 5), suggesting that selection on unobservable characteristics may be meaningful. Inclusion of fixed effects also reduces the gap (col. 6), and using our preferred labor productivity measure, the log wage (namely, the log of total earnings divided by hours worked), as the dependent variable nearly eliminates the gap: the coefficient estimate falls to 0.078 (standard error 0.021) in column 7, and further to 0.076 when considering the real log wage (adjusting for higher urban prices, col. 8).

We follow a similar approach for Kenya, where the raw agricultural productivity gap falls from 78 log points to 55 with the inclusion of controls (Table 2.5b, columns 1–4), and to 28 log points when including an individual fixed effect. Using the preferred hourly wage measure reduces the gap to 6.1 log points (col. 7), it falls further when adjusted with an urban price deflator (col. 8), and neither fixed effects wage estimate is significant at traditional levels of confidence.

Comparing column 1 (the raw gap) to column 7 (the preferred fixed effects estimate) in Table 2.5, the agricultural productivity gap is reduced by 88% in Indonesia and by 92% in Kenya. The standard errors are somewhat larger for Kenya, so the upper end of the 95% confidence interval includes a sizable gap of 37 log points, consistent with some non-trivial productivity gains to non-agricultural employment. That said, even this value remains far

[18]This comes from log transformed values from the "Average" row of GLW, Table 4, i.e. $\ln 2.6 - \ln 2.2 = 0.167$.

lower than the 108 and 71 log point effects that GLW estimate for Indonesia and Kenya, respectively, once they condition on observable labor characteristics (namely, hours worked and educational attainment). As noted in the introduction, these results for Indonesia and Kenya are presented graphically in Figure 2.1, Panels A and B and compared to GLW's estimated productivity gaps.[19]

Table 2.6 presents the closely related exercise of estimating the labor productivity gap between residents of urban and rural areas. While the existing empirical literature has sometimes conflated these two gaps, Table 2.1 shows that employment in rural areas is not exclusively characterized by agriculture. To the extent that residential migration is costlier than shifting jobs (but not homes), and the urban and non-agricultural wage premia are related but distinct parameters, one might suspect that an urban wage premium might even be more pronounced than the non-agricultural wage premium.

The microdata estimates from Indonesia and Kenya appear to be consistent with this view, at least at first glance: the raw gap reported in column 1 of Tables 2.6a and 2.6b are 54 and 86 log points for Indonesia and Kenya, respectively. Similar to the agricultural productivity gap, the urban-rural productivity gap falls when additional explanatory variables are added in columns 2, 3 and 4, but remains substantial and statistically significant. Focusing the analysis only on those who have earnings measures in both urban and rural areas (column 5) leads to a further reduction. Finally, the urban-rural earnings gap falls to 2.8 log points with the inclusion of individual fixed effects in Indonesia (column 6), and 0.2 log points for the preferred log wage measure (column 7). The analogous urban productivity effect estimate for Kenya is slightly larger at 16.5 log points (column 7). Thus, the productivity gap in Indonesia falls by 100% in Indonesia (to zero), and the reduction for Kenya is 81% (from 86.2 to 16.5 log points, across columns 1 and 7) with the inclusion of individual fixed effects. Once again, these results are summarized in Figure 2.1 (Panels C and D).[20] Urban productivity gaps in real wage terms (that account for higher urban prices) are further reduced in both countries (column 8).

The selection model (in Section 2.2) predicts that estimated productivity gaps would be higher among rural-to-urban migrants than for urban-to-rural migrants, given plausible patterns of selection bias. Table 2.7 explores this hypothesis in Indonesia by separately conditioning on birth location; Panel A contains those born rural and Panel B those born in urban areas. The same pattern of declining urban productivity gaps in each subsample is observed as additional controls are included (columns 1-3). In the preferred log wage specification in column 3, productivity gaps are indeed somewhat larger for those born in rural areas, as predicted by the sorting model. The estimated productivity gain to urban

[19]Similar patterns are obtained when using alternative definitions of non-agricultural employment, namely, classifying simultaneous work in both sectors as agriculture or as non-agriculture (Appendix Table A.5), a point we return to below.

[20]Appendix Table A.6 further explores the relationship between these two gaps by conditioning on observations in rural areas. The raw agricultural productivity gap and subsequent decline with the inclusion of controls and fixed effects is quite similar in Indonesia in rural areas (panel A), and even more pronounced in Kenya.

employment is 3.7 log points for those born in rural areas (Panel A) and -2.4 log points for those born urban (Panel B). The difference between estimates for those born in rural versus urban areas is significant (p-value=0.03), but is relatively small, suggesting tight bounds around zero.

There are a number of alternative measures of individual agricultural productivity that are worth considering to assess robustness of the main results. Appendix Figure A.3 illustrates how each source of agricultural productivity data in both the IFLS and KLPS contributes to the overall sample, and classifies measures into those that are more reliably measured (e.g., hourly wage work), and those that are less reliably measured (e.g., measures based on production in subsistence agriculture). We next assess robustness to different definitions of employment in agriculture, including if the majority of hours are in the sector (our main measure), as well as measures that classify an individual as working agriculture is any hours are in the sector, or alternatively if all hours are in the sector. We additionally explore robustness to the use of both wage earnings and self-employed profits in agriculture (main measure), versus measures that use only one or the other. For both Indonesia (Table 2.8a) and Kenya (Table 2.8b), estimated agricultural productivity gaps remain small and positive across five alternative measures, ranging from 1 to 12 log points in Indonesia and 0 to 16 log points in Kenya.[21]

Productivity versus Living Standards

The discussion above establishes at least an 80% reduction in estimated sectoral productivity gaps once individual fixed effects are included in the analysis (Figure 2.1). The wage measures presented thus far are closely related to the labor productivity parameters that are the focus of most existing macroeconomic empirical literature. However, productivity and "utility" may diverge for many reasons, including price differences across regions, amenities, unemployment, and other factors. For instance, there could be considerable individual heterogeneity in the taste for rural versus urban amenities, e.g., comforts of home, ethnic homogeneity, better informal insurance, etc., in rural areas versus cosmopolitan cities' better public goods and more novelty (but downsides too, such as crime). Moving itself may also impose large utility costs (Kleemans 2016).

Although it is impossible to fully capture these factors and convincingly measure individual welfare, to get somewhat closer to differences in living standards, we draw on consumption data from the IFLS. As described in Section 2.3, five rounds of the IFLS included questions on the value of household consumption which can be converted to per capita

[21]Further details and robustness checks are contained in Appendix Tables A.7–A.10. In Appendix Table A.7, we present estimates for Indonesia on a sample of individuals who are at most 30 years old, for greater comparability with the Kenya sample, which consists of young adults; the estimates remain similar. Appendix Tables A.8 and A.9 report results separately for wage earnings and self-employment earnings, respectively, and generate similar results. Appendix Table A.10 reports results for Kenya including subsistence agriculture even when the respondent is not the main decision maker for an activity (Panel A), and excluding subsistence agriculture entirely (Panel B), and results are robust.

consumption. In the main specification, we include all individuals who have such consumption data, even if they lack earnings measures. Consumption expenditures may also more accurately capture total household income in low income settings like ours with extensive subsistence agriculture, home production, and informal employment, all of which are challenging to measure, making it an attractive alternative to earnings data. The measure should also capture variation in total earnings caused by unemployment or job rationing.

The initial consumption gap between non-agriculture and agriculture is large and similar the productivity gap at 64 log points (Table 2.9a). The gap falls considerably when including time fixed effects and control variables in column 2, and falls to only 12.6 log points when also including individual fixed effects in column 3. A similar pattern is presented for the urban-rural consumption gap in columns 4, 5, and 6: the gap declines from 40 log points to 2.6 log points.[22]

We next explore the estimated urban consumption premium for those born in rural versus urban areas (Table 2.7, columns 4–6). In the preferred specification with individual fixed effects (col. 6), the urban consumption premium is larger for those born in rural areas (13.3 log points, Panel A), than those born in urban areas (-4.7 log points, Panel B), and the difference is highly significant (p-value$<$0.001). As with the earnings results, this is consistent with the predictions of the selection model (in Section 2.2) and suggests the urban premium is bounded rather tightly around zero.

The consumption proxy measure in the KLPS tells a similar story. The raw gap in meals eaten in Kenya between those in non-agriculture versus agricultural employment is positive and statistically significant, though smaller than the earnings gap (Table 2.9b); differences in magnitude are difficult to interpret given the different nature of the meals measure, and the possibility that it changes most at very low levels of income. Mirroring the broad pattern observed for labor productivity, this gap falls by almost half when including controls, and is actually slightly negative when including individual fixed effects (columns 1-3); a similar pattern holds for the urban-rural gap (columns 4-6).

Another dimension of welfare relates to patterns of unemployment. Appendix Table A.13 explores whether there are differences in unemployment rates and search behavior between urban and rural areas for Kenya, where this data is available. We find that unemployment (measured several ways) is either similar in urban and rural areas (Panel A, column 3) or somewhat higher in urban than in rural areas conditional on individual fixed effects (Panel A, column 6, and Panel B), strengthening the main finding that movers to urban areas may

[22] Appendix Table A.11 shows the gap in both food and non-food consumption (Panels A and B, respectively). The gaps in both components of consumption see reductions of 77-95% when including individual fixed effects. Appendix Table A.12 repeats the consumption analyses on the main analysis sample (i.e., those with earnings data) for total consumption (Panel A) and by food and non-food consumption (Panels B and C, respectively), and results are similar.

not experience large gains in total earnings.[23,24]

Sector-specific Productivity — Absolute and Comparative Advantage

In the conceptual framework, the richest model of human capital allowed for individual sector-specific productivity θ_{is}. Analysis of these productivities has been given renewed focus in Lagakos and Waugh (2013), who argue that self-selection on the basis of comparative advantage could play an important role. In their model, comparative advantage is positively correlated with absolute advantage, meaning that the most productive workers have the most to gain from selecting into non-agriculture.

Utilizing panel data, we estimate a modified version of equation 2.4 replacing the individual fixed effect with an individual-sector fixed effect.[25] We recover these estimates, and then normalize the mean of the fixed effects of permanent rural residents (non-movers) to be zero. Figure 2.3 presents the joint distributions of these estimated individual productivities by sector. Panel A includes Indonesians born in rural areas. It is apparent that rural-to-urban migrants are positively selected relative to non-migrants, with an average rural wage approximately 20 log points higher than non-migrants. These individuals experience only a 5 log point average increase in their wage upon migration to an urban area. Panel B presents the same exercise with Indonesians born in urban areas. Here, there appears to be negative selection into rural migration, with the average mover having approximately 20 log points lower wages when still in urban areas, and an increase of less than 1 log points in rural wages among movers. Panel C presents results in Kenya (all of whom were rural residents as children) that are analogous to panel A. Compared to Indonesia, there appears to be even more positive selection among urban migrants in Kenya (at 41 log points) as well as a moderate positive urban premium of roughly 16 log points, which is nearly identical to the regression adjusted estimate presented above.

Note that the realizations of roughly half of migrants fall below the 45 degree line in the three panels of Figure 2.3, which taken literally means that they experience higher earnings

[23] In this paper, we consider *mean differences* in productivity or consumption across sectors, but *variability of outcomes* could also be a determinant of individual wellbeing, as well as of migration choices (Munshi and Rosenzweig 2016). We test whether the variability of earnings in the agricultural (rural) sector is different than variability in the non-agricultural (urban) sector, conditional on individual fixed effects, and find mixed results. There are no statistically significant differences in variability across sectors in the Kenya sample. There is significantly more variability in agricultural (rural) wages and earnings in Indonesia relative to the non-agricultural sector (urban areas), although no significant differences in consumption variability (not shown). We leave additional exploration of these issues for future research.

[24] Results are unchanged when using alternative approaches to accounting for clustering (Bell and McCaffrey 2002, Cameron and Miller (2015), and Young (2016); see appendix Table A.14.

[25] This procedure is similar in spirit to the correlated random coefficient models utilized to analyze heterogeneous returns to hybrid seed adoption (Suri 2011) and labor unions' effects on wages (Card 1996, Lemieux 1998), although our approach makes fewer assumptions and is meant to be more descriptive.

in rural than urban areas. This is consistent with the empirical finding of zero or small positive sectoral productivity gaps.

This exercise is meant to be descriptive, and we interpret the relationships between the estimated individual urban and rural productivities with some caution, in part because the estimates are subject to measurement error and thus the fitted regression line may experience attenuation bias. With these caveats in mind, note that all three plots appear to show that absolute advantage plays a role in wage determination: individuals who have high rural productivity tend to have high urban productivity, and vice versa, indicated by the positive slope.

Dynamics of the Productivity Gap and Big City Effects

In unpacking the main result, we examine if dynamics and experience effects produce productivity gains that do not materialize right away. In particular, while the main specification includes time fixed effects which would account for overall growth of wages as the sample ages or year specific shocks, individuals may begin to earn more after spending time in urban areas. Figure 2.2 presents event study analyses of whether individuals earn more after migrating, where we estimate regressions of the form:

$$y_{it} = \theta_i + \delta_t + \boldsymbol{X}'_{it}\boldsymbol{b} + \sum_{\tau} \beta_{\tau(i,t)} D_{it} + \varepsilon_{it} \tag{2.8}$$

where τ indexes the number of periods relative to the first move to an urban area, and the β_τ's are the parameters of primary interest.

These regressions are estimated on an unbalanced panel of individual-time periods and include individual fixed effects θ_i, time fixed effects δ_t, and controls for squared age and dummy variables for time periods exceeding five years pre- and post-move. The Indonesian and Kenyan analyses both condition on individuals being born rural.

The β_τ parameters of primary interest are coefficients on indicators for time periods relative to the individuals' move to an urban area at $\tau = 0$. Estimates are relative to the year or month prior to the individuals' move in Indonesia and Kenya, respectively; we exclude an indicator for the period immediately prior to the individuals' move. These coefficients are identified by individuals who have adjacent productivity measures in both the period that they move to urban and the period immediately prior. We do not enforce a requirement that individuals are observed in every period five years prior- and post-move. If the extensive margin decision to exit the labor force entirely or attrit from the sample is correlated with urban labor market experiences, the results may be biased and we thus interpret them with caution. Nonetheless, the richness of the panel dataset is novel and worth exploring.

These parameters represent the difference in mean wages between movers and non-movers net of the difference that existed in the period prior to the urban move. This approach also allows us to assess wage dynamics *prior* to the move, which may give some clues about what precipitated the move – e.g., whether rural individuals are more likely to move following a negative earnings shock – and they also allow us to examine whether urban experience leads to gradually rising earnings there.

In Indonesia, urban wages do not change substantially relative to the year prior to moving, and even five years after the urban move, migrants see no average wage gain (Figure 2.2a). There are broadly similar results in Kenya relative to the month prior to the move; there is some suggestive indication of slightly rising wages in the first two years of residence in an urban area, but these are small (Panel B). There is no indication of meaningful pre-move trends in either country.

In this analysis, we consider wages for individuals who made an urban move regardless of whether they remained in cities or towns, or later moved back to rural areas. The bottom halves of both panels A and B show a "survival" rate in urban areas of between 50 to 60% after five years (in both countries), suggesting substantial return migration. Naturally, one might suspect that those with the worst economic outcomes in urban areas might return home, yet this does not appear to be the case: Appendix Figure A.4 separately plots post-move wages for those who remain in urban areas and those who return to rural areas, and we find no evidence of a significant divergence in earnings between these two groups (although note that this analysis has limited statistical power). This suggests a direction for future research in uncovering the reasons for these moves, including the role played by non-economic factors, including family reasons and heterogeneity in the taste for urban living.[26]

Other scholars have argued that job experience is particularly valuable in big cities and that residence in these cities may boost individual productivity over time (see de la Roca and Puga 2016 for the Spanish case). We examine this issue, first repeating the main urban productivity gap analysis (from Table 2.6) but including a breakdown into the five highest population cities in each country, in Table 2.10. In Indonesia, all five cities are larger than 2 million inhabitants, with the capital Jakarta at 10 million. Kenya's capital Nairobi has 3.4 million people, the second largest city (Mombasa) has nearly one million, while the other three cities in Kenya are smaller. The capitals are also the largest destinations for urban migrants in each country.

There is mixed evidence on the extent of big city productivity effects. There is no evidence for significantly larger effects in any of the largest cities in Indonesia, including Jakarta (column 4 of Table 2.10a). There is some evidence of significant positive urban productivity gains in the two largest Kenyan cities, Nairobi and Mombasa (Panel B). The total urban effect is moderate and statistically significant in the capital of Nairobi, at 23 log points.[27]

[26] We carry out an analogous event study of moves to rural areas among those born urban in Indonesia, and similarly find no evidence of significant dynamic impacts (see Appendix Figures A.5 and A.6).

[27] While this analysis finds mixed evidence of an overall big city effect in Indonesia and Kenya, we also assess whether effects might manifest over a longer time horizon by repeating the event study analysis over a five year time horizon separately for Jakarta and Nairobi. These figures show no clear evidence of differentially positive dynamic effects in capital cities: differences with other cities are imprecisely estimated and generally not significant (not shown).

2.5 Conclusion

Several influential recent studies document large sectoral productivity gaps in low-income countries and highlight an apparent puzzle, namely, "why so many workers remain in the agricultural sector, given the large residual productivity gaps with the rest of the economy" (Gollin, Lagakos, and Waugh 2014, p. 941). This study makes two main contributions using data from low-income countries with large populations (Indonesia and Kenya) located in two different regions. First, we show that estimating sectoral productivity gaps—both across non-agricultural and agricultural sectors, and across urban and rural areas—using panel data and including individual fixed effects leads to a reduction of over 80% in the estimated gaps. The second main empirical contribution lies in demonstrating that there is extensive individual selection across sectors, both along relatively easily observable dimensions such as educational attainment as well as measures of skill (here, a measure of cognitive ability) that most standard economic datasets lack.

Taken together, the findings point to the importance of individual selection in driving observed sectoral gaps and call into question strong causal interpretations. As a result, the puzzle of why the share of workers in rural agriculture remains high may not be as much of a puzzle as previously thought. Similarly, if gaps are mainly driven by selection, then policies to incentivize workers to move to urban areas (and out of agriculture), based on the logic of input misallocation, would not appreciably raise aggregate living standards and would not appear to be an appropriate policy direction.

An historical episode illustrates some of the potential risks of pro-urbanization policies. In the 1970s, Tanzania's authoritarian socialist government sought to move its rural population into larger villages and towns to speed up economic modernization. The underlying idea was that the provision of public services and the shift into non-agricultural work (including manufacturing) would be hastened if households would only leave their traditional homesteads, which were often highly spatially dispersed. After initial rhetorical encouragement and incentives by the government led to few moves, the government resorted to forced migration in certain regions in 1973, in the so-called "Operation Vijiji". The resulting economic and social dislocation is today widely viewed as a policy disaster within Tanzania (Stren, Halfani, and Malombe 1994). While one could argue that observers are unable to assess the true economic effects of the policy in Tanzania since the forced moves were quickly abandoned (within a year) in the face of large-scale resistance, at a minimum, the Tanzanian case indicates that it can sometimes be very costly from a welfare perspective to rapidly induce a large share of the population to move out of traditional rural agriculture.

As noted above, our main productivity gap estimates are derived from individual movers, namely, those with productivity measured in both sectors. Thus a logical way to reconcile our finding of small or even zero sectoral gaps with the existing macroeconomic empirical evidence of large average gaps is the possibility that productivity effects among non-movers would be much larger than those of movers. Given the nature of our data, it is impossible to rule out this possibility, and it clearly merits further investigation, although the lack of measured individual productivity in both sectors for non-movers naturally complicates the rigorous identification of these relationships.

However, several factors lean against this interpretation in our view, at least in the short-run. First, it is natural to think of the migration decision in terms of a Roy (1951) model, as we do above, in which those with the largest net utility benefits are most likely to move. This could lead our estimates to overstate gaps between sectors overall. While it is possible that those individuals who remain in the rural agricultural sector might receive large positive earnings gains from moving, their choice not to do so might simply reflect high financial or non-financial costs to migration. For instance, the amenities found in a large city are quite different than those in rural areas, and individuals may have strong and heterogeneous preferences for them, leading to large reductions in utility for some migrants even if wages rise. Poor individuals may also face credit constraints or other financial frictions that prevent them from exploiting wage gaps, and easing these constraints could boost migration rates, as argued for Indonesia by Kleemans (2016) and Bazzi (2017) and India by Munshi and Rosenzweig (2016). However, the long timeframes of both datasets used in this study help to, at least partially, mitigate this concern: some poor individuals with high returns to migration presumably had access to improved credit at *some* point during 1988-2015 in Indonesia or 1998-2014 in Kenya and managed to move.

A promising approach to estimating the returns to migration in low-income countries among those who are typically "non-movers" and may face such constraints is the Bryan et al. (2014) study in Bangladesh. They find that a modest subsidy induces 22 percent of households to send a migrant to towns and cities for temporary work during the agricultural low season; the relatively low rate of migration may indicate that the utility costs of migration are non-trivial. Among movers, there is an estimated increase in per capita consumption among the sending household of roughly 30% over two years and 25% average gain in earnings (not statistically significant) among those assigned to the subsidy. Overall, the study provides some indication that there are positive returns to temporary seasonal migration among rural workers who are typically non-movers. Nonetheless, the earnings gains are fairly modest in size and note that they are closer in magnitude to the small gaps we estimate in this paper than to those found in many other recent contributions. It is also worth noting that the subsidy was delivered during times of the year (the agricultural low season) when agricultural productivity was thought to be particularly meager and targeted to regions thought especially likely to benefit from seasonal moves, suggesting that the 30% consumption gain in Bryan et al (2014) might be an upper bound on the return to permanent urban migration in Bangladesh as a whole.

The case of urban-born non-movers is less well understood. Recall (from Table 2.2) that individuals raised in urban areas have much higher cognitive scores (on a test of fluid intelligence) than those raised in rural areas. This gap raises two intriguing possibilities. The first is that wave after wave of rural to urban (urban to rural) migration by positively (negatively) selected individuals over many decades, combined with partial heritability of cognitive ability, have reshaped the underlying ability distributions in these two sectors. This would simply be an inter-generational extension of the patterns of individual selection across urban and rural areas that we and Young (2013) document, and would not necessarily change the interpretation of our main results.

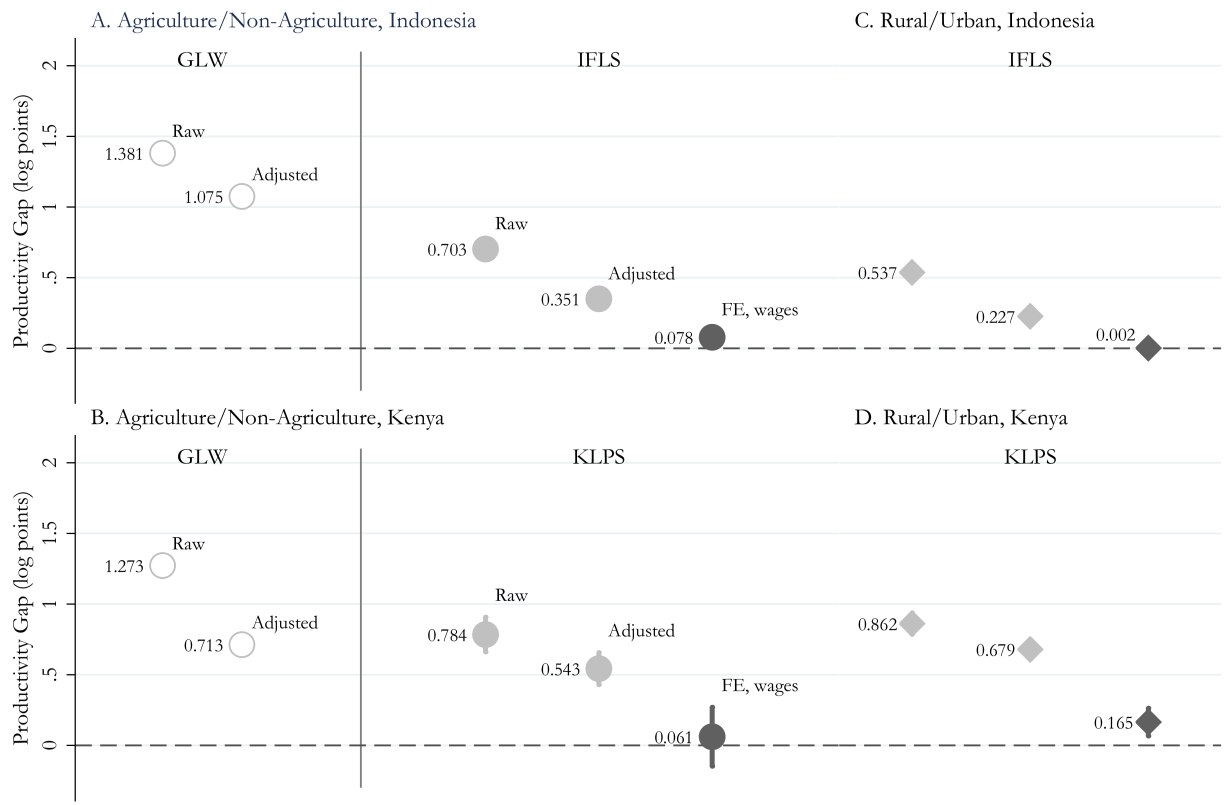
Another explanation, which is not mutually exclusive, is that there is a lower cost to skill acquisition in urban areas, either due to improved provision of schooling there or something about the nature of social interactions (e.g., the density of interactions or other forms of childhood intellectual stimulation). In other words, given the importance of early childhood circumstances for lifetime cognitive development (e.g., Gertler et al. 2014), growing up in a city might generate higher average adult skill levels. This would generate a positive causal effect of urban residence on labor productivity, albeit in the very long-run and on the movers' children; anecdotally, many migrants do claim to move in order to improve their children's wellbeing more than their own. These effects would not be captured in the five-year follow-up period that we consider in this study (in Figure 2.2), but could be contributing to large, persistent and real causal urban-rural productivity gaps overall. Indeed, Nakamura, Sigurdsson, and Steinsson (2017) study migration induced by a volcanic explosion in Iceland, and show that adult movers gain little from moving out of a rural area but their children earn far more in the long-run.

We tentatively assess this possibility using the IFLS data and a separate data set collected among young children of KLPS sample individuals, and find suggestive evidence that urban residence may have positive impacts on the cognitive ability of the next generation. In particular, young children born in urban areas have significantly higher cognitive scores than rural-born children, even conditional on the parent's own schooling and fluid intelligence (Table 2.11 and Appendix Figure A.7). The average gaps are meaningful, at approximately 0.15 standard deviation units in Indonesia (Table 2.11a, col. 5) and 0.25 in Kenya (Panel B, col. 5), although we admittedly cannot rule out the existence of some omitted variable bias in this more speculative cross-sectional comparison.

The study of sectoral productivity gaps remains an area ripe for further research, and some natural next steps include: extending long-run panel data analysis to new settings (as data becomes available); conducting more experiments that induce migration, thus generating "local" estimates in new sub-populations and improving understanding of the nature of constraints facing potential migrants; and further exploration of the inter-generational effects of sectoral and residential choice.

2.6 Figures

Figure 2.1: Productivity Gap in Total Earnings

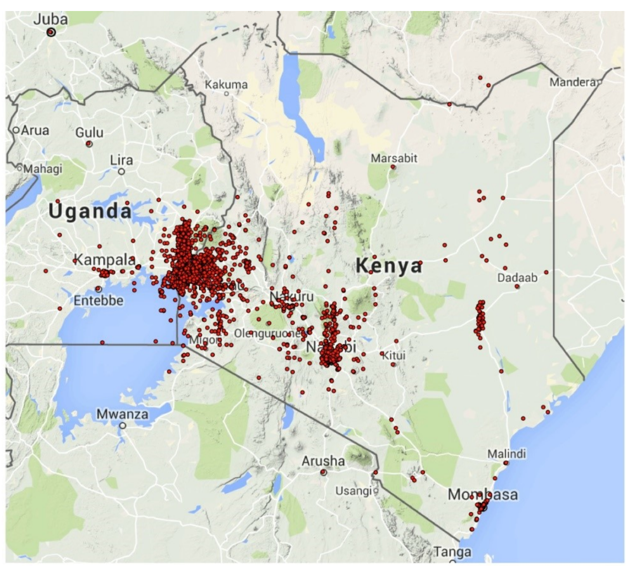


GLW refers to estimates from Gollin, Lagakos, and Waugh (2014), Online Appendix Table 4. For comparability, the figure reports log transformed numbers from their columns 4 and 5 for Indonesia and Kenya, respectively. Symbols here represent point estimates, and vertical lines represent 95% confidence intervals. Panel A estimates from the IFLS come from 2.5a: "Raw" is the mean difference estimate from column (1), "Adjusted" is the regression adjusted mean difference estimate from column (3), and "FE, wages" is the fixed effects regression estimate of wages on an urban indicator and squared-age from column (7). Corresponding estimates from the KLPS come from Table 2.5b. Estimates in panels C and D come from the same columns in Table 2.6, panels A and B, respectively. Note that the confidence intervals for the estimates from the IFLS are smaller than the size of the symbols and are therefore not visible.

Figure 2.2: Sample Areas

(a) Indonesia Family Life Survey

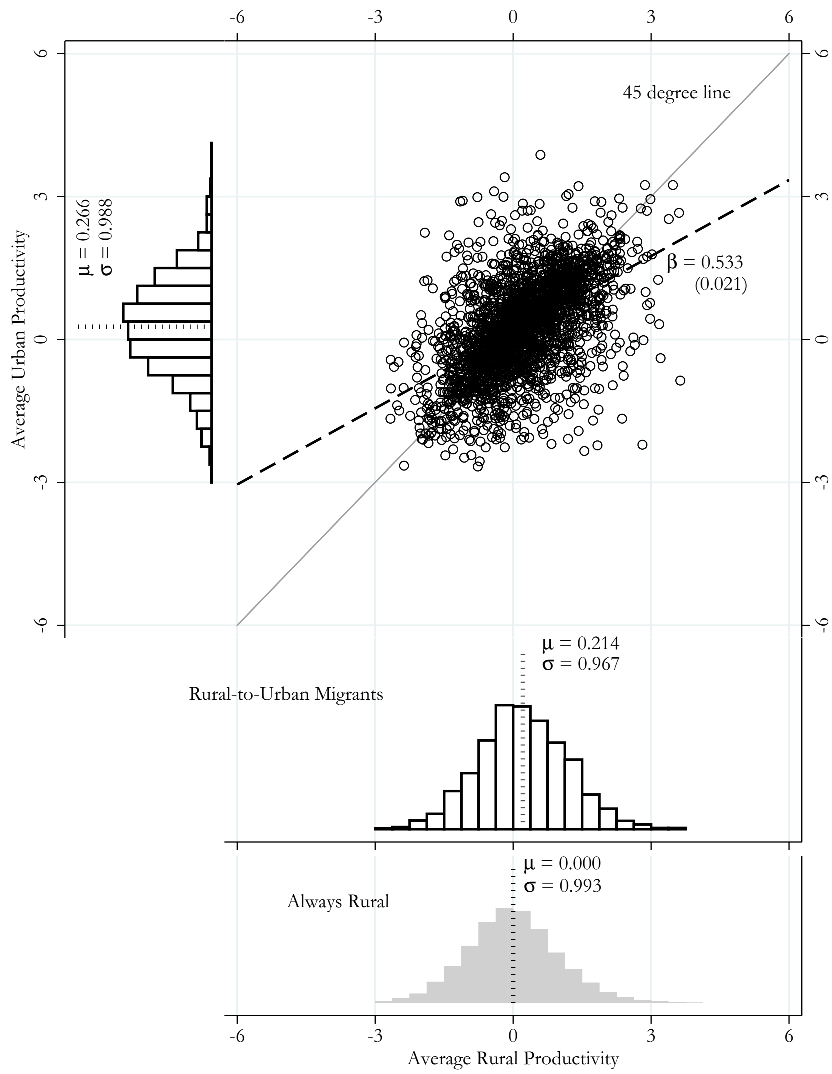
(b) Kenya Life Panel Survey



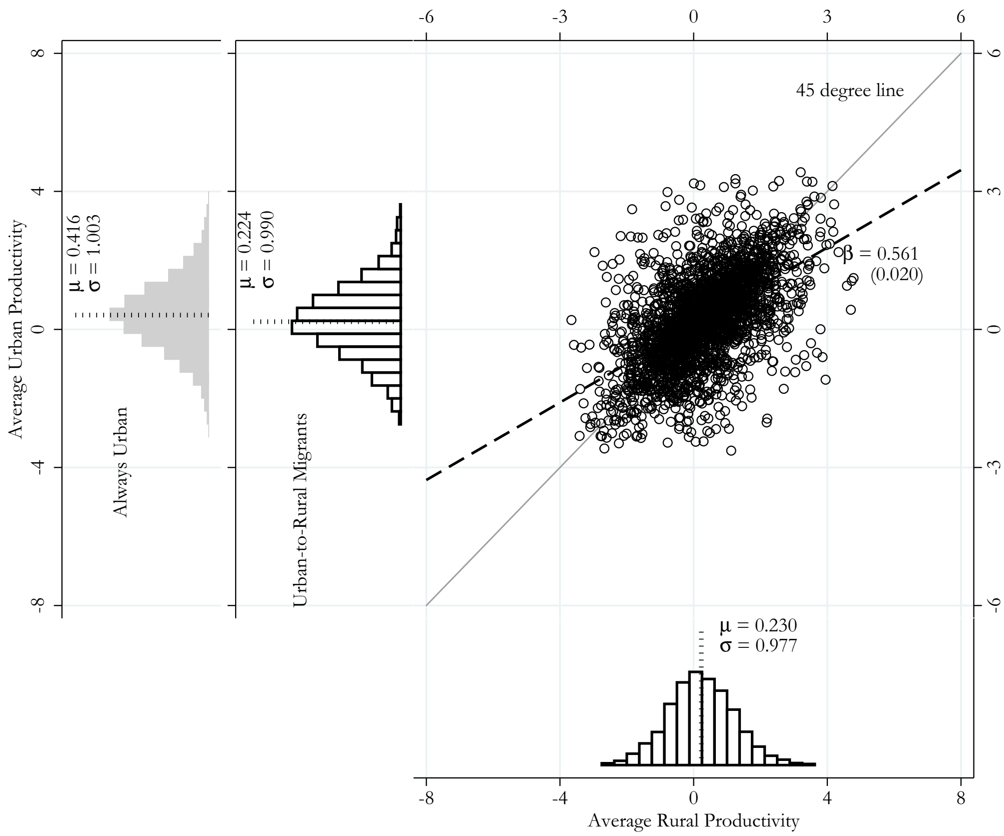
Panel A shows the residential locations of individuals during the 1988–2008 sample period of rounds 1–4 of the IFLS. For the Kenyan sample, Panel B shows individuals' residential locations during the 1998–2014 sample period that was collected during rounds 2 and 3 of the KLPS. Individuals living outside of Kenya are dropped from the analysis. The location information of both datasets are described in more detail in Section 2.3.

Figure 2.3: Joint Distribution of Rural and Urban Productivities

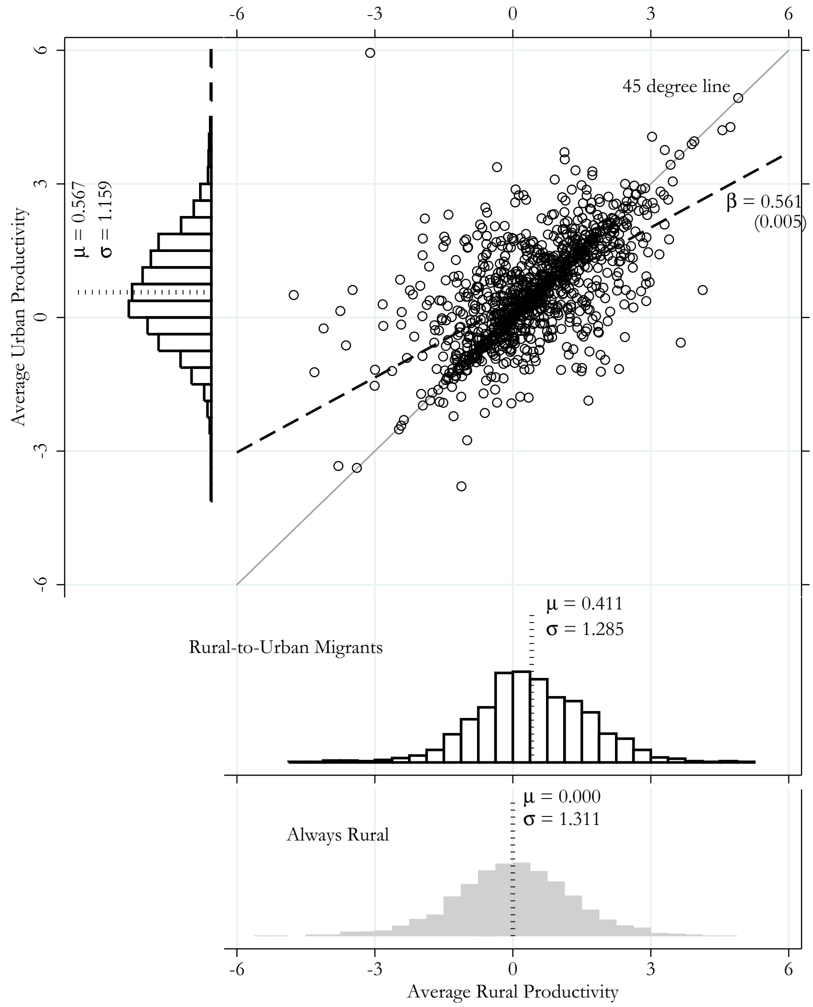
(a) Indonesia (Born Rural)



(b) Indonesia (Born Urban)



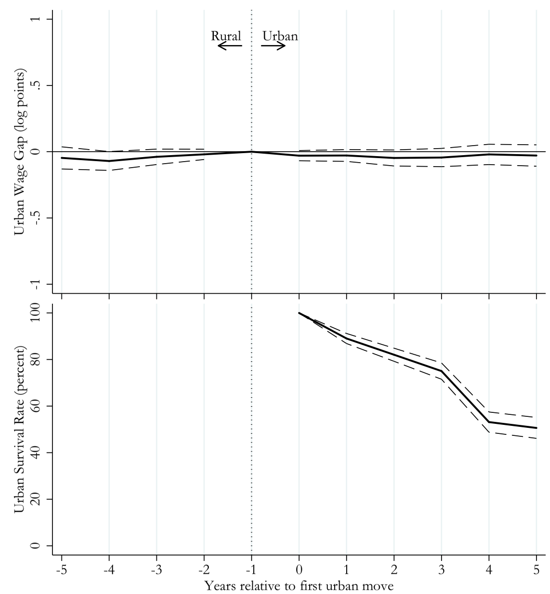
(c) Kenya (Born Rural)



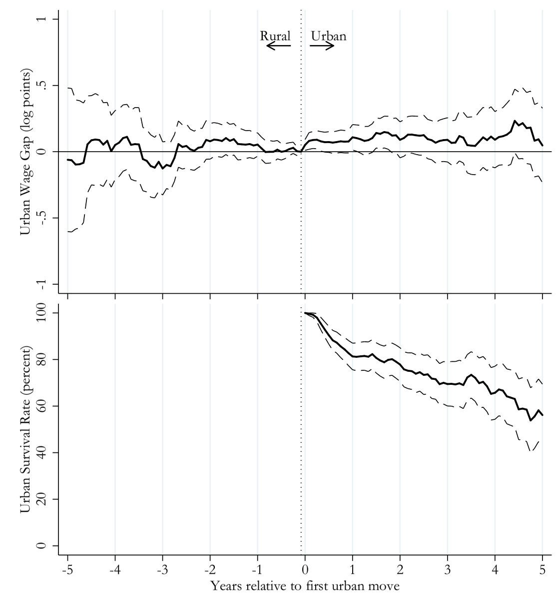
Productivities are recovered individual-urban status effects from a fixed effects regression of log wages on squared age and indicators for time period on the same sample used in Tables 2.5 and 2.6. Productivities are normalized such that the average productivity of rural non-migrants has zero mean. Histograms on the bottom of Panel A represent marginal distributions of rural productivities for "Always Rural" non-migrants (grey) and migrants (hollow). Marginal distribution of estimated urban productivities for migrants reported on the left (hollow). Means and standard deviations reported in log points. Scatterplot presents joint distribution for migrants with best fit line. Bootstrapped standard error of the slope reported in parentheses from 1,000 iterations of block sampling of individuals with replacement. Panel B presents a histogram of "Always Urban" urban productivities of non-migrants (grey) at the top left, an adjacent histogram of migrant urban productivities (hollow), and migrant rural productivities (grey) below. Joint distribution of urban and rural productivities and corresponding best fit line presented similar to panel A. Panel C mimics the format of Panel A except uses data from the KLPS.

Figure 2.2: Event Study of Urban Migration

(a) Indonesia (Born Rural)



(b) Kenya (Born Rural)



Panel A uses data from individuals in the IFLS who are born in rural areas, and Panel B uses data from the KLPS. Please refer to Section 2.3 for further details on the data. The top half of each panel reports event study coefficients β_τ from a regression of log wages described in equation 2.8 (in section 2.4). The solid line represents the point estimate, and the dashed lines represent the 95% confidence interval. Estimates represent the difference in mean wages between movers and non-movers net of the difference that existed in the period prior to the move. Regressions include individual fixed effects, time fixed effects, squared-age, and indicator variables that pool observations exceeding a five year window of the move. The lower half of each panel reports the fraction of people having no rural observations from period zero to the period of interest. (The estimated fraction of survivors can in principle increase due to sample composition changes as can be seen in the lower half of panel B.) In the IFLS, there are 923 individuals who have observed wages in the year of the move and the year prior; 482 of these individuals report wages 5 years later. In the KLPS, these numbers are 343 and 57, respectively.

2.7 Tables

Table 2.1: Non-Agriculture/Agriculture and Urban/Rural

(a) Indonesia (Main Analysis Sample)

	Rural	Urban	Total
Agriculture	42.6%	9.1%	30.9%
Non-Agriculture	57.4%	90.9%	69.1%
Number of Observations	179,756	95,844	275,600

(b) Kenya (Main Analysis Sample)

	Rural	Urban	Total
Agriculture	26.1%	5.4%	15.2%
Non-Agriculture	73.9%	94.6%	84.8%
Number of Observations	63,545	70,676	134,221

(c) Kenya (12 Months with Subsistence Agricultural Module

	Rural	Urban	Total
Agriculture	59.1%	9.1%	40.6%
Non-Agriculture	40.9%	90.9%	59.4%
Number of Observations	27,301	16,029	43,330

Panel A reports summary statistics from the Indonesia Family Life Survey (IFLS), and Panels B and C present data from the Kenya Life Panel Survey (KLPS); both are described in more detail in Section 2.3. Panel A shows the main Indonesian analysis sample of 275,600 individual-year observations, for individuals aged 16 and above for whom earnings measures are available. Panel B shows the main Kenyan analysis sample of 134,221 individual-month observations of individuals aged 16 and above for whom earnings measures are available. Panel C shows data from the 12 months where subsistence agriculture data is available and counts all agricultural activities: including when the person is not the main decision maker and when others work on the agricultural activity; in the case of the latter, the agricultural productivity is weighted by the share of hours that the individual supplies. Each cell reports the percentage of observations by agricultural and non-agricultural sector, and by rural and urban area. In both the IFLS and KLPS, individuals are characterized by the sector of their main employment. The urban indicator from the IFLS is obtained from survey responses to the question: “Is the area you live in a village, a town or a city?” If the person reports living in a town or city, the urban indicator variable equals 1. For the KLPS, the urban indicator equals 1 if the person reports living in a large town or city. Please see the text in section 2.3 for further details on this classification. The list of Kenyan urban areas and frequency of occurrence in the panel are given in Appendix Table A.4.

Table 2.2: Summary Statistics

(a) Indonesia

	All N=31843	Always Rural N=14737	Rural-to-Urban Migrants N=5287	Always Urban N=7594	Urban-to-Rural Migrants N=4218	Obs
Primary Ed.	0.865	0.791	0.939	0.957	0.862	31843
	[0.342]	[0.407]	[0.240]	[0.202]	[0.344]	
Secondary Ed.	0.393	0.255	0.452	0.614	0.400	31843
	[0.488]	[0.436]	[0.498]	[0.487]	[0.490]	
College	0.108	0.054	0.128	0.196	0.116	31843
	[0.311]	[0.226]	[0.334]	[0.397]	[0.321]	
Female	0.428	0.420	0.413	0.459	0.417	31843
	[0.495]	[0.494]	[0.492]	[0.498]	[0.493]	
Raven's Z-score	0.001	-0.143	0.082	0.185	0.081	22899
	[0.925]	[0.932]	[0.904]	[0.881]	[0.923]	

(b) Kenya

	All N=4791	Always Rural N=1639	Rural-to-Urban Migrants N=3152	Always Urban	Urban-to-Rural Migrants	Obs
Primary Ed.	0.734	0.637	0.785			4791
	[0.442]	[0.481]	[0.411]			
Secondary Ed.	0.353	0.240	0.412			4791
	[0.478]	[0.427]	[0.492]			
College	0.035	0.012	0.047			4791
	[0.184]	[0.107]	[0.212]			
Female	0.522	0.522	0.523			4791
	[0.500]	[0.500]	[0.500]			
Raven's Z-score	0.050	-0.143	0.149			4522
	[0.986]	[0.982]	[0.974]			

Panel A reports summary statistics from the IFLS and panel B reports summary statistics from the KLPS. Sample standard deviations reported in brackets below sample means. The sample is limited to respondents who report age, gender, and years of education and have at least one person-time observation that has income, hours, location of residence, and sector of occupation. In panel A (Indonesia), "Rural-to-Urban Migrants" are individuals born in rural areas and are observed in urban areas in our sample with data on wages, hours, and sector. "Urban-to-Rural Migrants" are defined similarly. In panel B (Kenya), all individuals are born rural; migrants are those who have subsequent observations with information on income, hours, and sector in urban areas. Rows correspond to the fraction within each column who have completed primary education, secondary education, and college; the fraction female; and the average score from a Raven's matrices exam, normalized to be mean zero and standard deviation one.

Table 2.3: Correlates of Urban Migration

(a) Indonesia (Born Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.203***					0.132***	0.153***
	(0.006)					(0.009)	(0.007)
Secondary Ed.		0.180***				0.108***	0.122***
		(0.007)				(0.009)	(0.008)
College			0.211***			0.0695***	0.0921***
			(0.013)			(0.016)	(0.015)
Female				-0.00565		0.0179*	0.0169**
				(0.006)		(0.007)	(0.006)
Raven's Z-score					0.0509***	0.0252***	
					(0.004)	(0.004)	
Constant	0.0954***	0.209***	0.249***	0.266***	0.271***	0.101***	0.0858***
	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)	(0.009)	(0.006)
Observations	20024	20024	20024	20024	14553	14553	20024

(b) Kenya (Born Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.170***					0.0902***	0.111***
	(0.016)					(0.019)	(0.018)
Secondary Ed.		0.169***				0.0969***	0.111***
		(0.014)				(0.017)	(0.016)
College			0.237***			0.121***	0.133***
			(0.025)			(0.027)	(0.027)
Female				0.000781		0.0168	0.0136
				(0.014)		(0.014)	(0.013)
Raven's Z-score					0.0674***	0.0308***	
					(0.007)	(0.008)	
Constant	0.533***	0.598***	0.650***	0.657***	0.657***	0.543***	0.525***
	(0.014)	(0.009)	(0.007)	(0.010)	(0.007)	(0.017)	(0.016)
Observations	4791	4791	4791	4791	4522	4522	4791

Please see Table 2 for sample restrictions and row variable definitions. Each cell reports a regression coefficient with an indicator for being an urban migrant as the dependent variable. Both panels are estimated on individuals who are born rural. Columns 6 and 7 report coefficients from multiple regressions with corresponding rows as included covariates. Column 7 omits the Raven's matrix exam to preserve sample size. Robust standard errors reported below in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.4: Correlates of Employment in Non-Agriculture

(a) Indonesia (Born Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.289***					0.208***	0.241***
	(0.009)					(0.013)	(0.010)
Secondary Ed.		0.226***				0.129***	0.157***
		(0.005)				(0.007)	(0.006)
College			0.225***			0.0481***	0.0622***
			(0.006)			(0.008)	(0.007)
Female				0.0533***		0.0801***	0.0872***
				(0.006)		(0.007)	(0.006)
Raven's Z-score					0.0633***	0.0348***	
					(0.004)	(0.004)	
Constant	0.512***	0.683***	0.736***	0.730***	0.790***	0.519***	0.462***
	(0.009)	(0.004)	(0.003)	(0.004)	(0.003)	(0.013)	(0.009)
Observations	20024	20024	20024	20024	14553	14553	20024

(b) Kenya (Born Rural)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.142***					0.106***	0.117***
	(0.012)					(0.014)	(0.013)
Secondary Ed.		0.101***				0.0419***	0.0523***
		(0.007)				(0.008)	(0.008)
College			0.0912***			0.0178*	0.0262***
			(0.007)			(0.008)	(0.008)
Female				0.0209*		0.0326***	0.0300***
				(0.008)		(0.009)	(0.008)
Raven's Z-score					0.0442***	0.0205***	
					(0.004)	(0.005)	
Constant	0.802***	0.870***	0.903***	0.895***	0.905***	0.794***	0.785***
	(0.011)	(0.006)	(0.004)	(0.006)	(0.004)	(0.013)	(0.012)
Observations	4791	4791	4791	4791	4522	4522	4791

Please see Table 2.2 for sample restrictions and row variable definitions. Each cell reports a regression coefficient with an indicator for being ever being employed in non-agriculture as the dependent variable. Panel A (Indonesia) is estimated on individuals who are born in rural areas, whereas panel B (Kenya) includes the full sample subject to previously defined sample restrictions. Please see Appendix Table A.1 for analogous regressions of individuals born urban in Indonesia. Columns 6 and 7 report coefficients from a multiple regression with corresponding rows as included covariates. Column 7 omits the Raven's matrix exam to preserve sample size. Robust standard errors reported below in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 2.5: Non-Agricultural/Agricultural Gap in Earnings

(a) Indonesia

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Log Wage	(8) Log Real Wage
Non-ag emp.	0.703***	0.572***	0.351***	0.349***	0.249***	0.252***	0.078***	0.076***
	(0.013)	(0.012)	(0.012)	(0.014)	(0.018)	(0.018)	(0.021)	(0.021)
Log hours		0.566***	0.437***	0.460***	0.279***	0.355***		
		(0.017)	(0.015)	(0.017)	(0.035)	(0.020)		
Log hours sq.		-0.022***	-0.011***	-0.014***	0.018***	-0.006		
		(0.003)	(0.003)	(0.003)	(0.007)	(0.004)		
Female			-0.468***	-0.465***	-0.530***			
			(0.011)	(0.012)	(0.028)			
Education			0.015***	-0.002	0.028***			
			(0.004)	(0.005)	(0.009)			
Education sq.			0.004***	0.005***	0.003***			
			(0.000)	(0.000)	(0.001)			
Ravens				0.065***				
				(0.007)				
Ravens sq.				0.014***				
				(0.005)				
Individual FE	N	N	N	N	N	Y	Y	Y
Time FE	N	Y	Y	Y	Y	Y	Y	Y
Switchers only					Y			
Obs	275600	275600	275600	201699	55802	275600	275600	275600
Individuals	31843	31843	31843	22899	4208	31843	31843	31843

(b) Kenya

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Log Wage	(8) Log Real Wage
Non-ag emp.	0.784***	0.548***	0.543***	0.559***	0.310***	0.283***	0.061	0.049
	(0.063)	(0.061)	(0.058)	(0.059)	(0.101)	(0.090)	(0.106)	(0.106)
Log hours		0.078	0.021	0.071	0.197	0.228		
		(0.196)	(0.172)	(0.174)	(0.377)	(0.258)		
Log hours sq.		0.041*	0.040**	0.035*	0.019	0.013		
		(0.021)	(0.019)	(0.019)	(0.045)	(0.028)		
Female			-0.562***	-0.533***	-0.708***			
			(0.038)	(0.039)	(0.137)			
Education			-0.012	-0.025	-0.041			
			(0.038)	(0.039)	(0.120)			
Education sq.			0.006***	0.006***	0.005			
			(0.002)	(0.002)	(0.007)			
Ravens				0.083***	0.034			
				(0.023)	(0.070)			
Ravens sq.				-0.046**	-0.153**			
				(0.020)	(0.070)			
Individual FE	N	N	N	N	N	Y	Y	Y
Time FE	N	Y	Y	Y	Y	Y	Y	Y
Switchers only					Y			
Obs	134221	134221	134221	128215	14922	134221	134221	134221
Individuals	4791	4791	4791	4522	341	4791	4791	4791

Panel A uses data from rounds 1–5 of the Indonesia Family Life Survey (IFLS), described in Section 2.3. Panel B uses data from rounds 2–3 of the Kenya Life Panel Survey (KLPS), also described in Section 2.3. The dependent variable in columns 1 to 6 is log earnings, which are the combined earnings from wage and self-employment, reported in Indonesian Rupiah. If an individual has multiple jobs in the same time period, earnings from all employment are included. The dependent variable in column 7 is log wage, which is obtained by dividing log earnings by total hours worked. The dependent variable in column 8 is log wage adjusted for differences in prices between urban and rural areas. The covariate “Non-agricultural employment” is an indicator variable which equals 1 if the person main employment is in the non-agricultural sector. The covariate log hours sums up hours worked in all employment. The sample size in column 4 is smaller in Panel A because the Raven's test was administered only for a subset of the sample. The sample size in column 5 is smaller because it only includes "switchers" who have at least one observation in both the non-agricultural and agricultural sector. Each regression in columns 2–8 include quadratic controls for age. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.6: Urban/Rural Gap in Earnings

(a) Indonesia

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Log Wage	(8) Log Real Wage
Urban	0.537***	0.462***	0.227***	0.217***	0.067***	0.028**	0.002	-0.094***
	(0.012)	(0.011)	(0.010)	(0.012)	(0.015)	(0.013)	(0.014)	(0.014)
Log hours		0.536***	0.417***	0.441***	0.465***	0.347***		
		(0.017)	(0.015)	(0.017)	(0.039)	(0.020)		
Log hours sq.		-0.012***	-0.003	-0.007**	-0.011	-0.003		
		(0.003)	(0.003)	(0.003)	(0.007)	(0.004)		
Female			-0.423***	-0.419***	-0.370***			
			(0.011)	(0.013)	(0.023)			
Education			0.023***	0.006	0.030***			
			(0.004)	(0.005)	(0.008)			
Education sq.			0.004***	0.004***	0.004***			
			(0.000)	(0.000)	(0.000)			
Ravens				0.070***				
				(0.007)				
Ravens sq.				0.012**				
				(0.005)				
Individual FE	N	N	N	N	N	Y	Y	Y
Time FE	N	Y	Y	Y	Y	Y	Y	Y
Switchers only					Y			
Obs	275600	275600	275600	201699	62944	275600	275600	275600
Individuals	31843	31843	31843	22899	5086	31843	31843	31843

(b) Kenya

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Log Wage	(8) Log Real Wage
Urban	0.862***	0.752***	0.679***	0.680***	0.367***	0.262***	0.165***	0.087*
	(0.039)	(0.038)	(0.035)	(0.036)	(0.055)	(0.047)	(0.050)	(0.050)
Log hours		0.106	0.050	0.084	0.506	0.259		
		(0.180)	(0.163)	(0.166)	(0.362)	(0.262)		
Log hours sq.		0.033*	0.034*	0.031*	-0.009	0.009		
		(0.020)	(0.018)	(0.018)	(0.039)	(0.029)		
Female			-0.514***	-0.485***	-0.368***			
			(0.037)	(0.038)	(0.069)			
Education			-0.025	-0.032	-0.104*			
			(0.035)	(0.037)	(0.060)			
Education sq.			0.006***	0.006***	0.010***			
			(0.002)	(0.002)	(0.003)			
Ravens				0.072***	0.078*			
				(0.022)	(0.040)			
Ravens sq.				-0.027	-0.014			
				(0.019)	(0.033)			
Individual FE	N	N	N	N	N	Y	Y	Y
Time FE	N	Y	Y	Y	Y	Y	Y	Y
Switchers only					Y			
Obs	134221	134221	134221	128215	39338	134221	134221	134221
Individuals	4791	4791	4791	4522	1037	4791	4791	4791

Panel A uses data from the IFLS and Panel B uses data from the KLPS. Please refer to Section 2.3 for further details on the data and to the notes of Table 2.5 for additional information on the variables. Please see the text in section 2.3 for details on the classification of the urban indicator. Column 5 only includes switchers, who are defined as individuals with at least one observation in both an urban and rural area. Each regression in columns 2–8 include quadratic controls for age. Standard errors clustered at the individual level are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.7: Gap in Earnings and Consumption for those Born in Rural and Urban Areas, Indonesia

(a) Indonesian individuals born in rural areas

	Dependent Variable: Log Earnings			Dependent Variable: Log Consumption		
	(1)	(2)	(3) Log Wage	(4)	(5)	(6)
Urban	0.635***	0.296***	0.037*	0.580***	0.307***	0.133***
	(0.018)	(0.015)	(0.020)	(0.018)	(0.013)	(0.020)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	179158	179158	179158	53840	53840	53840
Number of individuals	20010	20010	20010	22240	22240	22240

(b) Indonesian individuals born in urban areas

	Dependent Variable: Log Earnings			Dependent Variable: Log Consumption		
	(1)	(2)	(3) Log Wage	(4)	(5)	(6)
Urban	0.262***	0.091***	-0.024	0.206***	0.034***	-0.047***
	(0.020)	(0.016)	(0.020)	(0.020)	(0.012)	(0.015)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	96249	96249	96249	28363	28363	28363
Number of individuals	11784	11784	11784	12524	12524	12524

Columns 1–3 of Panels A and B repeat the analyses of Table 2.6a for those born in rural and urban areas, respectively, and columns 4–6 repeat the analyses of Table 2.9a for those born in rural and urban areas. Please refer to Section 2.3 for further details on the data and to the notes of Table 2.6 and 2.9 for additional information on the variables. Control variables include log hours worked, log hours worked squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only log hours worked, log hours worked squared, and age squared because the others are absorbed by the individual fixed effects. When using log wage as the dependent variable in column 3, only age squared is included as a control variable. Standard errors clustered at the individual level are in parentheses, *** $p<0.01$, ** $p<0.05$, *

$p<0.1$.

Table 2.8: Robustness to Alternative Agricultural Productivity Measures

(a) Indonesia

	Productivity Measure Includes...		Dependent variable:
Definition of Agriculture	Formal Wages	Self-Employed Profits	Log Wage
Majority of hours in agriculture			
Main Estimation	✓	✓	0.078***
			(0.021)
Any hours in agriculture	✓	✓	0.013
			(0.018)
All hours in agriculture	✓	✓	0.123***
			(0.020)
Majority of hours in agriculture	✓		0.021
			(0.026)
Majority of hours in agriculture		✓	0.084***
			(0.032)

(b) Kenya

	Productivity Measure Includes...		Dependent variable:
Definition of Agriculture	Formal Wages	Self-Employed Profits	Log Wage
Majority of hours in agriculture			
Main Estimation	✓	✓	0.061
			(0.106)
Any hours in agriculture	✓	✓	0.096
			(0.097)
All hours in agriculture	✓	✓	0.064
			(0.108)
Majority of hours in agriculture	✓		0.157
			(0.121)
Majority of hours in agriculture		✓	0.002
			(0.187)

Panel A uses data from the IFLS and Panel B uses data from the KLPS. Each row shows the robustness results of a regression of log wages (calculated as earnings per hour) on a non-agricultural indicator, age squared, and time and individual fixed effects. In each panel, the estimate in row 1 can be found in Appendix Table 2.5, column 7; row 2 can be found in Table A.5, column 4; row 3 in Appendix Table A.5, column 8; row 4 in Appendix Table A.8, column 4; and row 5 in Appendix Table A.9, column 4. All regressions report standard errors clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.9: Gaps in Consumption

(a) Indonesia

	Dependent variable: Log Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agr emp.	0.636***	0.250***	0.126***			
	(0.011)	(0.008)	(0.012)			
Urban				0.403***	0.139***	0.026**
				(0.011)	(0.008)	(0.012)
Individual FE	N	N	Y	N	N	Y
Controls and time FE	N	Y	Y	N	Y	Y
Obs.	82272	82272	82272	82272	82272	82272
Individuals	34820	34820	34820	34820	34820	34820

(b) Kenya

	Dependent variable: Log Meals Eaten					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-ag emp.	0.078***	0.059***	-0.090*			
	(0.016)	(0.017)	(0.049)			
Urban				0.029***	0.030***	-0.023
				(0.010)	(0.011)	(0.040)
Individual FE	N	N	Y	N	N	Y
Controls and time FE	N	Y	Y	N	Y	Y
Obs.	4203	4203	4203	4203	4203	4203
Individuals	3601	3601	3601	3601	3601	3601

Panel A uses data on total consumption from the IFLS, and Panel B uses data on meals eaten in the last day from the KLPS. Unlike previous tables, the sample includes individuals with and without earnings measures. Consumption data in the IFLS are obtained by adding up the value of food and non-food consumption in Indonesian Rupiah at the household level and dividing this by the number of household members. The data was collected for each of the five waves so each household has five observations at most. Separate analyses by food and non-food consumption in Indonesia can be found in Appendix Table A.11, and Appendix Table A.12 provides consumption analyses when using the sample with positive earnings measures. Data on meals eaten in Kenya are available from KLPS rounds 2 and 3 and refer to the day prior to the survey date. In the analysis sample, 0.6% of individual-time observations ate no meals in the prior day, 10.9% ate one meal, 53.2% ate two meals, 34.0% ate three meals, and 1.3% ate four or more. Control variables in both panels include age, age squared, years of education, years of education squared, and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared because the others are absorbed by the individual fixed effects. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.10: Urban/Rural Gap in Wages for Top 5 Cities

(a) Indonesia

	Dependent variable: Log Wages			
	(1)	(2)	(3)	(4)
Urban	0.356***	0.299***	0.081***	0.002
	(0.012)	(0.012)	(0.011)	(0.015)
Jakarta (population 10 million)		0.269***	0.243***	-0.033
		(0.020)	(0.018)	(0.034)
Surabaya (population 2.8 million)		0.017	-0.004	0.096
		(0.058)	(0.047)	(0.094)
Bandung (population 2.6 million)		0.262***	0.110**	0.229*
		(0.065)	(0.054)	(0.125)
Medan (population 2.5 million)		0.303***	0.251***	-0.069
		(0.047)	(0.044)	(0.104)
Bekasi (population 2.5 million)		0.628***	0.426***	0.112
		(0.063)	(0.056)	(0.080)
Individual fixed effects	N	N	N	Y
Control variables and time FE	N	N	Y	Y
Number of observations	275600	275600	275600	275600
Number of individuals	31843	31843	31843	31843

(b) Kenya

	Dependent variable: Log Wages			
	(1)	(2)	(3)	(4)
Urban	0.574***	0.373***	0.320***	0.071
	(0.040)	(0.060)	(0.055)	(0.063)
Nairobi (population 3.4 million)		0.324***	0.303***	0.156***
		(0.063)	(0.057)	(0.060)
Mombasa (population 1.2 million)		0.267***	0.265***	0.307***
		(0.079)	(0.075)	(0.094)
Kisumu (population 0.4 million)		0.014	0.090	0.099
		(0.153)	(0.142)	(0.244)
Nakuru (population 0.3 million)		0.252**	0.156*	0.136
		(0.118)	(0.094)	(0.155)
Eldoret (population 0.3 million)		0.148	0.105	-0.078
		(0.163)	(0.163)	(0.139)
Individual fixed effects	N	N	N	Y
Control variables and time FE	N	N	Y	Y
Number of observations	134221	134221	134221	134221
Number of individuals	4791	4791	4791	4791

Panel A uses data from the IFLS and Panel B uses data from the KLPS. Please refer to Section 2.3 for further details on the data and to the notes of Table 2.5a for additional information on the variables. The covariate "Urban" is an indicator variable that equals 1 if the person lives in an urban area, and five city indicators are included for the five most populous cities in Indonesia and Kenya, respectively. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 4, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 2.11: Integenerational Correlations of Cognitive Measures

(a) Indonesia

	Dependent variable: Normalized Ravens				
	(1)	(2)	(3)	(4)	(5)
Child Covariates:					
Born Urban	0.159***	0.169***	0.171***	0.146**	0.156***
	(0.056)	(0.056)	(0.060)	(0.058)	(0.059)
Female		-0.162***	-0.167***	-0.170***	-0.170***
		(0.055)	(0.055)	(0.053)	(0.054)
Parent (Averaged) Covariates:					
Born Urban			-0.015	-0.078	-0.076
			(0.068)	(0.066)	(0.066)
Age at Birth			-0.002	0.001	-0.000
			(0.002)	(0.002)	(0.003)
Years of Education				0.050***	0.041
				(0.008)	(0.030)
Normalized Ravens				0.063*	0.060
				(0.033)	(0.040)
Age, Education, and Ravens Squared	N	N	N	N	Y
Number of observations	861	861	861	861	861

(b) Kenya

	Dependent variable: Normalized Cognitive Ability Index				
	(1)	(2)	(3)	(4)	(5)
Child Covariates:					
Born Urban	0.344***	0.345***	0.369***	0.258***	0.258***
	(0.082)	(0.082)	(0.083)	(0.085)	(0.085)
Female		0.111	0.111	0.102	0.108
		(0.070)	(0.070)	(0.069)	(0.069)
KLPS Parent Covariates:					
Female			0.271***	0.308***	0.309***
			(0.075)	(0.073)	(0.074)
Age at Birth			0.002	0.013	0.270
			(0.014)	(0.014)	(0.195)
Years of Education				0.065***	0.068
				(0.014)	(0.067)
Normalized Ravens				0.055	0.053
				(0.040)	(0.041)
Age, Education, and Ravens Squared	N	N	N	N	Y
Number of observations	864	864	864	864	864

Panel A uses data from individuals in the IFLS, and Panel B uses data from a sample of children aged 3–5 of the KLPS adults. The Cognitive Ability index in Panel B is a composite of z-scores from six different tests of language, attention, memory, perception, and fine motor skills. Ravens matrices scores and the Cognitive Ability index are normalized to have mean zero and unit variance for full-year and six-month child age bins, respectively. In Panel A, parent covariates are averaged when both parents are available. In Panel B, only the covariates for the adult KLPS respondent are available (not for the spouse of the adult KLPS respondent). Regressions are clustered at the individual level in Panel A and at the parent level in Panel B. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Chapter 3

Government Decentralization Under Changing State Capacity: Experimental Evidence from Paraguay

3.1 Introduction

In standard models of delegation, devolution of decision-making from principals to agents is a way to take advantage of superior information that the latter are presumed to have (Aghion & Tirole, 1997; Dessein, 2002; Mookherjee, 2006). However, governments interested in taking advantage of these informational gains by decentralizing authority may face substantial administrative fixed costs. Justifying paying such costs depends on the total value of discretion, which in turn depends both on the presence of informational advantages and the planned scale of operation. For example, suppose that an organization plans to provide assistance to low-income families. If resources are sufficient to cover all households, then decentralizing the selection decisions has no value: the program can be rolled out from the center with universal coverage. If resources are insufficient to cover all households, then creating local branches to screen and prioritize households based on need may be more effective than using proxy-means testing and other crude information to ration access. But if resources are so meager that very few families can be covered in each district, the cost of developing the local branches may overshadow any potential improvement in the selection process.

How the state rolls out a new monitoring technology among its front-line providers—the subject of this paper—raises similar considerations. In 2014, the government of Paraguay decided to distribute GPS-enabled cell phones to supervisors to track their agricultural extension agents (AEAs). AEAs are tasked with visiting farmers scattered over large tracts of land and giving them access to various support services including timely information about prices and best farming practices. However, the central government suspected that AEAs were shirking due to the monitoring difficulties afflicting their supervisors and hypothesized that GPS phones could help mitigate the problem. Because the government did not have the resources to provide phones to all the AEAs at once, they faced two questions: 1) what should the extent of the roll-out be? and 2) who should decide which AEAs receive the phones?

In this paper, we develop an approach that allows us to measure not only the value of supervisors' information, but also how that value varies at different levels of coverage. First, we evaluate whether the new monitoring technology had an effect on shirking by estimating the impact on the share of assigned farmers that the AEA visited as well as measures of farmer satisfaction with the AEA services. Importantly, our design is well suited for evaluating whether supervisor-targeted AEAs were impacted more by the treatment. Specifically, prior to randomization, we elicited the preferences of supervisors regarding which half of their AEAs should be prioritized (selected) to receive the phone. We then randomly assigned phones to AEAs by their supervisors and effectively ignored the solicited priorities, generating a 2-by-2 treatment-by-selection matrix that forms the basis of our empirical analyses. Additionally, for a small, auxiliary, and random subset of AEAs, we implement the supervisor preferences and allocate phones to prioritized AEAs accordingly. This makes the preference solicitation incentive compatable.

Cell phones had a sizable effect on AEA performance, increasing the share of farmers visited in the last week by an average of 6 percentage points (pp). This represents a 22 percent increase over the AEAs in the control group. The cell phones also improve farmer satisfaction with and perceptions of their AEAs by 0.15 standard deviations. Because we do not find any impact of cell phones on AEAs who do not have supervisors, we interpret this effect to be a result of increased monitoring as opposed to the cell phones directly improving productivity. Also consistent with our interpretation, we find that treated AEAs perceive more monitoring from their supervisors. Finally, we find no evidence that treated AEAs increased the number of visits at the cost of conducting shorter ones.

Importantly, supervisor-chosen AEAs respond more to increased monitoring, entirely driving the average increase of 6 pp. Among these AEAs, treatment increased the likelihood that a farmer was visited in the past week by 15.4 pp compared to a statistically insignificant decrease of 3.6 pp among those who were not selected. This finding corroborates the notion that going down the hierarchy from the top program officers to local supervisors on the ground could allow the organization to leverage superior, dispersed knowledge about how best to allocate treatment.

While the research design divorces treatment and selection and allows us to evaluate heterogeneity along this policy-relevant dimension, supervisors in all cases maintain broad authority over AEAs, which may threaten interpreting treatment effects as being derived from monitoring if assignment to treatment results in Hawthorne effects or spillovers. We are able to rule out this threat using data from the group where supervisor preferences were implemented. In particular, we compare the performance of non-selected AEAs whose selected coworkers were either treated or not treated with phones. We find no significant differences. Correspondingly, we find no significant differences when we repeat the exercise with selected AEAs whose non-selected coworkers were either treated or not treated. Finally, we compare impacts in groups where supervisors had real discretion—several AEAs to choose from—to as few as a single AEA and find no significant differences.

Supervisors have superior information regarding AEA characteristics, only some of which are observable to the principal or an econometrician. Having collected a rich dataset on the AEAs that includes information on both cognitive and non-cognitive traits, we develop a two-step estimation procedure in the spirit of a sample selection model to decompose the value of information into observable and unobservable traits of an AEA. We use this to compute a series of marginal treatment effects under various selection rules and coverage rates. These marginal treatment effects are critical inputs to decide whether to decentralize the treatment assignment decision. In addition, the approach we develop would allow program leadership to optimize the program's roll-out scale.

We find that in general, both commonly observed demographic traits (e.g., gender) and even harder-to-measure characteristics such as cognitive ability or personality type do a poor job of explaining supervisors' targeting decisions. Among the few observable traits that predict targeting, the AEA's party affiliation is one of the most robust. Supervisors are much less likely to place members of the incumbent party under additional monitoring, suggesting that non-benevolent motives may have influenced, at least in part, their targeting decisions. Nevertheless, when we allow the treatment effects to vary by a rich set of observable characteristics, it is the unobservable component of the supervisors' choices that most robustly predicts the responsiveness of an AEA to the additional monitoring.

While our findings suggest that supervisors have valuable information, the decision of whether to decentralize depends on the counterfactual allocation regime of the principal, which depends on the information she has, the feasible allocation rules she can adopt, and the extent of available resources. In order to explore the potential for centralized versus decentralized assignment, we construct a number of counterfactuals corresponding to different degrees of sophistication of the central authority and scales of implementation. In particular, we use our model estimates for the distribution of treatment effects to compare the improvement in farmer visits under supervisor priorities to the improvement under four hypothetical allocation rules at varying coverage rates: 1) a totally uninformed principal who allocates randomly; 2) a minimally informed principal who targets AEAs who have to travel longer distances; 3) a more sophisticated principal who collects and analyzes baseline data on AEAs and targets predictably low productivity AEAs; and 4) the most sophisticated principal who pilots an experiment and thereafter targets AEAs in descending order of predicted responsiveness to treatment. We find that the value of supervisor information is substantial relative to a regime in which the principal simply allocates phones at random and that this difference in program impact is maximized at 53 percent coverage. At this coverage, the supervisor allocation increases the share of farmers visited by 6.9 pp versus an only 3.3 pp increase under random assignment. A slightly more effective approach compared to random assignment would be to simply allocate the phones to the AEAs who have to travel the longest distance to attend to their farmers. This method generally outperforms random assignment (a 2.0 pp advantage at 50 percent coverage), but the supervisor still outperforms this simple assignment mechanism.

A more effective centralized policy identifies and prioritizes the workers who are expected to be the least productive according to some observable characteristics without relying on reports from the supervisors. We operationalize this policy by estimating the relationship between AEA productivity and observable characteristics among the control AEAs without GPS phones. We then utilize this relationship to rank all AEAs based on their predicted productivity. Governments that have the information and capacity to prioritize the AEAs with the lowest predicted productivity are able to perform at least as well as, and in some cases better than, the supervisors. Such a regime succeeds because although the minimally informed principal does not have as much information as the supervisor, this procedure makes better use of the information that they do have. This in turn suggests that imperfect processing of information or bias prevents supervisors from being as effective as they could be.

The most effective but most information-demanding centralized policy we consider uses AEA observables to predict response to treatment rather than to predict baseline productivity. Under such a policy, the principal conducts a pilot experiment and uses the results to predict responsiveness among the remaining untreated AEAs. Treating AEAs in descending order of predicted responsiveness, even when lacking information on unobservables, vastly outperforms decentralized assignment by the supervisors. The high performance of these last two methods highlight that innovations in information and communication technologies as well as the introduction of experimental methods to inform policy can play a role in reducing information frictions and alter optimal organizational structure.

Our study speaks to several literatures. First and foremost, our study contributes to a large but mostly theoretical literature on why organizations decentralize decision-making authority.[1] Recently, some empirical progress has been made in understanding why private-sector firms decentralize. For instance, based on the insight by Aghion and Tirole (1997) that organizations are more likely to decentralize if the principal and agent have congruent preferences, Bloom, Sadun, and Van Reenen (2012) find that firms are more decentralized when located in regions that are judged to contain more trustworthy people by those in the headquarters location. The authors view trust as a proxy for congruency of preferences.

Given the standard assumption that agents are better informed than the principal, access to costly information can also determine a firm's decision to decentralize. For example, Acemoglu, Aghion, Lelarge, Van Reenen, and Zilibotti (2007) show using data on French and British firms in the 1990s that firms closer to the technological frontier, firms in more heterogeneous environments, and younger firms are more likely to choose decentralization—settings that presumably proxy for environments where learning is more difficult. Despite the progress that these and other studies have made, direct empirical evidence on the existence of superior information by agents is still lacking.

One notable exception is Duflo, Greenstone, Pande, and Ryan (2018) who conducted a

[1] Mookherjee (2006) provides an excellent review of the theory on decentralization.

field experiment that increased the frequency of inspections of industrial plants in Gujarat, India. In the control group, plants were audited as usual at the discretion of the inspectors, whereas in the treatment group, the audits were conducted more frequently but at random. They found that despite the increased regulatory scrutiny, the treatment plants did not significantly reduce pollution emissions. This is because the discretionary inspections targeted the plants with higher pollution signals. Because the largest penalties are reserved for extreme pollution violations, this is the population whose behavior is most likely to be impacted by audits.

We complement Duflo et al. (2018) in some important ways. Our experiment was designed to identify who the supervisors would target for monitoring without having to rely on strong functional form assumptions. As a result, we can experimentally identify the decentralized counterfactual to a centralized approach. Moreover, that counterfactual depends both on supervisors' informational advantage and on potential preference biases, which we allow for but are absent from the targeting rules in Duflo et al. (2018). Thus, we incorporate elements that are crucial to the evaluation of the relative merits of decentralization.

Similar to the public sector, private sector employers also need to monitor their employees. A paper by de Rochambeau (2017) discusses the roll-out of GPS tracking devices in a trucking company. She finds that managers choose to allocate the tracking device to drivers who perform less well at baseline and that these truckers benefit most from the device.

The problem of how best to deploy monitoring technology is similar to the issue of how best to target social programs. In this regard, our paper is most related to two studies. Alderman (2002) examines an Albanian social assistance program. He shows that even after controlling for the assets that local officials used to target beneficiaries, household consumption was still predictive of who received the program. The author interprets this as evidence that these local officials relied on their local information and discretion. Alatas, Banerjee, Hanna, Olken, and Tobias (2012) conducted a field experiment in 640 villages in Indonesia to compare proxy-means testing against community-based targeting of a social program and find that the former corresponds closer to consumption than the latter. They argue that this difference is not due to elite capture or local information, but rather a difference in how local communities define poverty. Similar to difficulties in the context of social programs, banks could benefit from community knowledge to help them lend to the most entrepreneurial people. Hussam, Rigol, and Roth (2017) find that community members have useful information on marginal returns and this information is useful above and beyond what a machine learning algorithm would predict from observables.

Our study also has clear parallels to the literature on applying marginal treatment effects (MTE) to construct policy-relevant counterfactuals (Heckman & Vytlacil, 2005). As in the MTE literature, we express our evaluation problem as a joint model of potential outcomes and selection as determined by a latent index crossing a threshold. In contrast with the standard MTE setup, our selection equation does not model an AEA's self-selection into treatment but rather the selection by a supervisor. Crucially, treatment is not contingent on being selected—only those AEAs who were randomized into treatment were in fact treated. Thus, when we compute the MTEs we do not have to extrapolate to subgroups of "always-takers" and "never-takers" because we only have compliers by design. In this respect, our approach implements a variant of the selective trial designs proposed by Chassang, Padró I Miquel, and Snowberg (2012). In that paper, the authors recast randomized control trials into a principal-agent problem and show theoretically how one can recover the MTEs necessary to forecast alternative policies and treatment assignments by eliciting subjects' willingness to pay for the treatment. Instead of eliciting our agents' willingness to pay for the treatment, we elicit the targeting preferences of the supervisor, who in our context is the relevant decision maker.

Finally, our study adds to a growing body of experimental evidence on the impact of new monitoring technologies for reducing shirking in the public sector. Similar to our setting, some of these studies involve weak or no explicit financial incentives. For example, Aker and Ksoll (2018) monitored teachers of adult education in Niger by calling both the teacher and the students to ask whether the class was held and how many students attended. They found that the calls led to fewer canceled classes and better student test scores. Callen, Gulzar, Hasanain, Khan, and Rezaee (2018) used a similar cell phone technology to monitor health facility inspectors and found that this increased the frequency of inspections, especially for those with 'better' personality traits.

Other studies have introduced new technologies for monitoring but have also overlaid financial incentives. For instance, Duflo, Hanna, and Ryan (2012) required teachers to take a picture of themselves with their students at the beginning and end of each school day using a camera with tamper-proof date and time functions, whereas Banerjee, Glennerster, and Duflo (2008) asked nurses to time-stamp a register at the beginning, middle, and end of the day. Both studies found these treatments increased teacher and nurse attendance, but in both cases, the impact was found to be mostly due to concomitant financial incentives. Dhaliwal and Hanna (2017) found that fingerprint readers in health centers decreased absence even though financial incentives provided by the monitoring technology were rather weak. Banerjee, Chattopadhyay, Duflo, Keniston, and Singh (2015) and Khan, Khwaja, and Olken (2016) look at on-the-job performance rather than attendance (among police and tax collectors respectively) and employ both monitoring and incentives. These papers do not give a definitive answer regarding whether most of the improvement in performance is due to the monitoring or the incentives. The first paper suggests a significant impact of monitoring alone, while the second suggests an insignificant impact.

We contribute to this literature by showing that a cell phone technology can be effective in reducing shirking for individuals such as agricultural extension agents whose job requires them to visit farmers who live out in rural areas, often quite far from the local agricultural ministry offices in town.

3.2 Background

Agricultural extension services in Paraguay are centered around the Ministry of Agriculture based in Asunción. Below the central ministry are 19 Centros de Desarrollo Agropecuario

(CDAs, which exist at the department level, similar to a state in the United States) and below the CDA level there are 182 Agencias Locales de Asistencia Técnica (ALATs, which are at the municipality level, similar to a county in the United States). The Paraguayan Ministry of Agriculture has close to 1000 agricultural extension agents working within ALATs spread across four main agencies. We work with the biggest of these agencies, Dirección de Extensión Agraria (DEAg).

The main job of extension agents is to help farmers access institutional services that will help them improve their production. The goal is to increase farmers' output directed both for own consumption as well as the market. Another goal is to increase farmers' connection to, and participation in, markets. The official thematic areas are soil improvement, food security, product diversification, marketing, improving quality of life, and institutional strengthening. Much of what extension agents do resembles the role of middlemen, connecting farmers with cooperatives, private enterprises, and specialists. Extension occurs both one-on-one and in group meetings. Extension agents conduct farm visits in which agricultural problems are diagnosed and addressed. Group meetings are used to lead demonstrations or talk about technical topics. AEAs also organize farmer field trips. Each extension agent is assigned to work with approximately 80 producers. Extension agents do not usually offer free goods or services to farmers. Although the headquarters for extension agents are in towns, most of their daily work involves driving out to visit farmers in the rural areas where these farmers live and work. Extension agents come from a variety of backgrounds including agricultural sciences, veterinary sciences, nutrition, law, and teaching.

Within every ALAT there is a supervisor who, in addition to working with his own farmers, must also monitor the other extension agents working in the ALAT. We will refer to individuals who work purely as agricultural extension agents as 'AEAs.' By this definition, DEAg has over 200 AEAs working within the organization at any time.

In June 2014, the Ministry of Planning, in association with the Ministry of Agriculture, decided to provide AEAs with GPS-enabled cell phones. While all AEAs already owned their own personal cell phones, these were not necessarily meant for work, and AEAs would have to pay to make calls or send messages on their personal phone. This initiative had several objectives. One was to improve coordination and communication between the AEAs and their supervisors. For example, it would give the AEA a mechanism to take a picture of a farmers' crop which was suffering from some pest, circulate it, and get a response for the farmer of how to deal with that pest. But crucially, it would allow the supervisors to see where AEAs were at all times, how long they spent in each place, and what they did there (since the AEA is supposed to document every meeting in which he participates). AEAs can submit reports and review reports they have already submitted through the phone. Supervisors can view reports submitted by all the AEAs they oversee.

In the terms of the hierarchical agency model we lay out in the next section, we view the ministerial leadership introducing the new technology as the principal, we will refer to the ALAT-level supervisors as "supervisors," and the AEAs as the "agents."

3.3 Model

Consider a hierarchy composed by a principal, a supervisor, and a continuum of agents with mass 1. The supervisor is responsible for monitoring the agents. In such a hierarchy there are two possible agency problems: that between the agent and his supervisor and that between the supervisor and the principal. We will focus mainly on the problem between agent and supervisor, and analyze how it changes when the agents are placed under a new monitoring technology. The question will be whether the principal can obtain better results by relying on supervisors in deciding how to deploy the technology.

Agents and monitoring Each agent caters to a mass 1 of farmers. A visit by an agent i yields a constant benefit B to each farmer. Agents receive a wage w and choose a share $s_i \in [0,1]$ of farmers to visit. The agent obtains an intrinsic motivation $m_i s_i$ from visiting a share s_i of farmers, but also incurs a cost $a_i s_i + b_i \frac{s_i^2}{2}$. The share s_i is a measure of agent effort, and because it directly constitutes a measure of service provision (visits to farmers), the principal cares about it. From now on, we will refer to s_i as effort and assume that it is noncontractible.

The supervisor operates a monitoring technology such that with probability $q_i \in (0,1)$, she learns s_i and reprimands the agent in proportion to the amount by which his effort falls short, $1 - s_i$. The agent gets a disutility from being reprimanded equal to $(1 - s_i)r_i$, with $r_i > 0$.[2] While monitoring allows the supervisor to obtain information about the agent's effort, it can potentially weaken intrinsic motivation. When monitored, the intrinsic motivation payoff of agent i becomes $(m_i - g_i)s_i$, which is potentially negative. It reflects the fact that agents may feel aggrieved to an extent $g_i \geq 0$ when under close supervision. In sum, wages and effort costs accrue to the agent regardless of supervision, while reprimand and intrinsic motivation payoffs accrue in relation to monitoring intensity q_i. Thus, agent i can be seen to maximize utility,

$$u_i\left(s_i\right) = w - a_i s_i - b_i \frac{s_i^2}{2} + q_i\left[\left(m_i - g_i\right)s_i - \left(1 - s_i\right)r_i\right] + \left(1 - q_i\right)m_i s_i,$$

or, collecting terms,

$$u_i\left(s_i\right) = \omega_i + \mu_i s_i - b_i \frac{s_i^2}{2} + q_i s_i \rho_i,$$

where $\omega_i \equiv w - r_i q_i, \mu_i \equiv m_i - a_i, \rho_i \equiv r_i - g_i$. Agent i chooses the share s_i of farmers to visit to maximize utility $u_i(s_i)$, and he does so after learning the level of monitoring intensity q_i

[2]Alternatively, one may assume that the supervisor draws a farmer at random, and finds he has not been visited with probability $1 - s_i$, in the event of which she proceeds to reprimand the agent with a fixed intensity r_i. It is also possible to extend the model to make q_i a function of monitoring effort by the supervisor. The choice of monitoring effort remains unmodeled here, in order to stick with the simplest formulation that will deliver the results of interest. Such an extension could also involve an agency problem in the supervisor's choice of monitoring effort without affecting the essence of our results. The only tension between supervisor and principal that may arise in our simpler setting relates to the deployment of the monitoring technology to be described below.

he is under. Because $u_i(s_i)$ is concave, agent i's optimal effort is $s_i^*(q_i) = \max\left\{0, \frac{q_i\rho_i+\mu_i}{b_i}\right\}$. Since b_i only affects effort through ratios involving ρ_i and μ_i, parameters that can be scaled arbitrarily, we normalize $b_i = 1$, yielding,

$$s_i^*(q_i) = \max\{0, \min\{q_i\rho_i + \mu_i, 1\}\}. \tag{3.1}$$

The term μ_i – a proxy for net-of-cost intrinsic motivation – is individual-specific and for some agents potentially negative. Even more important for our purposes, the term ρ_i, which captures both the agent's distaste for being reprimanded (which raises effort) and his resentment at being monitored (which lowers effort) is also potentially negative for some agents. We will assume ρ_i to be drawn from a continuous distribution $F(\rho_i)$ over a support $[\rho_l, \rho_h]$, where $\rho_h > 0$ but ρ_l is potentially negative.

New technology and treatment effects We assume that q_i can take one of two levels $\{q_l, q_h\} \in (0,1)$, with $q_h \equiv q_l + t_i\Delta q$, $\Delta q > 0$, where q_l denotes a status quo level of monitoring, and $t_i \in \{0,1\}$ reflects whether agent i is "treated" with a new monitoring technology.[3] In order to characterize treatment effects neatly and avoid awkward truncation issues, in what follows we will assume that $min\{q_h\rho_l, q_l\rho_l\} + \mu_i > 0$ and $q_h\rho_h + \mu_i < 1$, which guarantees interior solutions for s_i.

While μ_i and ρ_i both affect the level of effort, only ρ_i affects the response of effort to a change in monitoring technology. Thus, in the remainder of this section we will refer to different levels of ρ as agents' "types." Under increased monitoring, an agent of type ρ increases his effort by $\rho\Delta q \equiv T(\rho)$, which captures the treatment impact of the new technology for that agent. Note that since ρ_l can be negative, $T(\rho)$ can be negative for some types. To deploy the new monitoring technology on any given agent costs an amount c per agent. If the new technology is deployed over all agents, then all agents are "treated." Given a continuum of agents, the total (and average) treatment impact is $\int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho$, achieved at a total (and average) cost c. If the new technology is deployed on all agents with type above some level k, the total treatment impact over all agents is $\int_{k}^{\rho_h} T(\rho)f(\rho)d\rho$, achieved at total cost $c(1-F(k))$. Note that our definition of the total treatment abstracts from spillover effects across agents. These effects could be modeled, but as we report below we do not find evidence of spillover effects. Thus, we keep the theory consistent with our empirical approach to measuring marginal treatment effects, which will likewise abstract from spillovers.

Inspection of the expression for the treatment impact $\int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho$ yields the following:

Remark 1. *If $\rho_l \geq 0$, the total (and average) treatment impact is guaranteed to be positive. If $\rho_l < 0$, the total (and average) treatment impact is positive if and only if, given Δq, the density $f(\cdot)$ places enough weight on positive types.*

This remark highlights the conditions under which a new technology rolled out to all agents (or a representative sample of them) would yield positive results when assessed

[3] Here we assume treatment only affects the agent's problem by raising monitoring intensity, although it could in principle also affect μ_i via the agent's cost a_i. This is plausible as some technologies, like GPS phones, can be productivity-enhancing. However, as we will show later, the data do not support that possibility.

through a standard impact evaluation that estimates average treatment effects. In addition, the definition of $T(\rho_i)$ implies that equilibrium agent effort s_i^* (weakly) increases in monitoring technology q_i for all agents with $\rho_i > 0$, and an improvement in monitoring technology (an increase in q_i) has a more positive effect on the effort of agents with a higher type ρ_i.

The value of information and optimal decentralization To isolate a central advantage of decentralization, we assume that the principal knows the distribution of types $F(\rho)$, but does not know the type of any specific agent. The supervisor, in contrast, knows both $F(\rho)$ and agents' individual types – this constitutes the information advantage of the supervisor vis-a-vis the principal. Both principal and supervisor know all other model parameters. The thought experiment of interest is whether, given a new monitoring technology, the principal would want to delegate to supervisors the choice of which agents to treat.

Centralization We take a centralized regime to be one in which the principal makes all decisions without any further input beyond what she already knows, namely $F(\rho)$.[4] Thus, she can only make a general decision about whether to adopt the new technology or not and cannot determine whether any one agent is more profitably treated than another. We denote the scale of adoption with m (for the *measure* of the treated, not to be confused with the individual-specific intrinsic motivation m_i used earlier). If roll-out has scale m, and treated agents are selected at random, the total treatment impact will be $m \int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho$, which increases linearly in m as illustrated by the strictly increasing diagonal line in Figure 3.1 depicting the impact of random treatment assignment. The cost will be mc. The principal will adopt whenever $\int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho \geq c$ (breaking indifference in favor of adoption), i.e., whenever the average treatment effect of the new technology is larger than its marginal and average cost. If this condition is met, a roll-out at 100% would produce a total treatment impact equal to the average treatment effect, recommending not only adoption, but also adoption at full scale.

Decentralization In the decentralized regime, the principal can pay a cost $d \geq 0$ to delegate to the supervisor the decision over which agents to place under the new monitoring technology.[5] For simplicity, we focus on a well-meaning supervisor who deploys the new

[4]We equate centralization with a regime where the principal makes all decisions based upon her own information, and decentralization to one where the principal delegates decisions to supervisors, or, equivalently, one where supervisors submit information that mechanically drives the principal's decisions. Thus, we abstract from the interesting distinctions made by Dessein (2002) between delegation and strategic communication.

[5]The delegation cost can arise due to the need to transfer certain administration means to the supervisor or from establishing additional communication and administration channels to track the supervisor's recommendations and/or technology deployment decisions. In some empirical settings, like the one discussed in the introduction on screening candidates for income support, the costs of decentralization are fixed and likely large. The reason is that identifying the best units to treat–even if they are just a few–may require deploying a nation-wide organizational operation. In other settings costs may be small, and in others even negative, since centralization may at times be costlier. In the latter cases, there will be no tension – decentralization is both informationally advantageous *and* cheaper – and therefore of less analytical interest to us. In the case in which decentralization costs are variable rather than fixed there will be quantitative differences in terms

technology to maximize agent output. Our empirical approach allows for potential supervisor bias. If the marginal cost of the new technology is lower than the treatment effect for the type with highest type ρ_h, a benevolent supervisor will place agents under the new monitoring system starting from the highest type ρ_h and work downwards. How far down he goes depends on the scale of roll-out for the new technology. If the supervisor chooses which agents to treat but the scale of roll-out is fixed, he will choose the highest types to fill the quota. Thus, if the supervisor is told to place a share m of agents under the new technology, he will treat every agent with type $\rho \in [\rho_m \equiv F^{-1}(1-m), \rho_h]$. This implies,

Remark 2. *If supervisors know agents' types and assign treatment with a benevolent intent to maximize visits to farmers, treatment effects on agents selected by supervisors will be higher than treatment effects on agents selected at random.*

If, against our assumptions, the supervisor is not well informed, then the treatment effects among agents selected by the supervisor could be similar to those among agents selected at random. If the supervisor has mistaken views or is not benevolent, treatment effects among agents selected by him could be even lower than among those selected at random: a supervisor who ranks types to be treated in an inverse way (i.e., starting with ρ_l and working upwards) would in fact minimize the impact of technology adoption.

Returning to the case of an informed and benevolent supervisor, it is helpful to consider the situation in which the supervisor also has control over the scale of adoption. In this situation, she will choose the lowest treated type k to maximize,

$$\int_k^{\rho_h} T(\rho) f(\rho) d\rho - c(1 - F(k)),$$

which yields $T(k^*) = c$. In words, the supervisor will choose to treat every agent down to a type k^* whose marginal treatment effect from the new technology equals the marginal cost.

Optimal decentralization Consider the case where the supervisor, under decentralization, has authority over the selection of agents to be treated but not over the scale of technology adoption.[6] Given a scale of adoption m, the principal will choose to decentralize if and only if $\int_{\rho_m}^{\rho_h} T(\rho) f(\rho) d\rho - cm - d \geq \left(\int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho - c\right) m$, or equivalently, iff,

$$\iota(m) \equiv \int_{F^{-1}(1-m)}^{\rho_h} T(\rho) f(\rho) d\rho - m \int_{\rho_l}^{\rho_h} T(\rho) f(\rho) d\rho \geq d, \tag{3.2}$$

where $\iota(m)$, graphed in the bottom panel of Figure 3.1, captures the informational gain from decentralization. This gain is the difference between the total treatment effect that can be attained through the centralized and decentralized approaches.

of the roll-out rates that make centralization preferred to decentralization, but the basic point will remain that marginal treatment effects and roll-out rates matter.

[6] A realistic example fitting our empirical setting is when a new technology is acquired by government and is made available to an agency in a fixed amount.

Note that when $m = 0$ and $\rho_m = \rho_h$, the marginal gain from expanding roll-out under the decentralized scheme is at a maximum since the supervisor would treat the most responsive agent first. But because the new technology would be applied to very few agents, the value of the informational gain from decentralization is zero and does not justify paying a fixed positive cost d to decentralize. On the other extreme, where $m = 1$ and $\rho_m = \rho_l$, the difference in value again goes to zero because the advantage of treating the more responsive agents first is completely diluted. Since all agents will be treated, there is no need to decentralize.

For every value of m strictly between 0 and 1, the total treatment effect attained by a supervisor who treats the most responsive agents first is larger than that which can be attained by assigning treatment at random. As illustrated in Figure 3.1, the value of information $\iota(m)$ is decreasing in m near $m = 1$ (or, equivalently, increasing in ρ_m near ρ_l). This is true up to a type $\bar{\rho}$ for whom the marginal treatment effect $T(\bar{\rho})$ is equal to the average treatment effect $\bar{T} \equiv \int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho$.[7] The value of information $\iota(m)$ then increases in m as m approaches 0 (or equivalently decreases in ρ_m as ρ_m approaches ρ_h). Given these considerations, and recalling continuity of F, standard intermediate value theorem arguments imply:

Proposition 1. *(i) If $0 < \iota(\bar{\rho}) < d$, then decentralization is never optimal and if $d = 0$, decentralization is always optimal. (ii) If $0 < d < \iota(\bar{\rho})$, there exist two values $m' < m''$ in $[0, 1]$ such that for any scale of roll-out of the new technology $m \in [m', m'']$ the principal prefers decentralization to centralization; whereas for $m \notin [m', m'']$, centralization is preferred.*

This proposition establishes that the case for decentralization rests on the value of its informational gain relative to its cost, which in turn depends crucially on the scale at which the new technology is to be adopted.

According to our model, there are interventions which can never yield positive value if implemented centrally and/or fully, but could deliver value if implemented in a decentralized manner with a limited roll-out. To see this, suppose an intervention satisfies $\int_{\rho_l}^{\rho_h} T(\rho)f(\rho)d\rho < c$, yielding an average treatment impact below marginal and average cost, so adoption at a 100% scale would yield a loss. But suppose also that treatment impact is larger than cost for a set of highest types, so that $T(\rho) > c$ for all types in $\rho \in (\rho', \rho_h]$, with $T(\rho') = c$, and that $\int_{\rho'}^{\rho_h} T(\rho)f(\rho)d\rho > c(1 - F(\rho')) + d$. In other words, there is a set of types for whom treatment effects are larger than the marginal cost by more than the cost of decentralization. In this situation, the principal would gain by delegating to the supervisor the adoption decision if the latter will treat only those types in $(\rho', \rho_h]$. This suggests that impact evaluation should not abstract from the extent of roll-out and its implementation mode, i.e., centralization versus decentralization, since the implementation mode affects who gets treated. In other words, determining whether a technology is valuable – presumably the ultimate goal of an impact evaluation – requires assessing the likely total treatment impact under different roll-out extents under both the centralized and decentralized approaches.

[7] Differentiating $\iota(m)$ we get $\left(T(\rho_m) - \bar{T}\right) f(\rho_m)$, where $f(\cdot) > 0$.

Our empirical study will investigate three claims stemming from the two remarks and proposition derived in this section. First, does the intervention at hand deliver a positive treatment impact on average? Second, do supervisors have valuable knowledge (net of potential limitations in benevolence) about which agents ought to be treated given partial roll-out? If so, we should observe treatment effects among those selected by supervisors to be larger than among agents who were selected into treatment at random. Third, could the scale of roll-out alter the relative advantage of decentralization vs centralization? Answering these questions will require developing a method for ascertaining the marginal treatment impact on different types of agents while characterizing centralization approaches with varying levels of information.

3.4 Research Design

Our experiment was conducted on 180 local technical assistance agencies (ALATs, Agencia Local de Asistencia Técnica). On average, each ALAT consists of a supervisor and three agricultural extension agents (AEAs). Many ALATs have a single AEA, but 48 ALATs have at least 2 AEAs. We asked the supervisors of the latter group to indicate which half of her AEAs should receive the phones first given the program's objective to increase worker performance. We refer to these AEAs as "selected." AEAs were not told that their supervisor was asked to make such a decision, and were not told who was selected. These 48 ALATs were then randomly assigned into three groups according to how and when the agents would receive their phones.

The main group of ALATs is in cells A, B, C, and D in Figure 3.2. The ALATs in cells B and D (a quarter of the ALATs), serve as the treatment group. In these ALATs all AEAs, both selected and non-selected, received the GPS-enabled cell phone which increased monitoring. The ALATs in cells A and C, (half of the ALATs), serve as our control group as no AEAs received the phones in these groups. The average difference in performance between AEAs in cells B and D and AEAs in cells A and C estimates the average impact of treatment. And, the difference-in-differences computed as the performance by AEAs in cells $(B - A) - (D - C)$ estimates whether the impact on selected AEAs is larger than the impact on non-selected AEAs. This difference-in-differences allows us to determine whether supervisors had valuable information on how to direct treatment. A third group of ALATs (cells E and F) received partial treatment. Only those AEAs who had been selected by their supervisors for treatment were treated immediately (cell E). This design helped make the elicitation of supervisors' preferences credible and relevant. Eight months after the delivery of these phones, a second wave of phones were delivered to the non-selected AEAs in those ALATs, group F.

The difference in performance between the AEAs in cell F and in cell C provides a test of whether allocating phones to the selected AEAs can also affect the performance of non-selected AEAs in the same ALAT. This would be the case if the supervisors also responded to the treatment of those in cell E by monitoring more intensively the AEAs without the cell phone in cell F.

Taking the Theory to the Data

Recall that the performance of AEAs whenever interior is given by equation (3.1):

$$s_i^* = q_i \rho_i + \mu_i.$$

To operationalize this equation, recall that the level of monitoring for each AEA, q_i, is a function of the monitoring technology t_i, according to the expression $q_i = q_l + \Delta q t_i$ where t_i takes a value of 0 when AEAs do not get a cell phone and 1 when they do. Because our objective is to see AEAs respond to exogenous changes in monitoring q_i, we normalize $q_l = 0$ and can rewrite the expected disutility of being reprimanded (net of monitoring grievance) to be $q_i \rho_i = \beta_i t_i$, where $\beta_i = \Delta q \rho_i$.

The central goal of our approach is to model various selection criteria and estimate the marginal treatment effects under each criterion for varying levels of roll-out. A key element in this approach will be to consider different degrees of observability of the individual parameters (μ_i, β_i), in an individual AEA's effort function in equation $s_i^* = \mu_i + \beta_i t_i$. In particular, we will map these parameters into a vector of fixed characteristics $(\boldsymbol{X}_i)$ and two independently random characteristics (ε_i, η_i), to write: $\mu_i(\boldsymbol{X}_i, \varepsilon_i)$ and $\beta_i(\boldsymbol{X}_i, \eta_i)$, so that,

$$s_i^* = \mu_i(\boldsymbol{X}_i, \varepsilon_i) + \beta_i(\boldsymbol{X}_i, \eta_i) t_i. \tag{3.3}$$

While the vector $\boldsymbol{X}_i$ may be observable to both the principal and supervisor, the elements (ε_i, η_i), may only be partially observed by the supervisor.

Average treatment effect We can estimate the average treatment impact of the cell phone on effort by imposing some familiar (but mild) structure on individual parametric heterogeneity as follows: $\mu_i = \mu' \boldsymbol{X}_i + \varepsilon_i$ and $\beta_i = \beta_0$. An individual AEA's effort function becomes

$$s_i^* = \mu' \boldsymbol{X}_i + \beta_0 t_i + \varepsilon_i, \tag{3.4}$$

where s_i^* measures the share of farmers AEA i visited in the past week. The coefficient β_0 provides a causal estimate of the difference in performance between AEAs in both treated cells B and D relative to AEAs in the control cells A and C. Thus, the first theoretical claim that the intervention yields positive value is captured by contrasting the null hypothesis of $\beta_0 = 0$ against the alternative that $\beta_0 > 0$.

Given our research design, we cluster the standard errors at the ALAT level and also report p-values based on a score bootstrap procedure procedure to account for the fact that we have relatively few clusters (Kline & Santos, 2012; Wu, 1986) as well as randomization inference p-values. In estimating equation (3.3), we can also include the single AEA ALATs, which were assigned phones at random. In this randomization, one-third of AEAs initially received a phone, with two-thirds serving as a control. When including these ALATs, the vector $\boldsymbol{X}_i$ contains an indicator for whether or not the ALAT has a single AEA.

Average treatment effect by supervisor's choice To test whether supervisors are able to select those AEAs whose effort would most increase when monitored, we can simply re-parameterize $\beta_i = \beta_0 + \beta_1 D_i^S$, where D_i^S is an indicator for whether AEA i was selected to receive a phone. Equation (3.3) then becomes

$$s_i^* = \mu' \boldsymbol{X}_i + \beta_0 t_i + \beta_1 (D_i^S \times t_i) + \varepsilon_i, \tag{3.5}$$

where included within the vector $\boldsymbol{X}_i$ is the indicator D_i^S. With this specification, we can compare the difference in performance between selected AEAs in the treatment and control groups (cells $B - A$) net of the difference in non-selected AEAs in the treatment and control group (cells $D - C$). Thus, the second theoretical claim that supervisors have valuable information about which AEAs should be targeted is captured by contrasting the null hypothesis $\beta_1 = 0$ against the alternative $\beta_1 > 0$. We directly observe s_i^* and can thus estimate μ' and $\boldsymbol{\beta} = (\beta_0, \beta_1)$ via ordinary least squares since t_i is randomly assigned. This is because the supervisor's selection D_i^S is elicited in a way that does not affect treatment assignment in cells A, B, C, and D.

Estimating the Marginal Treatment Effects of the Program

A strictly positive value for β_1 in estimating equation (3.5) is a necessary condition for a decentralized approach to be preferred, but it is not a sufficient condition. Two other considerations are pertinent. First, is the value β_1 large enough to justify paying the cost d of decentralization? Second, what would the average treatment effect be at scales other than 50 percent? We asked supervisors to select half of their AEAs but this pilot implementation does not directly tell us what β_1 would be at different selection shares. In this section we develop a method for tracing out the impact for all possible roll-out scales under different implementation regimes that vary the degree of informational advantage associated with decentralization.

Marginal treatment effects under different selection models

In order to lay out the main intuitions surrounding the value of decentralization, our theory considered the stark contrast between a totally uninformed principal and a fully informed, benevolent supervisor. We will allow for intermediate cases in our empirical approach – the econometric operationalization of the theory will in fact extend it in two directions. First, we allow for supervisors to be less than fully benevolent. Second, we allow them to be less than perfectly informed about the responsiveness of AEAs to treatment. In addition, this framework will allow us to consider a principal who is partially informed.

Each organizational situation – decentralization or centralization under different informational capabilities of the principal – will be modeled as leading to the selection of AEAs according to a suitably defined latent index model.

When we implement empirically the study of supervisors' choices, how worthy of treatment a particular AEA is in the eyes of the supervisor will be seen as a function of observables

X_i and unobservables u_i according to the function $\Gamma' X_i + u_i$. In what follows we develop some structure to link this empirical object to the theory.

In the case of decentralization, supervisors select AEAs according to some value they perceive from treating supervisor i,

$$v_i = \beta_i(\boldsymbol{X}_i, \eta_i) + \psi_i(\boldsymbol{X}_i, \zeta_i), \tag{3.6}$$

where v_i is AEA i's desirability for selection as seen by the supervisor, $\beta_i(\boldsymbol{X}_i, \eta_i)$ represents the heterogeneous effect of receiving the cell phone and ψ_i is a preference for treating AEA i that depends on $\boldsymbol{X}_i$ and an independent, idiosyncratic preference term ζ_i. A benevolent supervisor would only select AEAs based on an index $v_i = \beta_i(\cdot)$. Thus, the additional term ψ_i captures the potential non-benevolence of the supervisor. In addition, supervisors may not observe η_i perfectly but, instead, observe a signal $\theta_i = \eta_i + \xi_i$, where $\xi_i \sim F_\xi(\cdot)$ is a white noise (hence mean zero) term; as the variance of ξ_i goes to zero, the supervisor gets closer to being perfectly informed. Given the random element (ξ), the supervisor faces uncertainty. A risk neutral supervisor will assign monitoring technology to AEAs depending on the expected value $\mathbb{E}\{v_i | \boldsymbol{X}_i, \theta_i, \zeta_i\}$. The expectation is taken over ξ, and conditional on ζ, since to the supervisor the former represents noise while the latter may capture preferences.

Given a selection criterion (such as v_i), and a well-defined measure of diversity across AEAs as given by a joint distribution over $(\boldsymbol{X}_i, \theta_i, \zeta_i)$, it is possible for the supervisor to rank order all AEAs according to the value $\mathbb{E}\{v_i | \boldsymbol{X}_i, \theta_i, \zeta_i\}$, with minimum element $\underline{Ev}$ and maximum element $\overline{Ev}$. We assume there is enough variation that the rank order is strictly monotonic. Therefore, any roll-out of scale m under a selection criterion based on v_i implies treating all AEAs who satisfy $\mathbb{E}\{v_i | \boldsymbol{X}_i, \theta_i, \zeta_i\} \geq c_p(m)$, where $c_p(m)$ is a putative cost (hence the subscript). This cost is putative in the sense that it is the cost of treatment that the supervisor would have to perceive in order to decide to treat a share m of AEAs. Thus, $c_p(m)$ satisfies $\frac{dc_p}{dm} < 0$, $\lim_{m \to 0} c_p(m) = \overline{Ev}$, and $\lim_{m \to 1} c_p(m) = \underline{Ev}$. These conditions say that for the supervisor to want to treat more AEAs, the putative cost of treatment must be lower; for the supervisor to treat no AEAs, the putative marginal cost of treating a single AEA must exceed the benefit of treating the most valuable AEA; and that for the supervisor to treat all AEAs, the expected desirability of treating the least valuable AEA must cover the putative cost. When $\mathbb{E}\{v_i | \boldsymbol{X}_i, \theta_i, \zeta_i\} \geq c_p(m)$ is true, the selection indicator denoted by $D_i^M(\boldsymbol{X}_i, \theta_i, \zeta_i, c_p)$ takes the value 1, and 0 otherwise.

The fundamental difference between $\boldsymbol{X}_i$ and η_i is that elements in the vector $\boldsymbol{X}_i$ are potentially observable by a sophisticated principal who can gather and analyze data. Elements in $\boldsymbol{X}_i$ could contain AEA-related demographic and psychometric data. The term η_i is fully unobservable to the principal, and can potentially be known only to a supervisor who establishes a more personal connection with the AEA. Thus, decentralization has two potential informational advantages: supervisors may (or may not) know and use data on $\boldsymbol{X}_i$ better than the principal, and they are the only ones who can potentially know something about η_i. To the extent that η_i enters the function $\beta_i(.)$ the supervisor will have an unassailable informational advantage over the principal.

To make further progress, we need to parameterize the dependence of $\mu_i(\cdot)$, $\beta_i(\cdot)$, and $\psi_i(\cdot)$ on $\boldsymbol{X}_i$. We parameterize each of these linearly. Slightly abusing notation, and anticipating our assumption that η is mean zero, we can re-write equations (3.3) and (3.6) respectively as,

$$\begin{aligned} s_i^* &= \underbrace{(\boldsymbol{\mu}'\boldsymbol{X}_i + \varepsilon_i)}_{\mu_i(\cdot)} + \underbrace{(\boldsymbol{\beta}'\boldsymbol{X}_i + \eta_i)}_{\beta_i(\cdot)} t_i \\ &= \boldsymbol{\mu}'\boldsymbol{X}_i + (\boldsymbol{\beta}'\boldsymbol{X}_i) \times t_i + \varepsilon_i + \eta_i \times t_i. \end{aligned} \tag{3.7}$$

and

$$v_i = \underbrace{(\boldsymbol{\beta}'\boldsymbol{X}_i + \eta_i)}_{\beta_i(\cdot)} + \underbrace{(\boldsymbol{\psi}'\boldsymbol{X}_i + \zeta_i)}_{\psi_i(\cdot)}. \tag{3.8}$$

Effects under an uninformed principal An uninformed principal knows nothing about individual values of β_i, so she can only select which AEAs should be placed under the new technology at random. Using equation (3.7), given a scale of roll-out m (the share of AEAs to be treated), the total treatment effect on expected performance is

$$\int_{\boldsymbol{X}_i} ((E_{\varepsilon,\eta}(s^*|t=1, \boldsymbol{X}_i) - E_{\varepsilon,\eta}(s^*|t=0, \boldsymbol{X}_i))m) d\Xi(\boldsymbol{X}_i) = m\boldsymbol{\beta}'\bar{\boldsymbol{X}},$$

where Ξ is a cumulative distribution function describing variation in the vector $\boldsymbol{X}$, which is unobservable to a fully uninformed principal. This equation says that if no AEAs are treated, the total gains are zero. If all AEAs are treated, the total gains are equal to the average treatment effect of the intervention. If a partial measure $m \in (0, 1)$ is treated, the total gains are proportional to roll-out m, and the marginal impact of enhancing roll-out is always the average impact $\boldsymbol{\beta}'\bar{\boldsymbol{X}}$.

Effects under decentralization A supervisor observes each AEA's characteristics $(\boldsymbol{X}_i, \theta_i, \zeta_i)$, and selects AEAs to treat according to the value of the expected index $\mathbb{E}\{v_i|\boldsymbol{X}_i, \theta_i, \zeta_i\}$ as given by,

$$\begin{aligned} \mathbb{E}\{v_i|\boldsymbol{X}_i, \theta_i, \zeta_i\} &= \underbrace{(\boldsymbol{\beta}'\boldsymbol{X}_i + \mathbb{E}\{\eta_i|\boldsymbol{X}_i, \theta_i\})}_{\mathbb{E}\{\beta_i(\cdot)|\theta_i, \boldsymbol{X}_i\}} + \underbrace{(\boldsymbol{\psi}'\boldsymbol{X}_i + \zeta_i)}_{\psi_i(\cdot)} \\ &= \underbrace{(\boldsymbol{\beta}' + \boldsymbol{\psi}')}_{\boldsymbol{\Gamma}'} X_i + \underbrace{(\mathbb{E}\{\eta|\boldsymbol{X}_i, \theta_i, \zeta_i\} + \zeta_i)}_{u_i}. \end{aligned} \tag{3.9}$$

This equation is important for our linking the theory with the empirics of AEA selection by supervisors. The AEA observables in X_i matter both because they affect response to treatment (through β', as in the theory), but also because supervisors may have biases (through ψ'). Unobservables in u_i may also reflect components that affect response to treatment (through η) and biases (through ζ).

A key hurdle is that we do not have a direct measure of $\mathbb{E}\{v_i|\boldsymbol{X}_i, \theta_i, \zeta_i\}$, but we only observe the supervisor selection decision D_i^S. To recover $\boldsymbol{\Gamma}$, we further assume that η_i, ξ_i, and ζ_i are mean zero, normally distributed random variables with variances σ_η^2, σ_ξ^2, and σ_ζ^2, respectively. Given all of these distributional assumptions, the variable u_i can be characterized as drawn from Φ, a cumulative Normal $(0, \sigma_u^2 = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\xi^2}\sigma_\eta^2 + \sigma_\zeta^2)$ (this stems from the fact that the supervisor is Bayesian and updates her expectation of η upon observing θ). This, in turn, implies that D_i^S takes the familiar form of a probit model:[8]

$$\boldsymbol{Pr}\{D_i^S = 1|\boldsymbol{X}_i\} = \Phi\left(\frac{1}{\sigma_u}(\boldsymbol{\Gamma}'\boldsymbol{X}_i - c_p(m))\right)$$

Under these assumptions, standard arguments yield $\mathbb{E}\{\eta_i|u_i\} = \frac{\sigma_{\eta u}}{\sigma_u^2}u_i$, where $\sigma_{\eta u} = \left(\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\xi^2}\sigma_\eta^2\right)$. It follows that $\mathbb{E}\{\eta_i|D_i^S, \boldsymbol{X}_i, m\} = \frac{\sigma_{\eta u}}{\sigma_u}\frac{\phi(\frac{1}{\sigma_u}(\boldsymbol{\Gamma}'\boldsymbol{X}_i - c_p(m)))}{D_i^S - \Phi(\frac{1}{\sigma_u}(\boldsymbol{\Gamma}'\boldsymbol{X}_i - c_p(m)))} \equiv \frac{\sigma_{\eta u}}{\sigma_u}\lambda(D_i^S, X_i, m)$. Using these expressions after taking conditional expectations in equation (3.7), we get [9]

$$\begin{aligned}\mathbb{E}\{s_i^*|\boldsymbol{X}_i, D_i^S, t_i, m\} &= \boldsymbol{\mu}'\boldsymbol{X}_i + (\boldsymbol{\beta}'\boldsymbol{X}_i) \times t_i + \mathbb{E}\{\varepsilon_i|\boldsymbol{X_i}, D_i^S\} + \mathbb{E}\{\eta_i|\boldsymbol{X_i}, D_i^S, m\} \times t_i \\ &= \boldsymbol{\mu}'\boldsymbol{X}_i + (\boldsymbol{\beta}'\boldsymbol{X}_i) \times t_i + \frac{\sigma_{\eta u}}{\sigma_u}\lambda(D_i^S, X_i, m) \times t_i. \qquad (3.10)\end{aligned}$$

Note that ε_i and η_i are independent of $\boldsymbol{X}_i$ *by definition* and independent of t_i by way of the randomized experiment. Thus we can estimate equation (3.10) via a two-step procedure using OLS. The first step allows us to estimate the selection model that will yield $\lambda(D_i^S, X_i, m)$ and the second step yields estimates for the coefficients in equation (3.10).

Equation (3.10) is the crucial resource to estimate the marginal treatment impact of the intervention under different scenarios of decentralization and informational advantage. To see this, consider first the simplest case where neither the principal nor the supervisor can observe any AEA traits so the vector $\boldsymbol{X}_i$ is constant. The expected index on which the supervisor selects is $\mathbb{E}\{v_i|\theta_i, \zeta_i\} = \mathbb{E}\{\eta|\theta_i, \zeta_i\} + \zeta_i = u_i$. Given the 50 percent roll-out in the experiment, we know that under decentralization, the total treatment impact of 50 percent roll-out is $\beta_0 + \beta_1$ from OLS estimation of equation (3.5). In order to trace the marginal treatment impact at any other roll-out m, we only need to consult the value of u_i at the m percentile in the Normal distribution of u_i.[10] Thus, it is possible to trace the total treatment gain from following the supervisor's selection criterion for all m.

As the expression $\mathbb{E}\{\eta|\theta_i, \zeta_i\} + \zeta_i = u_i$ makes clear, we cannot tell whether a supervisor's selection is due to information on unobservables that affect true responsiveness to treatment

[8] In our estimation, c_p is not separately identified from the constant vector in $\boldsymbol{X}_i$ and thus we normalize it to zero. We revisit c_p in section 3.7.

[9] We do not impose any restrictions on u_i and ε_i and so also include $\lambda(D_i, X_i, m)$ as a main effect without any interaction with t_i. This parameter (along with $\boldsymbol{\mu}$) is not of direct interest to us and is not required for identification, but may improve the efficiency of the other estimates.

[10] We do not recover separate values for σ_u and σ_η, since all parameters are scaled by σ_u in the probit regression.

(η) as opposed to unobservables that make the supervisor select an AEA for other reasons (ζ). But if $\beta_1 > 0$ we know the supervisor gets a precise enough signal on η, and places enough weight on it, so that even if she is biased in her choices, her selection yields higher treatment impact than selecting the AEAs at random.

In most situations, supervisors will know characteristics of their AEAs, and so the expected index $\mathbb{E}\{v_i|\boldsymbol{X}_i, \theta_i, \zeta_i\}$ on which supervisors select will indeed be a function of $\boldsymbol{X}_i$. In this situation, each expansion of roll-out will imply extending treatment to new AEA types, where the type space as seen by the supervisor is some unidimensional path in a higher dimensional space of traits $\boldsymbol{X}_i$ and the supervisor-only observed u_i. The analyst does not observe u_i, but can form an expectation of it conditional on an AEA with traits $\boldsymbol{X}_i$ being selected. Knowing traits $\boldsymbol{X}_i$ and a conditional expectation on u_i for an AEA being selected at a given level of roll-out, equation (3.10) delivers the treatment impact. Thus, it is possible to derive the total treatment gain from following the supervisor's selection criterion for all m.

Further uses of the model: evaluating supervisors, and the potential for sophisticated centralization We have now described ways to obtain marginal treatment impacts at varying roll-outs for the cases of an uninformed centralized principal and an informed supervisor. But the selection model laid out in this section can be put to other uses. First, it is possible to evaluate the supervisors in a more complete way than simply saying whether they have an informational advantage over the principal. We can ask the extent to which their advantage is related to their knowledge of elements that are potentially observable to the principal ($\boldsymbol{X}_i$) versus things the principal cannot expect to learn (η). Moreover, the analyst can econometrically evaluate the extent to which supervisors make optimal use of observable data in $\boldsymbol{X}_i$.

Second, with gains in the ability to gather and process data, a principal could learn some traits of her AEAs, captured by $\boldsymbol{X}_i$. This opens up consideration to a new class of counterfactuals, with a natural one involving the marginal treatment impact for varying m for a decision maker that knows $\boldsymbol{X}_i$ but does not observe θ_i. Thus, we can ask whether a sophisticated centralized principal can emulate or surpass the performance of supervisors despite her informational disadvantage. We perform these exercises in Section 3.7.

Discussion We have presented a heterogeneous treatment effect model where supervisors have private information about the treatment effects. Equation (3.10) shares the same functional form as the "Heckit" selection model. However, in most settings where the Heckit is applied, $t_i = D_i^S$. In settings that mirror ours where D_i^S is assigned according to ε_i or η_i, inclusion of the $\lambda(\cdot)$ control function in estimation is required for identification because of non-random censoring of potential outcomes, the raison d'etre for the literature on selection correction. However, control functions require credible instruments; without an instrumental variable that could be excluded from one equation or the other, if one instead assumed a uniform distribution for u_i and then used OLS in the first stage for D_i^S), $\lambda(\cdot)$ would be collinear with the vector X_i (Olsen, 1980). Even in contexts where there are credible instruments that generate experimental variation, selection models have been used to extrapolate treatment effect heterogeneity among never-takers and always-takers from instrument-implied local average treatment effects (Heckman & Vytlacil, 2005; Kline & Walters, 2019).

In our context, however, t_i is independently and randomly assigned, and not equal to D_i^S. While supervisor preferences are elicited, they are not used to determine assignment in our main sample. This means that we neither have censored potential outcomes nor always-takers and never-takers. Instead, we have a randomized experiment with full compliance as well as information about supervisor preferences that were not implemented, and so we are able to credibly estimate treatment effects along the full distribution of η_i, an exercise that requires no extrapolation. Because we observe treatment effects for non-selected AEAs (i.e., those with $D_i^S = 0$), even if misspecified, $\lambda(D_i^S, X_i)$ is just a transformation of D_i^S and X_i , and with inclusion of controls, its independent variation is driven primarily by D_i^S.[11]

3.5 Data

We collected two main sources of data. The first is a survey of AEAs. Each AEA and supervisor independently filled out answers on a paper questionnaire with survey enumerators available to answer any questions. The survey contains questions regarding the AEAs' demographics and work history, the digit span test measuring cognitive ability, and the Big-5 inventory (John, Naumann, & Soto, 2008). We combine this five-dimensional inventory into two higher-order personality traits called stability and plasticity. Stability combines neuroticism, agreeableness, and conscientiousness and therefore keeps track of traits that have been found to predict earnings and job attainment, such as the tendency to remain emotionally stable and motivated and be organized and thorough. Plasticity, which aggregates extraversion and openness, is a measure of a person's gregariousness and openness to new experiences. These two meta-traits tend to account for much of the shared variance among the lower order dimensions (DeYoung, 2006).

The second source of data we have is two rounds of farmer phone surveys. We called farmers who were beneficiaries of the AEAs and asked questions about their interactions with the AEAs such as how often they saw the AEA and how satisfied they were with his work.

The timeline of events is as follows. In March of 2014, the ALAT-level supervisors chose which AEAs they would like to prioritize for receiving a phone with the objective of expanding effort in service to farmers. The first round of phones was distributed to the AEAs between April 30, 2014 and July 16, 2014. Individuals from the central ministry office traveled across the country to meet with the AEAs who were scheduled to receive phones, distribute the

[11] Thus, if one wanted to stick with OLS in the first stage rather than a Probit, while continuing to assume a linear conditional expectation function $\mathbb{E}\{\eta_i|u_i\} \propto u_i$, the coefficient on $\lambda(\cdot)$ in the second stage would be numerically equivalent to estimating an OLS regression in one step with $\frac{D_i^S}{2}$ in place of $\lambda(\cdot)$. This is a result that follows immediately from the Frisch-Waugh theorem. Its unique role in our context derives from its tagging supervisors' choices so that it may reflect their perceptions of how AEAs respond to treatment.

phones to them, and teach them how to use the phones. This process took over two months because it involved 19 meetings spread across the country.

After the first round of phones was distributed, we conducted two types of data collection. From July 7 through September 7, 2014 we conducted the first round of farmer phone surveys. Additionally, during September 2014, we conducted the survey of all AEAs as well as their supervisors. We treat AEA characteristics such as sex, age, years of education, and the personality indices as being fixed and not affected by the roll-out of the phones. On the other hand, we treat variables such as the AEAs' perceptions of whether their supervisors know where they are during the working week as potentially being affected by the roll-out of the phones. In the control group, those ALATs where no AEAs received phones, these responses should not be impacted by the roll-out of the phones.

After completing the first round of surveying, the second round of phones was distributed between February 10 and March 13, 2015. We then conducted a second round of farmer phone surveys between March 24 and May 7, 2015. The Ministry of Agriculture planned to distribute phones to all AEAs who had not yet received one before the end of 2015 but in the end did not do so.

The ministry did not give any phones to AEAs who were not on our randomized list. There were a few cases in which phones broke down or sick AEAs were not able to pick up their phones. For this reason we look at intent-to-treat (ITT) estimates using our initial random assignment.

In early 2014, we were given full information, including job title, job location, and client names and phone numbers for 368 agricultural extension agents - 139 supervisors and 229 AEAs. In late 2014, we were able to interview 301 of these – 119 supervisors and 182 AEAs. We interviewed 79% of the AEAs in our original administrative data, 15% no longer worked for DEAg, and 6% were absent the day of the surveying.

The job description of an AEA involves working with 80 farmers. Thus, it is no surprise that the median AEA in our data listed the names of 80 farmers with whom he worked; the mean of the distribution is 75 with a standard deviation of 26. The median AEA in our data listed phone numbers for 78% of the farmers he served, while the mean share listed is 73%. These numbers vary very little for AEAs versus supervisors.

In total, we called 2,635 farmers in the first round and 2,642 in the second round for the 182 AEAs who responded to the AEA survey.[12] Of those, 68% led to completed surveys.[13] Conditional on completing the survey, 70% of farmers confirmed that the AEA that had

[12]We conducted two rounds of farmer phone surveys, but we wanted to leave open the possibility of conducting three rounds. For AEAs and supervisors in multi-AEA ALATs who listed 75 or more farmer phone numbers, we randomly chose 75 farmers to call and then randomly divided them to call 25 farmers in each of three rounds. For those who listed fewer than 75 farmer phone numbers, we randomly divided their farmers into thirds to call in each of the three rounds. Similarly, for AEAs and supervisors in single-AEA ALATs who listed 24 or more farmer phone numbers, we randomly chose 24 farmers to call and then randomly divided them to call 8 farmers in each round. For those who listed fewer than 24 farmer phone numbers, we randomly divided their farmers into thirds to call in each of the tree rounds.

[13]In 18% of cases, we reached voicemail on all five tries, 7% of cases were wrong numbers, 4% were out-of service phone numbers, and 2% of farmers did not agree to complete the survey.

provided their number worked with them and thus were asked more detailed questions about their interaction with that AEA.[14] This leads to 2,519 completed farmer phone surveys.

Table 3.1 presents sample means and a randomization check of the cellphone assignment for various AEA characteristics. The table distinguishes between treated and control small single-AEA ALATs (columns 1 and 2) and treated and control large multi-AEA ALATs (columns 3 and 4). On average, AEAs are 37 years old, and 76% of them are male. The AEAs were able to recall an average of 5.2 digits in the memory digit span test, which is a commonly-used measure of cognitive ability.[15] AEAs are also required to travel on average 12 kms to visit a given farmer. Overall the results in Table 3.1 suggest that the treatment, which was randomized at the ALAT level, was done in a balanced way.

In Appendix Table B.1, we also check for balance on a set of ALAT-level characteristics extracted from the population and agricultural censuses, as well as the 2013 presidential elections. We look at 18 comparisons, and only one shows significant imbalance across treatment and control. The results in the table again suggest that the randomization led to balance across treatment and control. Across a different dimension, he most noticeable difference between small and large ALATs is that large ALATs are located in districts with larger populations, and a lower rural population share.

3.6 Results

In this section, we begin by estimating the impact of the cell phones on AEA performance. According to the model, under certain conditions (cell phones improve monitoring and there are sufficiently many AEAs who respond positively to it), the increase in monitoring induced by the phones should boost the effort levels of the AEAs and thus increase the number of farmers visited. Subsequently, we test whether the impact of cell phones was higher among the AEAs who were selected by the supervisors, which would be the case if supervisors were able and chose to target the AEAs with highest responsiveness to treatment. Next we look for, but fail to find, evidence of spillovers and Hawthorne effects. Finally, we estimate heterogeneous treatment effects, which we use to evaluate impacts under various counterfactual scenarios with different scales of roll-out.

Increased Monitoring and Performance

As we discussed in Section 3.2, the primary task of an AEA is to visit farmers. In columns (1) through (5) of Table 3.2, we estimate the impact of the phone on whether the farmer

[14] We first asked the farmers to talk about any AEAs with whom they worked and did not offer up the name of the AEA we had on record for them. We only asked the farmer about the specific name we had on record if either the farmer worked with an AEA whose name he couldn't remember or if he did not list the name of the AEA we had on record on his own.

[15] For the digit span test, the enumerator read out loud a random number that the AEA was then required to recite back. The test began with a number that was two digits long and then increased incrementally in the number of digits until the AEA could no longer recall a number correctly on both of two chances.

reported having been visited by his AEA in the last week. In columns (1) through (3), our estimation sample includes all AEAs in the small and large ALATs, excluding those randomized into the partial treatment cells (cells E and F). In column (1), we present the estimates without any additional controls. In column (2), we add a set of basic controls (e.g., age and gender), and in column (3), we further augment the specification to include controls measuring AEA personality type (e.g., Big 5 meta-traits and Digit Span). In column (4), we re-estimate the specification presented in column (3), excluding the single-AEA ALATs.

We find that the increase in monitoring leads AEAs to visit their farmers more often. They are approximately 6 percentage points more likely to have visited a given farmer in the past week, which is an increase of 22% over the control group. As expected given the random assignment, the estimated impact is robust across the various specifications, and when we restrict the estimation to only multi-AEA ALATs, which will be our main sample moving forward. The bottom row of the table shows randomization inference p-values which help account for the small number of ALATs under study, and the coefficient on treatment status retains its significance. Overall, the demographic and personality-based controls have little predictive power.[16]

Supervisors are in charge of both supervising the AEAs in their ALAT as well as serving their own farmers. In column (5), we test the impact of the phone on the visits to those farmers who are served by a supervisor. We find a small and insignificant impact (point estimate = -0.008; clustered standard error = 0.036). This suggests that the impact of the phone is related to the greater monitoring ability it gives supervisors and not due to productivity-enhancing functions of the phone (e.g., ease in communication), which would have the same effect on both supervisors and AEAs. As a further check, AEAs were asked whether they agreed with the statement that their supervisor usually knows where they are during the work week. In column (6), we see that having a phone significantly increased the extent to which AEAs agreed with this statement.

While the treatment led to more visits, this does not necessarily imply that the AEAs are exerting more effort. AEAs could be making more visits but making them shorter. In column (7), we test for this possibility but do not find evidence to support the idea. The point estimate, which suggests that treated AEAs spend only 1.6 percent less time on each visit, or approximately one and a half minutes, is small and statistically insignificant.

In Appendix Table B.2, we examine the effects of the treatment on other dimensions of AEA performance. We consider four additional measures: 1) how satisfied the farmer is with the AEA (1=very, 2=somewhat, 3=not at all); 2) an indicator for whether the farmer thought the AEA conducted helpful training sessions; 3) an indicator for whether the farmer did not find the AEA helpful at all; 4) and the first principal component for the three measures, with higher values indicating worse performance. All four measures are significantly correlated with AEA visits (see Appendix Table B.3). Farmers who receive more visits from their AEAs are also more likely to think the AEA conducts helpful training

[16] In results not shown here we look separately at short-run versus long-run impacts of the phones, and find that they are quite similar. The impact of the phones does not diminish over time.

sessions, are more satisfied with their AEAs, and more likely to find the AEA helpful in some way. In general, we find that the additional monitoring improved performance along these dimensions as well, although the estimates are measured with less precision. Based on our principal component measure, the treatment improved aggregate performance by 0.14 of a standard deviation (standard error = 0.07).

Do Supervisors Have Useful Information?

Recall that our model assumes AEAs differ in their responsiveness to enhanced monitoring and that supervisors know this information. If supervisors wish to increase the number of farmers visited, then when tasked with the responsibility of assigning increased monitoring, they should target the AEAs for whom a larger increase in performance ought to be expected. Our research design allows us to test this directly.

Prior to the randomization, supervisors identified which half of their AEAs they believed should receive the phones. Given these selections, we test for the value of information using a simple difference-in-differences estimator for our sample of large ALATs. We compare the performance of AEAs who were selected and received the phone against those who were selected but did not receive the phone, net of the difference in performance between those who were not selected and received the phone against those who were not selected and did not receive the phone (i.e., $(B - A) - (D - C)$).

From Table 3.3, we see that the effects of the phones on performance are entirely driven by the effects on the AEAs prioritized to receive the phone prior to the randomization. These AEAs increased the share of farmers visited in the last week by approximately 15 percentage points. Compared to the prioritized AEAs in the control group, this effect represents a substantial increase of 54 percent. From column (2), we also see that prioritized AEAs in the control group are 3.3 percentage points less likely to have visited their farmers relative to the non-selected. This potentially suggests that supervisors target the least productive AEAs, although this difference is not statistically significant.

Spillover and Hawthorne Effects The theory we laid out in Section 3.3 assumes the absence of spillovers across agents. An advantage of our experimental design is that we can test this assumption directly. Recall from Figure 3.2 that for the ALATs assigned to the 50% coverage treatment, only the selected AEAs (cell E) received phones in the first wave. Therefore we can test for the presence of spillover effects by comparing the performance of the AEAs in cell F (i.e., non-selected AEAs who did not receive a cell phone but whose selected coworkers did) to the AEAs in cell C (i.e., the non-selected in the control group who, like their selected coworkers, did not receive a phone).

We do this in column (5) of Table 3.3. Here we focus on the first wave of the survey and add two new controls to the original specification presented in column (4). These are indicators for the AEAs in the ALATs where only the selected received phones (cells E and F, labeled *Treat 50*) and an interaction of this variable with whether the AEA was selected and thus received a phone in the first wave (cell E only, labeled *Treat 50* $\times$ *Selected*).[17] The excluded group continues to be the non-selected in the control group (cell C).

The non-selected do not appear to perform differently whether or not their selected coworkers are treated with phones. The point estimate comparing the AEAs in cell F to cell C is small at 0.0138 and not statistically significant. This finding suggests the absence of spillovers, either directly in the form of peer effects, or through effort substitution by supervisors. Meanwhile in the 50% coverage arm, the treatment effect on those who were selected is similar in magnitude (0.133) to the treatment effect for the selected in the 100% coverage arm (0.151). This is again consistent with our interpretation that supervisors can supervise all AEAs with phones regardless of what share of their AEAs have phones, and that supervisors possess useful information about their AEAs.

An alternative story, akin to a Hawthorne effect, is that supervisors may have increased their supervision effort on those treated AEAs who they had selected. They might do this to signal that their selection was judicious, making sure their selected AEAs performed better under the new monitoring technology. In this case, our findings would not necessarily reflect supervisors' informational advantage but rather their additional effort.

There are three reasons we find this explanation unlikely. First, an ancillary test points to the absence of Hawthorne effects. We exploit the fact that some ALATs are small (a single AEA), and thus supervisors were not asked to identify which half of the AEAs ought to be treated. If Hawthorne effects are present, the average treatment effect in multi-AEA ALATs should be stronger than in the single-AEA ALATs where no selection was made, as the signaling motive is absent in the latter. We estimate the effects of the phone distinguishing between single and multi-AEA ALATs. As we see in column (1) of Table 3.4, we cannot reject the hypothesis that the treatment effects are similar across small and big ALATs. If anything, the point estimate of the interaction term is positive, suggesting that the treatment effect might be slightly larger in the single-AEA ALATs. This test assumes that the propensity to have been "selected" is balanced across small and large ALATs. In column (2) we test whether AEAs are systematically different across the two types of ALATs. Based on our rich set of observables, we do not find any evidence of AEAs sorting by ALAT size.

The data speak against the potential for Hawthorne effects in a second way. If Hawthorne effects were present, then we might expect them to be more pronounced in the 50% coverage treatment arm (where supervisors' selection rule was enforced) relative to the 100% coverage treatment. But as we noted from the results presented in column (5) of Table 3.3, the effects for the selected AEAs across the two treatment arms are similar, with the effect in cell B (selected AEAs in ALATs where all are treated) being 0.133 and cell F (selected AEAs in ALATs where only the selected are treated) being 0.151.[18]

[17] Unfortunately, after the randomization we discovered that some of the AEA characteristics were not balanced between the 50% coverage treatment arm and the no treatment and full treatment arms. Although we do not find evidence of spillover effects, our point estimates are sensitive to control variables.

[18] Alternatively, one could conjecture that the Hawthorne effects would be less pronounced in the 50% coverage treatment arm if, for instance, supervisors in the 100% coverage treatment felt slighted. But again, the fact that the two treatment effects are similar in magnitude rejects this possibility.

Third, the structure of incentives in the field makes Hawthorne effects unlikely. Supervisors had very little incentive, if any, to prove to the Ministry that they could select the most responsive AEAs. They had no reason to think that the Ministry would or even could evaluate their selection process, let alone reward or punish supervisors for the selections made. Those beliefs would have been correct since the Ministry never performed, nor planned to perform, such an evaluation. Given the lack of incentives, it seems unlikely that supervisors would exert costly effort to make sure the selected AEAs who were treated performed better.

In sum, we find strong evidence that the phones increase AEA effort and that supervisors possess useful information regarding which AEAs' performance will improve most after receiving a phone. This of course begs the question of what characteristics the supervisors used to create their prioritized list and the extent to which supervisors used information on characteristics analysts could hope to obtain. The next subsection answers these questions.

Heterogeneous Treatment Effects

In Table 3.5, we present estimates from a Probit regression, in which the dependent variable is an indicator for whether the AEA was prioritized by the local supervisor. Based on standard observable characteristics, we find that supervisors tended to prioritize AEAs who were younger, married, and had to travel further distances to visit their farmers (although this last characteristic is only significant at an 88 percent level of confidence). In terms of their personality traits, supervisors were more likely to select AEAs with lower levels of the Big-5 Stability meta-trait. Individuals with higher stability scores may be more likely to stay motivated and have better relationships with their supervisors.

Interestingly, we also find that supervisors of the large ALATs, who except for one supervisor are all registered with the incumbent political party, are significantly less likely to place AEAs who are registered with the incumbent party under increased monitoring. This suggests that either supervisors are acting non-benevolently or, as we will subsequently test, that party affiliation serves as a marker for those who are less likely to respond to treatment.

Despite the richness of our data, our ability to predict the choices of the supervisors is fairly low: the highest pseudo R^2 is only 18.5%. This opens the possibility that supervisors are also selecting AEAs based on unobservable but productive characteristics (η) or unobservable and idiosyncratic characteristics (ψ), features that are not captured by demographic traits or even indicators of cognitive and non-cognitive ability. Ultimately, the only way to determine whether there could be an advantage to decentralization is to rely on our experimental design, and ask whether supervisors select AEAs who will be more responsive to treatment.

In Table 3.6, we present a series of second stage estimates based on Equation 3.10. In column (1), we present a specification without any additional controls or interaction terms, whereas in columns (2) and (3) we include additional controls along with their interactions with the treatment indicator. For columns (2) and (3), the first stage regressions correspond to the ones presented in Table 3.5.

The key finding in Table 3.6 concerns the inverse Mills ratio and particularly its interaction with treatment. The inverse Mills ratio captures the expected unobservable traits that recommended an AEA for selection by the supervisor. Because no controls were included in column (1), the coefficient on the inverse Mills ratio interacted with treatment in that column replicates the findings from Table 3.3 that supervisors are selecting individuals with higher treatment effects.

When we allow the effects of the treatment to vary by the characteristics that we found were predictive of the likelihood of selection (columns (2) and (3)), we find that the inverse Mills ratio is still highly predictive of responsiveness to treatment, direct evidence that the unobservable reasons supervisors are selecting AEAs are productive rather than non-germane. In addition to the unobservable traits, the treatment effect also varies by the cognitive ability of the AEAs; those who performed worse on the digit span test exerted more additional effort in response to the treatment. Moreover, in column (2) it appears that members of the incumbent party respond less to the treatment. But, once we account for the differential effects of cognitive and non-cognitive ability in column (3), this effect goes away. This suggests that non-benevolent motives may have influenced the supervisors' targeting.

The results so far suggest several questions. What is the basis for the supervisors' informational advantage? At a given cost of decentralization, does the informational advantage justify the cost? To answer these questions, one needs to know two elements. First, what is the scale of roll-out anticipated. Under decentralization, the anticipated scale of roll-out should be whatever is optimal, and this motivates the need to identify that optimal level. Second, how much information does the central authority have? In the next section, we apply the framework introduced in Section 3.4 to provide answers to these questions.

3.7 Counterfactuals

In this section, we exploit our heterogeneous treatment effects model to compute counterfactual treatment effects under alternative selection rules. This allows us to assess the benefits of decentralization relative to centralization under different informational assumptions.

The first step is to define a counterfactual aggregate benefit under an arbitrary selection rule D_i^{CF} as:

$$\begin{aligned}\Delta Y^{CF} &= \mathbb{E}\{\underbrace{\beta_i(X_i,\eta_i)}_{\text{how much?}} \times \underbrace{D_i^{CF}}_{\text{who?}}\} \\ &= \int \mathbb{E}\{\beta(X_i,\eta_i)|D_i^{CF}=1\}\boldsymbol{Pr}\{D_i^{CF}=1\}d\boldsymbol{X}_i \qquad (3.11)\end{aligned}$$

In keeping with the rest of our notation, we write our arbitrary selection rule as a threshold problem, $D_i^{CF}(X_i,u_i) = \mathbf{1}[\tilde{\boldsymbol{\Gamma}}'\boldsymbol{X}_i + \tilde{u}_i \geq c_p]$; because we have not made any distributional assumptions about $\tilde{u}_i$, this does not impose additional assumptions. Note that the assumed cost c_p is not directly observable, and the threshold problem is not a unique representation of the selection rule—any monotonic transformation of the latent index and c_p will yield the same choices. However, we are not trying to directly obtain either of these objects: only the consequences $\boldsymbol{Pr}\{D_i^{CF} = 1|X_i, c_p\}$ and $\mathbb{E}\{\beta(X_i, \eta_i)|D_i^{CF}\}$, which map into the scale of roll-out m and the aggregate counterfactual impact ΔY^{CF}.

One example of a selection rule is the one implicitly applied by supervisors, $D_i^S(X_i, u_i)$, which anchors our portrait of what can be achieved under decentralization. Note that from our estimation of Equation 3.10, we have recovered $\mathbb{E}\{\beta_i(\cdot)|X_i, u_i\}$, and under distributional assumptions, the selection rule under decentralization, $D_i^S(X_i, u_i)$. Given this, we can use Equation 3.11 to trace out the expected treatment effects of the cell phones under decentralization for any given threshold c_p or, by extension, any scale of roll-out m. But, we can impose any other selection rule capturing different counterfactual scenarios corresponding to different forms of centralized assignment and trace out the expected treatment effects for all roll-out levels in each scenario.

Uninformed Principal A natural, if extreme, benchmark is that of a principal who does not have any information about how best to target roll-out. In this situation, the selection rule is random allocation. At a roll-out level m, a fraction m of all AEAs receives a cell phone, and the expected total treatment effect is $m\%$ of the average treatment effect (considering, as in the theory section, a large number of AEAs who can then be approximated by a continuum).

The dotted line in Figure 3.3 plots this counterfactual selection rule at various roll-out levels. For instance if the principal decided to allocate the phones to everyone, then the expected aggregate treatment of the program would be 6.4 percentage points, which corresponds to the average treatment effect in column (3) in Table 3.6. If instead she decided to treat only half of the AEAs, then we would expect an aggregate treatment effect of only 3.2 percentage points. Thus, it is easy to see that with a random selection rule, we get a set of counterfactuals that traces a straight line from zero to the average treatment effect. In Table 3.7, we present our estimated treatment effects at different roll-out levels for the various allocation rules we consider. The number displayed in bold represents the largest treatment effect under a given allocation rule.

Supervisor We can contrast the random allocation rule with the aggregate benefits based on the supervisor's selection rule. In this case, the selection rule is given by $\boldsymbol{Pr}\{D_i^S = 1|\boldsymbol{X}_i\} = \Phi(\frac{1}{\sigma_u}(\boldsymbol{\Gamma}'\boldsymbol{X}_i - c_p))$ and the expected aggregate treatment effect is $\Delta\mathbb{E}\{s_i^*|\boldsymbol{X}_i, D_i^S, T_i\} = \boldsymbol{\beta}'\boldsymbol{X}_i + \frac{\sigma_{\eta u}}{\sigma_u}\lambda(D_i^S, X_i)$. This counterfactual is depicted in Figure 3.3 with the solid line. Note that by construction, the curve must cross three points: the origin, 0.064 at 100% roll-out, and 0.070 at 53.8% roll-out which corresponds to the share of AEAs that received the phones under the actual research design.

The difference between the supervisor counterfactual and the random allocation rule measures the benefits of decentralization at each level of roll-out under the assumption that the principal does not possess any information. As we can see from the figure, the difference between the random allocation and the supervisor rule is maximized at a roll-out threshold of 53% where the additional treatment effect is over 3.5 percentage points. The optimal scale of roll-out under decentralization is not 53%, however, but 77%, at which level the total treatment effect is 7.7 percentage points. The total treatment effect starts to decline at a roll-out scale of 77% as we begin to assign the treatment to individuals for whom the treatment effect is negative. The existence of individuals for whom the treatment effect is negative is consistent not only with our model (as ρ is allowed to be negative), but with the findings in de Rochambeau (2017), who showed that the introduction of a new monitoring device for truck drivers in Liberia lowered the productivity of the intrinsically motivated.

What underlies the informational advantage of the supervisor over an uninformed principal? One way to tackle this question is to ask how much of the supervisor's advantage is predicated on the use of information on observables $\boldsymbol{X}_i$ versus information on unobservables η_i. The dot-dash line in Figure 3.3 traces out the counterfactual treatment effect under the assumption that the supervisor does not use his signal of η. In other words, the dot-dash line tells us what the treatment effects would be under a supervisor who cannot use information on unobservables. In this case, the selection rule and expected treatments are only computed based on the observable (to the econometrician) traits, setting $\lambda = 0$. The dot-dash curve is much closer to the one under random assignment. This suggests that, in our setting, most of the supervisor's informational advantage is driven by access to information that is likely hard to collect for a centralized authority lacking personal contact with the AEAs.

Giving Centralization A Chance: Counterfactual Treatment Effects With A Partially Informed Principal

Minimally informed principal: Assignment based on distance traveled Thus far, we have assumed that the principal does not have any prior information about how AEAs will respond to the program which, though extreme, may not be a wholly unreasonable approximation to the situation facing the leadership of government programs in low state capacity contexts. This does not suggest however that adopting a sensible heuristic might not outperform a random assignment mechanism, which would of course affect the centralization versus decentralization calculus. One such heuristic might be to simply allocate the phones to the AEAs who have to travel the farthest in order to visit their farmers. This requires some information on the work environment of AEAs, and it constitutes the case we associate with a minimally informed principal. This counterfactual is displayed in Figure 3.4 with a dashed light gray line. We find that this method generally outperforms random assignment (a 2.0 p.p. advantage at 50 percent coverage), but it cannot beat the supervisor at any roll-out level.

Significantly informed principal: Assignment based on predicted baseline performance We consider a second type of partially informed principal that has the capacity to gather information on individual AEA characteristics, and can map them onto their baseline productivity. To this end, we run a simple prediction model in which, among the AEAs in the control ALATs, we regress the share of farmers visited on our set of basic and cognitive controls (see Appendix Table B.4 for the estimation results). Using the estimated coefficients, we can then compute an AEA's expected productivity based on his or her observable traits. Given this information, a sensible centralized policy would be to assign cell phones starting with the AEAs who had the lowest predicted productivity, and as roll-out increases, expand coverage to AEAs with higher productivity. As we see in Figure 3.4, under this approach centralization would dominate decentralization at virtually all levels of roll-out. It is worth noting that the data requirements to estimate our performance-prediction model are not trivial and often beyond the capacity of government programs in several developing countries.

A sophisticated principal: Experimentation and assignment based on response to treatment For all its data demands, the approach in the previous section that assigns phones based on baseline performance prediction does not exhaust the possibilities open to a central authority who has the capacity to gather and analyze data. The key shortcoming of that approach is that baseline performance is not always a great predictor of responsiveness to treatment. While baseline performance can reflect individual heterogeneity in, say, linear terms of the effort cost function, response to treatment depends on other cost drivers, such as the disutility from receiving a reprimand.

To overcome these difficulties, a sophisticated principal can conduct a pilot experiment at a low roll-out level and establish a map between AEA observable characteristics and response to treatment. Then it is possible to construct an assignment rule $D_i^{CF}(X_i)$ that allocates phones starting with those AEAs who are predicted to have the highest response to treatment and work downwards to treat progressively less responsive AEAs, tracing out the total treatment effect for each roll-out level. Note that we are privileging principals in the "sophisticated" case because they would need to have digit span and Big-5 measures for all of their workers, which may be as much of a data constraint as running the pilot RCT.

As shown in Figure 3.4, this approach outperforms all others by a wide margin. The largest gap relative to the decentralized supervisor-choice approach is above 1.7 percentage points and occurs at a roll-out level of 38.4%. A sophisticated principal would be more interested in setting roll-out at its optimal scale: the maximum total treatment effect for an 'experimenting principal' is 9.0 percentage points and is achieved at a roll-out of 70%.

Note that relative to "blind centralization," which treats everyone and attains a total treatment effect of 6 percentage points, this arrangement saves on almost a third of the phones and attains almost 1.5 times the total treatment impact. Relative to the decentralized supervisor choice, the sophisticated centralized approach distributes roughly 10% fewer phones and attains roughly 1.3 additional percentage points in total treatment effect.

3.8 Conclusions

One of the primary benefits of decentralization is that mid-level supervisors are presumably better informed than their principals about how to implement a particular task. But the importance of this superior information will often depend on the scale of the task at hand. Because decentralization is costly, the decision to devolve decision-making powers to supervisors requires knowing not only the value of their information, but this value at different scales of roll-out. Despite the fact that the informational advantage of middle managers is a maintained assumption in much principal-agent theory, evidence of the presence and extent of that advantage has been scarce. We also have little evidence on the effects of decentralization in a context in which the scale of implementation affects the average treatment effects.

In this paper, we establish that middle managers in the government hierarchy do have information which can improve targeting of an intervention. We develop an approach to trace out the total treatment effects of the intervention at all levels of roll-out. The context is an initiative by the federal government in Paraguay to introduce a new monitoring device that enables supervisors in rural areas to track their agricultural extension agents.

Our experimental design randomly assigned monitoring devices across AEAs and independently elicited the preferences of their supervisors as to which AEAs should be prioritized for monitoring. Crucially, in the main sample, treatment assignment was kept independent of supervisor recommendations. This allows us to establish that supervisors have valuable knowledge because the AEAs selected by them are far more responsive to treatment. We find that the informational advantage of supervisors is tied to information other than observables that analysts might reasonably collect, and argue theoretically that the value of this information advantage varies with the scale of anticipated roll-out for the new technology. In addition, we estimate the full schedule of marginal treatment effects as roll-out scale is expanded from 0 to 100 percent. We do this for the selection rule that supervisors are seen to have used as well as several other counterfactual assignment rules.

Our counterfactual assignment rules approximate what principals with varying levels of information might achieve when targeting AEAs for treatment in a centralized fashion. In our setting, impacts resulting from treatment decided upon by a minimally informed principal are not as high as those attained under decentralization. However, a reasonably well-informed principal can approach the level of impact of decentralization. And in the best case scenario for centralization, a principal who can conduct a pilot RCT to obtain predictors of individual response to treatment can outperform supervisor choices; such a principal would substantially reduce the roll-out scale and still attain larger aggregate gains in AEA performance.

Overall our findings suggest that as information and communication technologies continue to improve the capabilities of government and organizations more generally, the informational benefits that lower level agents bring become less clear. Although studies have shown that innovation in information technologies can lead to more decentralization (Bloom, Garicano, Sadun, & Van Reenen, 2014; Bresnahan, Brynjolfsson, & Hitt, 2002), our findings suggest the opposite may occur, particularly if these technologies primarily serve to reduce the information gap between principals and agents (or middle managers such as supervisors).

Of course, the value of the information that supervisors possess is specific to the task and context, which may raise concerns of external validity. But while our findings may not be generalizable, our method is, as it can be easily exported to other settings. Our approach is designed for settings in which spillovers across treatment units are minimal. Thus, it would be interesting to extend our framework to incorporate the potential effects of spillovers in the calculus to decentralize. We view this as a potential avenue for future research.

3.9 Figures

Figure 3.1: Treatment effects, roll-out extent, and the value of information

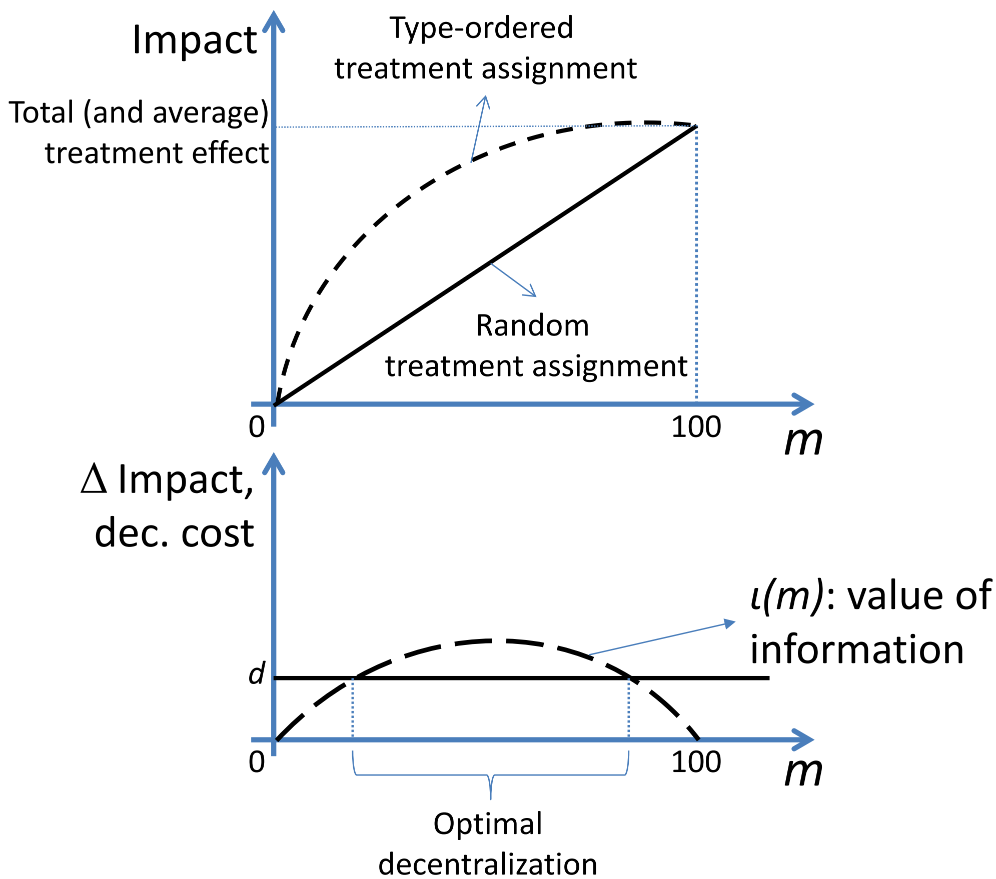
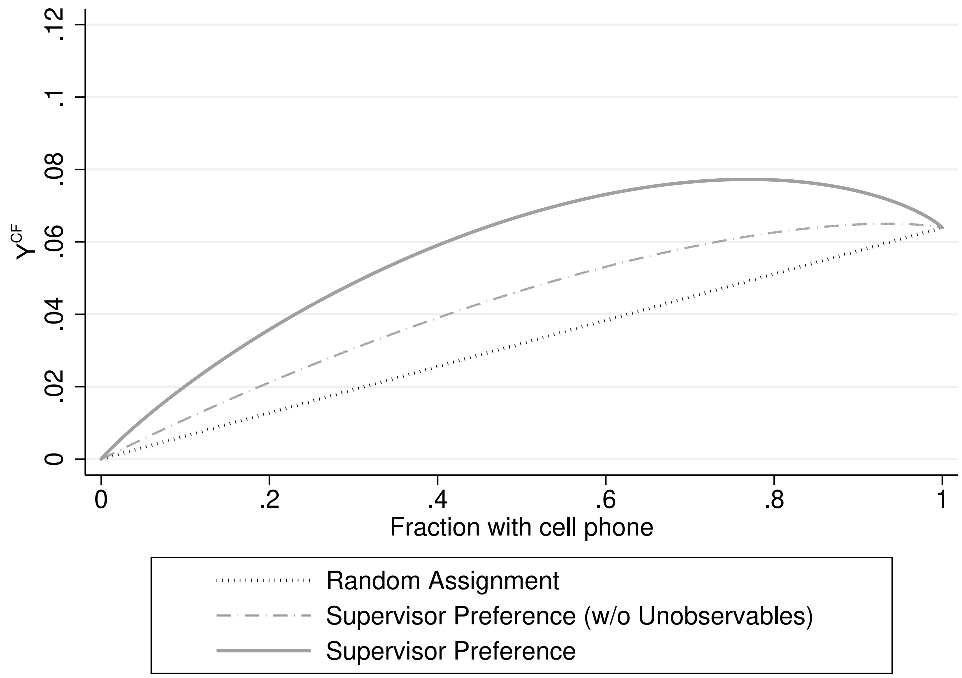


Figure 3.2: Experimental Design

	ALATs		
	Control Group	100% Coverage	50% Coverage
Selected AEA	A	B	E
Not selected AEA	C	D	F

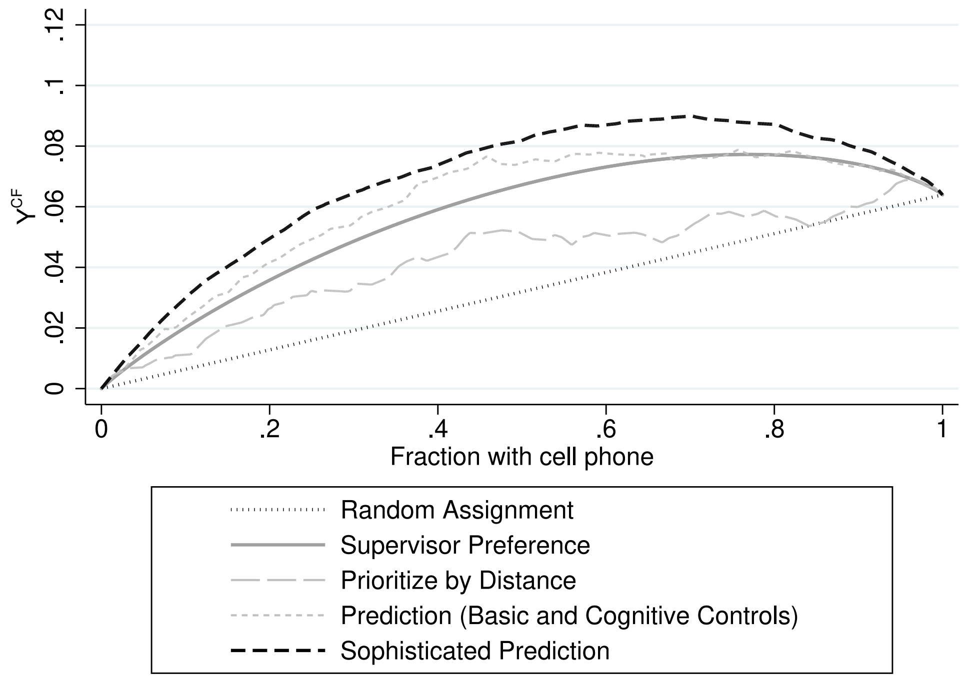
The columns correspond to ALATs, and the rows correspond to AEAs. For the ALATs in the column labeled Coverage 100% (cells B and D), every AEA received a cell phone independently of whether or not they had been selected by their supervisor. For the ALATs in the column labeled Coverage %50 (cells E and F), only the selected AEAs received cell phones (i.e. those in cell E). The control group contains cells A and C, where none of the AEAs received a cell phone.

Figure 3.3: Supervisor versus Random Assignment



The y-axis shows the total treatment effect at different scales of roll-out under different assignment rules. With supervisor preference the treatment assignment is what would be achieved under decentralization if supervisors made the assignment decision based on all the information they had; supervisor preference without unobservables refers to the case in which supervisors made their decision based purely on the observable AEA characteristics; and under random assignment the treatment assignment is made randomly.

Figure 3.4: Supervisor versus Alternative Allocation Rules



The y-axis shows the total treatment effect at different scales of roll-out under different assignment rules. Under random assignment the treatment assignment is made randomly; with supervisor preference the treatment assignment is what would be achieved under decentralization if supervisors made the assignment decision based on all the information they had; prioritize by distance is what would happen if treatment were assigned first to those AEAs whose beneficiaries live further from the local ALAT office; prediction (basic and cognitive controls) uses the control group to predict baseline performance using the observable variables and then treats first those AEAs who are predicted to be the worst performers in the baseline; sophisticated prediction runs a pilot experiment at low roll-out to establish a map between treatment response and observables and then treats first those AEAs who are predicted to have the highest treatment response.

3.10 Tables

Table 3.1: Covariate Balance Across AEAs

	Small ALATs		Large ALATs	
	(1) Control	(2) Difference (T-C)	(3) Control	(4) Difference (T-C)
Male	0.611	0.139	0.750	-0.135
	[0.502]	{0.326}	[0.436]	{0.387}
Age	36.889	0.153	37.838	4.393
	[11.386]	{0.964}	[10.815]	{0.109}
Married	0.278	0.056	0.441	0.059
	[0.461]	{0.702}	[0.500]	{0.540}
Average Distance	10.425	3.156	12.678	-1.355
	[8.410]	{0.302}	[9.206]	{0.442}
Incumbent Party	0.611	0.014	0.559	-0.020
	[0.502]	{0.934}	[0.500]	{0.903}
Digit Span	5.333	0.333	5.191	0.270
	[0.840]	{0.251}	[1.069]	{0.339}
Big 5 — Stability	-0.083	0.345	-0.057	0.112
	[1.126]	{0.295}	[1.126]	{0.664}
Big 5 — Plasticity	-0.496	0.713*	-0.140	0.325
	[1.208]	{0.050}	[1.111]	{0.137}
Selected			0.603	-0.026
			[0.493]	{0.689}
Number of AEAs	18	24	68	26
Number of ALATs	17	23	22	11
p-value from Joint Test		0.644		0.425

"Control" and "Treatment" for small ALATs refer to ALATs that received cell phones in round 3 and rounds 1 and 2, respectively. The number of AEAs and ALATs in columns (1) and (3) correspond to the respective numbers in the control group. Those in columns (2) and (4) correspond to the respective numbers in the treatment group. The fraction of selected AEAs exceeds 50% because when ALATs had an odd number of AEAs, supervisors were told to round up. The joint test in the bottom row runs a regression of treatment assignment on all listed covariates. The joint test p-value is from a wild bootstrapped F-test imposing the null hypothesis that all coefficients equal zero. Standard deviations reported in square brackets and p-values from a Wu, 1986 wild bootstrap procedure with 100,000 replication draws reported in curly braces. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table 3.2: Average Effects of Cell Phones on Productivity

	Farmer was visited in the last week					Supervisor knows	Log length of meeting (mins)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.066**	0.063**	0.063**	0.057*	-0.008	0.185*	-0.010
	(0.031)	(0.026)	(0.026)	(0.030)	(0.035)	(0.103)	(0.059)
	{0.029}	{0.023}	{0.023}	{0.091}	{0.844}	{0.063}	{0.865}
Selected	0.011	0.015	0.017	0.011		-0.080	-0.034
	(0.033)	(0.027)	(0.029)	(0.029)		(0.138)	(0.036)
	{0.743}	{0.614}	{0.605}	{0.728}		{0.536}	{0.362}
Male		0.039	0.039	0.032	-0.029	0.301*	-0.047
		(0.027)	(0.030)	(0.033)	(0.046)	(0.147)	(0.042)
		{0.222}	{0.283}	{0.484}	{0.543}	{0.050}	{0.310}
Age		0.003*	0.003*	0.003	-0.000	-0.008	0.004**
		(0.001)	(0.001)	(0.002)	(0.002)	(0.008)	(0.002)
		{0.055}	{0.077}	{0.138}	{0.942}	{0.361}	{0.035}
Married		-0.004	-0.003	0.020	-0.041	0.014	-0.053
		(0.030)	(0.030)	(0.031)	(0.034)	(0.141)	(0.042)
		{0.878}	{0.918}	{0.560}	{0.266}	{0.924}	{0.257}
Distance to Farmers (log)		0.039	0.037	0.032	-0.069**	-0.139	-0.019
		(0.027)	(0.028)	(0.034)	(0.028)	(0.119)	(0.047)
		{0.187}	{0.239}	{0.445}	{0.028}	{0.250}	{0.713}
Incumbent Party		-0.022	-0.021	-0.040	0.058	0.254*	0.002
		(0.024)	(0.025)	(0.025)	(0.057)	(0.151)	(0.052)
		{0.404}	{0.440}	{0.149}	{0.362}	{0.092}	{0.971}
Digit Span			0.010	0.013	0.030	0.051	0.013
			(0.012)	(0.013)	(0.018)	(0.052)	(0.018)
			{0.440}	{0.371}	{0.153}	{0.355}	{0.557}
Big 5 — Stability			0.012	0.011	0.035	0.141**	0.018
			(0.015)	(0.016)	(0.020)	(0.050)	(0.024)

Continued on next page

Table 3.2 – *Continued from previous page*

	Farmer was visited in the last week					Supervisor knows	Log length of meeting (mins)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			{0.517}	{0.590}	{0.108}	{0.011}	{0.540}
Big 5 — Plasticity			-0.011	-0.011	-0.024	0.111*	-0.018
			(0.010)	(0.010)	(0.017)	(0.064)	(0.017)
			{0.357}	{0.322}	{0.210}	{0.079}	{0.313}
Servicer	AEA	AEA	AEA	AEA	Supervisor	AEA	AEA
Mean of Control Dep. Var	.271	.271	.271	.274	.308	4.593	4.414
R^2	0.005	0.012	0.013	0.012	0.032	0.199	0.012
Number of Phone Surveys	1842	1842	1842	1584	1173		1819
Number of AEAs	136	136	136	94	107	126	136
Number of ALATs	71	71	71	33	107	65	71
RI p-value on **Treated**	.013	.007	.007	.047	.843	.198	.903
Includes Small ALATs	✓	✓	✓		✓	✓	✓

Big 5 stability and plasticity measures normalized to have mean zero and unit variance. Outcomes in all columns other than (6) are from the farmer phone survey. These regressions include unreported indicators for small ALATs, survey wave, and an interaction of the two. Outcome in column (6) is from the AEA survey and is from a regression of a five-value index ranging from (1) "Strongly Disagree" to (5) "Strongly Agree." Cluster robust standard errors and p-values from a Wu, 1986 wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values. Randomization inference (RI) p-values are based on 1000 replication draws using Stata's ritest command (see Heß, 2017).

Table 3.3: Do Supervisors Have an Informational Advantage?

	Farmer was visited in the last week				
	(1)	(2)	(3)	(4)	(5)
Treat	0.0607*	-0.0233	-0.0397	-0.0361	-0.0127
	(0.034)	(0.043)	(0.035)	(0.038)	(0.061)
	{0.0820}	{0.608}	{0.286}	{0.399}	{0.838}
Treat × *Select*		0.142**	0.161**	0.154**	0.151
		(0.058)	(0.051)	(0.050)	(0.087)
		{0.0360}	{0.0170}	{0.0310}	{0.148}
Select	0.0113	-0.0332	-0.0445	-0.0409	-0.0519
	(0.033)	(0.036)	(0.027)	(0.028)	(0.045)
	{0.733}	{0.475}	{0.190}	{0.296}	{0.348}
Treat 50					0.0138
					(0.049)
					{0.808}
Treat 50 × *Select*					0.133
					(0.072)
					{0.138}
R^2	0.00	0.01	0.02	0.02	0.02
Number of Phone Surveys	1584	1584	1584	1584	1110
Number of AEAs	94	94	94	94	134
Number of ALATs	33	33	33	33	44
RI p-value (1) = 0	.039	.652	.356	.474	.803
RI p-value (2) = 0		.032	.011	.036	.097
RI p-value (4) = 0					.11
RI p-value (5) = (2)					.969
Includes Basic Controls			✓	✓	✓
Includes Cognitive Controls				✓	✓

Regressions also include survey wave indicators. Basic controls include gender, age, marital status, and average distance to farmers. Cognitive controls include digit span, the Big 5 stability meta-trait, and the Big 5 plasticity meta-trait. Columns (1)-(4) include cells A through D. Column (5) adds cells E and F but only includes the first round of phone surveys since AEAs in cell F were given a phone before the second round of surveys. Cluster robust standard errors and p-values from a Wu, 1986 wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values. Randomization inference (RI) p-values are based on 1000 replication draws using Stata's ritest command (see Heß, 2017). The RI p-values respectively test: (1) the coefficient on **Treat** = 0; (2) the coefficient on **Treat** × *Select* = 0; (4) the coefficient on **Treat 50** = 0; and (5) = (2) equality of the coefficients on **Treat** × *Select* and **Treat50** × *Select*

Table 3.4: Testing for Hawthorne Effects

	Farmer was visited in the last week	AEA works in small ALAT
	(1)	(2)
Treated	0.058*	
	(0.029)	
	{0.066}	
Treated × small ALAT	0.031	
	(0.074)	
	{0.698}	
small ALAT	-0.021	
	(0.042)	
	{0.624}	
Male	0.036	0.004
	(0.029)	(0.099)
	{0.306}	
Age	0.003*	-0.003
	(0.001)	(0.005)
	{0.084}	
Married	-0.001	-0.089
	(0.030)	(0.088)
	{0.982}	
Distance to Farmers (log)	0.038	-0.014
	(0.027)	(0.082)
	{0.224}	
Incumbent Party	-0.026	0.097
	(0.025)	(0.102)
	{0.322}	
Digit Span	0.010	0.037
	(0.011)	(0.034)
	{0.419}	
Big 5 — Stability	0.012	0.039
	(0.015)	(0.040)
	{0.509}	

Continued on next page

Table 3.4 – *Continued from previous page*

	Farmer was visited in the last week	AEA works in small ALAT
	(1)	(2)
Big 5 — Plasticity	-0.011	-0.026
	(0.010)	(0.035)
	{0.349}	
Mean of Control Dep. Var	.271	.309
R^2	0.01	0.04
Number of Phone Surveys	1842	
Number of AEAs	136	136
Number of ALATs	71	71
RI p-value on **Treated**	0.025	
RI p-value on **Treated** × small ALAT	0.689	

Outcome in column (1) is from the farmer phone survey and in column (2) is from administrative data. Big 5 stability and plasticity measures normalized to have mean zero and unit variance. These regressions include unreported indicators for small ALATs, survey wave, and an interaction of the two. Cluster robust standard errors and p-values from a Wu, 1986 wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values. Randomization inference (RI) p-values are based on 1000 replication draws using Stata's ritest command (see Heß, 2017).

Table 3.5: First Stage Probit Regressions

	(1)	(2)
Male	-0.519	-0.511
	(0.349)	(0.395)
	{0.144}	{0.242}
Age	-0.036**	-0.037**
	(0.018)	(0.017)
	{0.024}	{0.028}
Married	0.863**	0.791*
	(0.459)	(0.441)
	{0.036}	{0.054}
Average Distance	0.331	0.318
	(0.173)	(0.197)
	{0.125}	{0.118}
Incumbent Party	-0.808**	-0.840**
	(0.359)	(0.367)
	{0.048}	{0.047}
Digit Span		-0.111
		(0.142)
		{0.466}
Big 5 — Stability		-0.234*
		(0.124)
		{0.097}
Big 5 — Plasticity		0.122
		(0.154)
		{0.437}
Pseudo R^2	0.159	0.185
Number of AEAs	94	94
Number of ALATs	33	33

Coefficients are from a probit regression of an indicator for the AEA being selected on AEA characteristics. Cluster robust standard errors and p-values from a Kline and Santos, 2012 wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table 3.6: Treatment Effect Heterogeneity on Observable and Unobservable Characteristics

	(1)	(2)	(3)
Main Effects:			
Inverse Mills Ratio	-0.021	-0.019	-0.016
	(0.023)	(0.016)	(0.016)
Average Treatment Effect	0.062*	0.058*	0.064*
	(0.035)	(0.026)	(0.028)
	{0.092}	{0.070}	{0.069}
Interactions with Treatment:			
Inverse Mills	0.088**	0.064**	0.064**
	(0.036)	(0.022)	(0.025)
	{0.036}	{0.026}	{0.043}
Male		-0.021	-0.004
		(0.051)	(0.066)
		{0.719}	{0.972}
Age		-0.003	-0.005
		(0.003)	(0.003)
		{0.330}	{0.214}
Married		-0.084	-0.094
		(0.061)	(0.052)
		{0.251}	{0.133}
Average Distance to Farmers (log)		0.101	0.104
		(0.068)	(0.077)
		{0.253}	{0.380}
Incumbent Party		-0.081*	-0.069
		(0.038)	(0.048)
		{0.050}	{0.165}
Digit Span			-0.064**
			(0.027)
			{0.036}
Big5 — Stability			-0.038
			(0.049)
			{0.566}

Continued on next page

Table 3.6 – *Continued from previous page*

	(1)	(2)	(3)
Big5 — Plasticity			0.035
			(0.025)
			{0.329}
R^2	0.009	0.023	0.028
p-value for Observable Interactions		0.206	0.155
p-value for Observable Interactions (Not Wild)		0.031	0.001
Number of Phone Surveys	1584	1584	1584
Number of AEAs	94	94	94
Number of ALATs	33	33	33
RI p-value on **Treated**	.039	.054	.049
RI p-value on **Treated** × *Inverse Mills*	.023	.098	.15
Basic Controls		✓	✓
Cognitive Controls			✓

Left-hand side variable is whether the farmer was visted in the last week. The inverse Mills ratio is the generalized residual—the expected value of the error term—from a probit regression of being selected on the corresponding controls from the column. Regressions also include survey wave indicators. Main effects of basic and cognitive controls omitted for space. The p-value for all observable interactions reported in the bottom rows is the implied p-value from a wild bootstrapped F-test for coefficients on treatment interacted with age, married, average distance, male, digit span, and Big 5 measures. Cluster robust standard errors and p-values from a Wu, 1986 wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values. Randomization inference (RI) p-values are based on 1000 replication draws using Stata's ritest command (see Heß, 2017).

Table 3.7: Treatment Effects by Roll-Out Levels and Allocation Rules

Rollout	Random	Supervisor	Distance	Prediction	Sophisticated
0.25	0.016	0.043	0.032	0.049	0.059
0.54	0.034	0.070	0.051	0.075	0.085
0.70	0.045	0.077	0.054	0.076	**0.090**
0.75	0.048	0.077	0.058	0.078	0.088
0.76	0.048	0.077	0.058	**0.079**	0.088
0.77	0.049	**0.077**	0.057	0.077	0.088
0.97	0.062	0.069	**0.069**	0.069	0.070
1.00	**0.064**	0.064	0.064	0.064	0.064

This table displays estimated treatment effects at the different roll-out levels shown in the first column for the various allocation rules we consider. The number displayed in bold represents the largest treatment effect under a given allocation rule. The allocation rules shown are random - treatment assignment is made randomly; supervisor - treatment assignment is what would be achieved under decentralization with supervisors making the assignment decision; distance - treatment is assigned first to those AEAs whose beneficiaries live further from the local ALAT office; prediction - uses the control group to predict baseline performance using the basic and cognitive controls and then treats first those AEAs who are predicted to be the worst performers in the baseline; and sophisticated - runs a pilot experiment at low roll-out to establish a map between treatment response and observables and then treats first those AEAs who are predicted to have the highest treatment response.

Bibliography

Acemoglu, D., Aghion, P., Lelarge, C., Van Reenen, J., & Zilibotti, F. (2007). Technology, information, and the decentralization of the firm. *Quarterly Journal of Economics*, *122*(4), 1759–1799. doi:10.1162/qjec.2007.122.4.1759. eprint: /oup/backfile/content_public/journal/qje/122/4/10.1162/qjec.2007.122.4.1759/2/122-4-1759.pdf

Aghion, P., & Tirole, J. (1997). Formal and real authority in organizations. *Journal of Political Economy*, *105*(1), 1–29. Retrieved from http://www.jstor.org/stable/2138869

Aker, J. C., & Ksoll, C. (2018). *Call me educated: Evidence from a mobile phone experiment in Niger*. Center for Global Development Working Paper 406.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., & Tobias, J. (2012). Targeting the poor: Evidence from a field experiment in Indonesia. *American Economic Review*, *102*(4), 1206–40. doi:10.1257/aer.102.4.1206

Alderman, H. (2002). Do local officials know something we don't? Decentralization of targeted transfers in Albania. *Journal of Public Economics*, *83*(3), 375–404. doi:https://doi.org/10.1016/S0047-2727(00)00145-6

Altonji, J. G., & Card, D. (1991). The effects of immigration on the labor market outcomes of less-skilled natives. In *Immigration, trade, and the labor market* (pp. 201–234). University of Chicago Press.

Alvarez, J. A. (2018). The agricultural wage gap: Evidence from brazilian micro-data. Working paper.

Au, C.-C., & Henderson, J. V. (2006). How migration restrictions limit agglomeration and productivity in china. *Journal of Development Economics*, *80*(2), 350–388.

Baird, S., Hamory, J., & Miguel, E. (2008). Tracking, Attrition and Data Quality in The Kenyan Life Panel Survey Round 1 (KLPS-1). *University of California CIDER*. Working Paper.

Bajari, P., & Kahn, M. E. (2005). Estimating housing demand with an application to explaining racial segregation in cities. *Journal of Business & Economic Statistics*, *23*(1), 20–33.

Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D., & Singh, N. (2015). *Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy and training*. Unpublished Working Paper.

Banerjee, A., Glennerster, R., & Duflo, E. (2008). Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system. *Journal of the European Economic Association*, *6*(2-3), 487–500.

Basile, P. F., Landon-Lane, J., & Rockoff, H. (2010). *Money and interest rates in the united states during the great depression* (Working Paper No. 16204). National Bureau of Economic Research. doi:10.3386/w16204

Bayer, P., Casey, M., Ferreira, F., & McMillan, R. (2017). Racial and ethnic price differentials in the housing market. *Journal of Urban Economics*, *102*, 91–105.

Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, *115*(4), 588–638.

Bayer, P., McMillan, R., & Rueben, K. (2004). *An equilibrium model of sorting in an urban housing market* (Working Paper No. 10865). National Bureau of Economic Research.

Bayer, P., & Timmins, C. (2005). On the equilibrium properties of locational sorting models. *Journal of Urban Economics*, *57*(3), 462–477. doi:https://doi.org/10.1016/j.jue.2004.12.008

Bazzi, S. (2017). Wealth heterogeneity and the income elasticity of migration. *American Economic Journal: Applied Economics*, *9*(2), 219–55.

Bazzi, S., Gaduh, A., Rothenberg, A. D., & Wong, M. (2016). Skill transferability, migration, and development: Evidence from population resettlement in indonesia. *American Economic Review*, *106*(9), 2658–2698.

Beegle, K., Weerdt, J. D., & Dercon, S. (2011). Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey. *Review of Economics and Statistics*, *93*(3), 1010–1033.

Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Bertrand, M., & Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, *94*(4), 991–1013. doi:10.1257/0002828042002561

Bloom, N., Garicano, L., Sadun, R., & Van Reenen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, *60*(12), 2859–2885.

Bloom, N., Sadun, R., & Van Reenen, J. (2012). The organization of firms across countries. *Quarterly Journal of Economics*, *127*(4), 1663–1705. doi:10.1093/qje/qje029. eprint: /oup/backfile/content_public/journal/qje/127/4/10.1093/qje/qje029/2/qje029.pdf

Boustan, L. P. (2010). Was postwar suburbanization “white flight?” evidence from the black migration. *The Quarterly Journal of Economics*, *125*(1), 417–443.

Bresnahan, T., Brynjolfsson, E., & Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quarterly Journal* *of Economics*, *117*(1), 339–376. Retrieved from https://EconPapers.repec.org/RePEc:oup:qjecon:v:117:y:2002:i:1:p:339-376.

Brock, W. A., & Durlauf, S. N. (2002). A multinomial-choice model of neighborhood effects. *American Economic Review*, *92*(2), 298–303.

Bryan, G., Chowdhury, S., & Mobarak, A. (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, *82*(5), 1671–1748.

Bryan, G., & Morten, M. (2017). The aggregate productivity effects of internal migration: Evidence from indonesia.

Callen, M., Gulzar, S., Hasanain, A., Khan, M. Y., & Rezaee, A. (2018). *Data and policy decisions: Experimental evidence from Pakistan*. Unpublished Working Paper.

Cameron, A. C., & Miller, D. L. (2015). A practioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372.

Card, D. (1996). The effects of union on the structure of wages: A longitudinal analysis. *Econometrica*, *64*(4), 957–979.

Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, *19*(1), 22–64.

Card, D., Mas, A., & Rothstein, J. (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics*, *123*(1), 177–218.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, *13*(2), 185–213.

Caselli, F. (2005). Accounting for Cross-Country Income Differences. In P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth* (Chap. 9, Vol. 1, pp. 679–741). Handbook of Economic Growth. Elsevier. Retrieved from https://ideas.repec.org/h/eee/grochp/1-09.html

Chassang, S., Padró I Miquel, G., & Snowberg, E. (2012). Selective trials: A principal-agent approach to randomized controlled experiments. *American Economic Review*, *102*(4), 1279–1309. doi:10.1257/aer.102.4.1279

Chetty, R., & Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, *133*(3), 1107–1162. doi:10.1093/qje/qjy007. eprint: /oup/backfile/content_public/journal/qje/133/3/10.1093_qje_qjy007/2/qjy007.pdf

Chetty, R., & Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, *133*(3), 1163–1228. doi:10.1093/qje/qjy006. eprint: /oup/backfile/content_public/journal/qje/133/3/10.1093_qje_qjy006/1/qjy006.pdf

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states *. *The Quarterly Journal of Economics*, *129*(4), 1553–1623. doi:10.1093/qje/qju022. eprint: /oup/backfile/content_public/journal/qje/129/4/10.1093_qje_qju022/4/qju022.pdf

Chiquiar, D., & Hanson, G. H. (2005). International Migration, Self-Selection, and The Distribution of Wages: Evidence from Mexico and the United States. *Journal of Political Economy*, *113*(2), 239–281.

Clark, G. (2014). *The son also rises: Surnames and the history of social mobility*. Princeton University Press.

Coleman, J. S. (1966). Equality of educational opportunity.

Coleman, J. S., Kelly, S. D., & Moore, J. A. (1975). Trends in school segregation, 1968-73.

Cook, L. D., Logan, T. D., & Parman, J. M. (2014). Distinctively black names in the american past. *Explorations in Economic History*, *53*, 64–82.

Cutler, D. M., & Glaeser, E. L. (1997). Are ghettos good or bad? *The Quarterly Journal of Economics*, *112*(3), 827–872. doi:10.1162/003355397555361. eprint: /oup/backfile/content_public/journal/qje/112/3/10.1162/003355397555361/2/112-3-827.pdf

Cutler, D. M., Glaeser, E. L., & Vigdor, J. L. (1999). The rise and decline of the american ghetto. *Journal of Political Economy*, *107*(3), 455–506.

de la Roca, J., & Puga, D. (2016). Learning by working in big cities. *Review of Economic Studies*. Forthcoming.

de Mel, S., McKenzie, D., & Woodruff, C. (2009). Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics*, *88*(1).

de Rochambeau, G. (2017). *Monitoring and intrinsic motivation: Evidence from Liberia's trucking firms*. Unpublished Manuscript.

Dessein, W. (2002). Authority and communication in organizations. *Review of Economic Studies*, *69*(4), 811–838. doi:10.1111/1467-937X.00227. eprint: /oup/backfile/content_public/journal/restud/69/4/10.1111/1467-937x.00227/2/69-4-811.pdf

DeYoung, C. G. (2006). Higher-order factors of the big five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*(6), 1138–1151.

Dhaliwal, I., & Hanna, R. (2017). The devil is in the details: The successes and limitations of bureaucratic reform in India. *Journal of Development Economics*, *124*, 1–21.

DiPasquale, D. (1999). Why don't we know more about housing supply? *The Journal of Real Estate Finance and Economics*, *18*(1), 9–23. doi:10.1023/A:1007729227419

Duflo, E., Greenstone, M., Pande, R., & Ryan, N. (2018). The value of regulatory discretion: Estimates from environmental inspections in India. *Econometrica*, *86*(6), 2123–2160.

Duflo, E., Hanna, R., & Ryan, S. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, *102*(4), 1241–78.

Epple, D., & Sieg, H. (1999). Estimating equilibrium models of local jurisdictions. *Journal of political economy*, *107*(4), 645–681.

Gertler, P., Heckman, J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., ... Grantham-McGregor, S. (2014). Labor market returns to an early childhood stimulation intervention in jamaica. *Science*, *344*(6187), 998–1001.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2018). *Bartik instruments: What, when, why, and how* (Working Paper No. 24408). National Bureau of Economic Research. doi:10.3386/w24408

Goldstein, J. R., & Stecklov, G. (2016). From patrick to john f.: Ethnic names and occupational success in the last era of mass migration. *American Sociological Review*, *81*(1), 85–106. PMID: 27594705. doi:10.1177/0003122415621910

Gollin, D., Lagakos, D., & Waugh, M. E. (2014). The agricultural productivity gap. *Quarterly Journal of Economics*, *129*(2), 939–993.

Gollin, D., Parente, S., & Rogerson, R. (2002). The Role of Agriculture in Development. *American Economic Review*, *92*(2), 160–164. Retrieved from https://ideas.repec.org/a/aea/aecrev/v92y2002i2p160-164.html

Graham, B. S., & Temple, J. R. (2006). Rich nations, poor nations: How much can multiple equilibria explain? *Journal of Economic Growth*, *11*(1), 5–41.

Graser, A., Schmidt, J., Roth, F., & Brändle, N. (2017). Untangling origin-destination flows in geographic information systems. *Information Visualization*, 1473871617738122.

Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, *73*(3), 669–738. doi:10.1111/j.1468-0262.2005.00594.x

Hendricks, L., & Schoellman, T. (2017). *Human capital and development accounting: New evidence from wage gains at migration.*

Herrendorf, B., Rogerson, R., & Valentinyi, Á. (2014). Chapter 6 - growth and structural transformation. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 2, pp. 855–941). Handbook of Economic Growth. Elsevier.

Herrendorf, B., & Schoellman, T. (2018). Wages, human capital, and structural transformation. *American Economic Journal: Macroeconomics*. Forthcoming.

Heß, S. (2017). Randomization inference with Stata: A guide and software. *Stata Journal*, *17*(3), 630–651. Retrieved from http://www.stata-journal.com/article.html?article=st0489

Hsieh, C.-T., & Klenow, P. (2009). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics*, *124*(4), 1403–1448.

Hussam, R., Rigol, N., & Roth, B. (2017). *Targeting high ability entrepreneurs using community information: Mechanism design in the field*. Unpublished Manuscript.

Jaeger, D. A., Ruist, J., & Stuhler, J. (2018). *Shift-share instruments and the impact of immigration* (Working Paper No. 24285). National Bureau of Economic Research. doi:10.3386/w24285

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big-five trait taxonomy: History, measurement, and conceptual issues. In R. W. R. Oliver P. John & L. A. Pervin (Eds.), *Handbook of personality: Theory and research*. New York, NY: Guilford Press.

Johnston, B., & Kilby, P. (1978). *Agriculture and structural transformation: Economic strategies in late-developing countries*. Oxford University Press.

Khan, A. Q., Khwaja, A. I., & Olken, B. A. (2016). Tax farming redux: Experimental evidence on performance pay for tax collectors. *Quarterly Journal of Economics*, *131*(1), 219–271.

Kleemans, M. (2016). Migration choice under risk and liquidity constraints. Working Paper.

Kleemans, M., & Magruder, J. (2017). Labor market changes in response to immigration: Evidence from internal migration driven by weather shocks. *Economic Journal*.

Kline, P., & Walters, C. R. (2019). On Heckits, LATE, and numerical equivalence. *Econometrica*. Forthcoming.

Kline, P. (2010). Place based policies, heterogeneity, and agglomeration. *American Economic Review*, *100*(2), 383–87.

Kline, P., & Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, *1*(1), 23–41. Retrieved from https://ideas.repec.org/a/bpj/jecome/v1y2012i1p23-41n4.html

Kremer, M., Miguel, E., & Thornton, R. (2009). Incentives to learn. *The Review of Economics and Statistics*, *91*(3), 437–456.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Kuznets, S. (1973). Modern economic growth: Findings and reflections. *American Economic Review*, *63*(3), 247–258. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=4507611&site=ehost-live

Lagakos, D., & Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, *103*(2), 948–980.

Lee, K., Brewer, E., Christiano, C., Meyo, F., Miguel, E., Podolsky, M., ... Wolfram, C. (2015). Barriers to electrification for "under grid" households in rural kenya. *Development Engineering*, *1*, 26–35.

Lemieux, T. (1998). Estimating the effects of unions on wage inequality in a panel data model with comparative advantage and nonrandom selection. *Journal of Labor Economics*, *16*(2), 261–291. Retrieved from http://www.jstor.org/stable/10.1086/209889

Lewis, W. (1955). *The theory of economic growth*. Homewood: R. D. Irwin. Retrieved from https://books.google.com/books?id=HcvtkWZfsLMC

Lieberson, S. (1980). *A piece of the pie: Blacks and white immigrants since 1880*. Univ of California Press.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Massey, D. S., Alarcón, R., Durand, J., & González, H. (1987). *Return to aztlan: The social process of international migration from western mexico*. University of California Press.

Massey, D. S., & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.

McKenzie, D., Gibson, J., & Stillman, S. (2010). How Important Is Selection? Experimental vs. Non-Experimental Measures of The Income Gains from Migration. *Journal of the European Economic Association*, *8*(4), 913–945.

Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, *72*(1), 159–217.

Mookherjee, D. (2006). Decentralization, hierarchies, and incentives: A mechanism design perspective. *Journal of Economic Literature*, *44*(2), 367–390. doi:10.1257/jel.44.2.367

Mora, R., & Ruiz-Castillo, J. (2010). *A kullback-leibler measure of conditional segregation*.

Mora, R., & Ruiz-Castillo, J. (2011). Entropy-based segregation indices. *Sociological Methodology*, *41*(1), 159–194.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69–85. Retrieved from http://www.jstor.org/stable/1913646

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the u. s. labor market*. *The Quarterly Journal of Economics*, *118*(2), 549–599. doi:10.1162/003355303321675455

Munshi, K., & Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, *106*(1), 46–98.

Nakamura, E., Sigurdsson, J., & Steinsson, J. (2017). The gift of moving: Intergenerational consequences of a mobility shock. *Working Paper*.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, *69*(2), 307–342.

Olivetti, C., & Paserman, M. D. (2015). In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940. *American Economic Review*, *105*(8), 2695–2724.

Olsen, R. J. (1980). A least squares correction for selectivity bias. *Econometrica*, *48*(7), 1815–1820. Retrieved from http://www.jstor.org/stable/1911938

Ondrich, J., Stricker, A., & Yinger, J. (1998). Do real estate brokers choose to discriminate? evidence from the 1989 housing discrimination study. *Southern Economic Journal*, 880–901.

Ondrich, J., Stricker, A., & Yinger, J. (1999). Do landlords discriminate? the incidence and causes of racial discrimination in rental housing markets. *Journal of Housing Economics*, *8*(3), 185–204.

Pacini, D., & Windmeijer, F. (2016). Robust inference for the two-sample 2sls estimator. *Economics Letters*, *146*, 50–54.

Pack, H. (1972). *Employment and productivity in kenyan manufacturing*. Institute for Development Studies, University of Nairobi.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, *110*(4), 705–729. doi:10.1086/340779. eprint: https://doi.org/10.1086/340779

Piazza, A., Rendine, S., Zei, G., Moroni, A., & Cavalli-Sforza, L. L. (1987). Migration rates of human populations from surname distributions. *Nature*, *329*(6141), 714.

Porzio, T. (2017). Cross-country differences in the optimal allocation of talent and technology. Working Paper.

Preobrazhensky, E. (1921). *The crisis of soviet industrialization(1980 ed.)* London: MacMillan. Retrieved from https://books.google.com/books?id=Sx-TAAAAIAAJ

Reber, S. J. (2005). Court-ordered desegregation successes and failures integrating american schools since brown versus board of education. *Journal of Human Resources*, *40*(3), 559–590.

Restuccia, D., & Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous plants. *Review of Economic Dynamics*, *11*(4), 707–720. Retrieved from https://ideas.repec.org/a/red/issued/07-48.html

Restuccia, D., Yang, D. T., & Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, *55*, 234–250.

Rosenstein-Rodan, P. N. (1943). Problems of industrialisation of eastern and south-eastern europe. *Economic Journal*, *53*(210/211), 202–211.

Rostow, W. W. (1960). *The stages of economic growth: A non-communist manifesto*. Cambridge: Cambridge University Press. Retrieved from https://books.google.com/books?id=XzJdpd8DbYEC

Rothstein, R. (2017). *The color of law: A forgotten history of how our government segregated america*. Liveright Publishing.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers*, *3*(2), 135–146.

Rubalcava, L., Teruel, G., Thomas, D., & Goldman, N. (2008). The healthy migrant effect: New findings from the mexican family life survey. *American Journal of Public Health*, *98*(1), 78–84.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (n.d.). 1940 census: Instructions to enumerators. https://usa.ipums.org/usa/voliii/inst1940.shtml. Accessed: November 11, 2018.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., & Sobek, M. (2018). Ipums usa: Version 8.0. doi:10.18128/d010.v8.0

Saiz, A. (2003). Room in the kitchen for the melting pot: Immigration and rental prices. *The Review of Economics and Statistics*, *85*(3), 502–521. doi:10.1162/003465303322369687. eprint: https://doi.org/10.1162/003465303322369687

Saiz, A. (2010). The geographic determinants of housing supply*. *The Quarterly Journal of Economics*, *125*(3), 1253–1296. doi:10.1162/qjec.2010.125.3.1253. eprint: /oup/backfile/content_public/journal/qje/125/3/10.1162/qjec.2010.125.3.1253/2/125-3-1253.pdf

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*(2), 143–186.

Schelling, T. C. (1978). *Micromotives and macrobehavior*. WW Norton & Company.

Schultz, T. P. (1988). Handbook of development economics. In H. Chenery & T. N. Srinivasan (Eds.), (Chap. 13 Education Investments and Returns, Vol. 1, pp. 543–630). Elsevier.

Shertzer, A., & Walsh, R. P. (2016). *Racial sorting and the emergence of segregation in american cities* (Working Paper No. 22077). National Bureau of Economic Research.

Strauss, J., Witoelar, F., & Sikoki, B. (2016). The fifth wave of the indonesian family life survey (ifls4): Overview and field report. *WR-1143/1-NIA/NICHD*.

Stren, R., Halfani, M., & Malombe, J. (1994). Beyond capitalism vs. socialism in kenya and tanzania. (Chap. Coping with Urbanization and Urban Policy). Colorado: Lynne Rienner.

Stuart, B. A., & Taylor, E. J. (2017). Migration networks and location decisions: Evidence from us mass migration.

Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica*, *79*(1), 159–209. Retrieved from http://dx.doi.org/10.3982/ECTA7749

Tabellini, M. (2018). Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration.

Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., & Suriastini, W. (2012). Cutting the costs of attrition: Results from the indonesia family life survey. *Journal of Development Economics*, *98*(1), 108–123.

Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. University of Chicago Press.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, *14*(4), 1261–1295. doi:10.1214/aos/1176350142

Yang, D. (2006). Why Do Migrants Return to Poor Countries? Evidence from Philippine Migrants' Responses to Exchange Rate Shocks. *Review of Economics and Statistics*, *88*(4), 715–735. Retrieved from https://ideas.repec.org/a/tpr/restat/v88y2006i4p715-735.html

Yinger, J. (1986). Measuring racial discrimination with fair housing audits: Caught in the act. *The American Economic Review*, *76*(5), 881–893.

Young, A. (2013). Inqeuality, the urban-rural gap, and migration. *Quarterly Journal of Economics*, *128*(4), 1727–1785.

Young, A. (2016). *Improved, nearly exact, statistical inference with robust and clustered covariance matrices using effective degrees of freedom corrections.*

Zei, G., Barbujani, G., Lisa, A., Fiorani, O., Menozzi, P., Siri, E., & Cavalli-Sforza, L. L. (1993). Barriers to gene flow estimated by surname distribution in italy. *Annals of Human Genetics*, *57*(2), 123–140.

Zei, G., Guglielmino, C. R., Siri, E., Moroni, A., & Cavalli-Sforza, L. L. (1983). Surnames as neutral alleles: Observations in sardinia. *Human Biology*, 357–365.

Appendix A

Appendix to Chapter 2

A.1 Additional Figures

Figure A.1: Log GDP per Capita and Agricultural Share

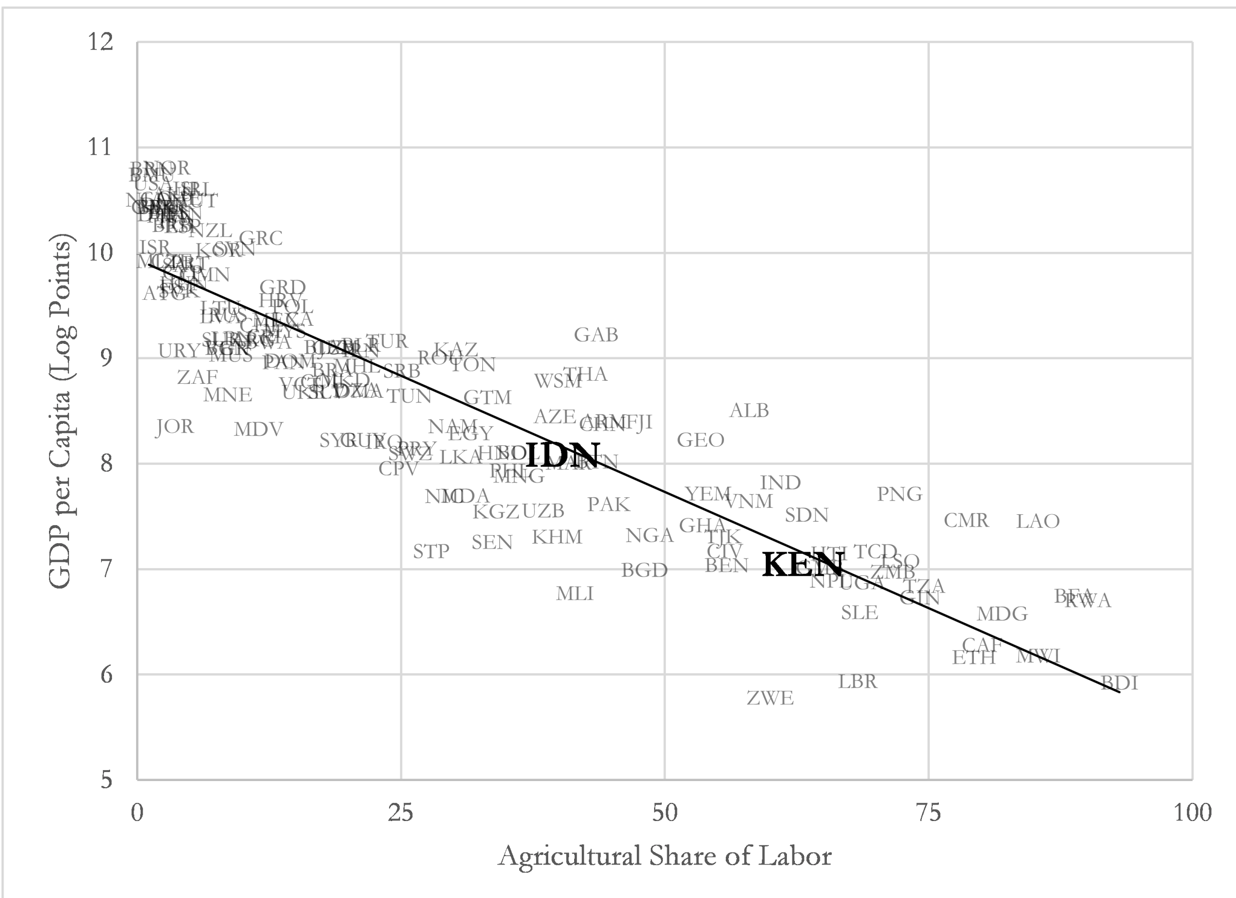


Table source data is from Gollin, Lagakos, and Waugh (2014), Online Appendix Table 4. Kenya (KEN) and Indonesia (IDN) are highlighted.

Figure A.2: Agricultural Share and Agricultural Productivity Gap

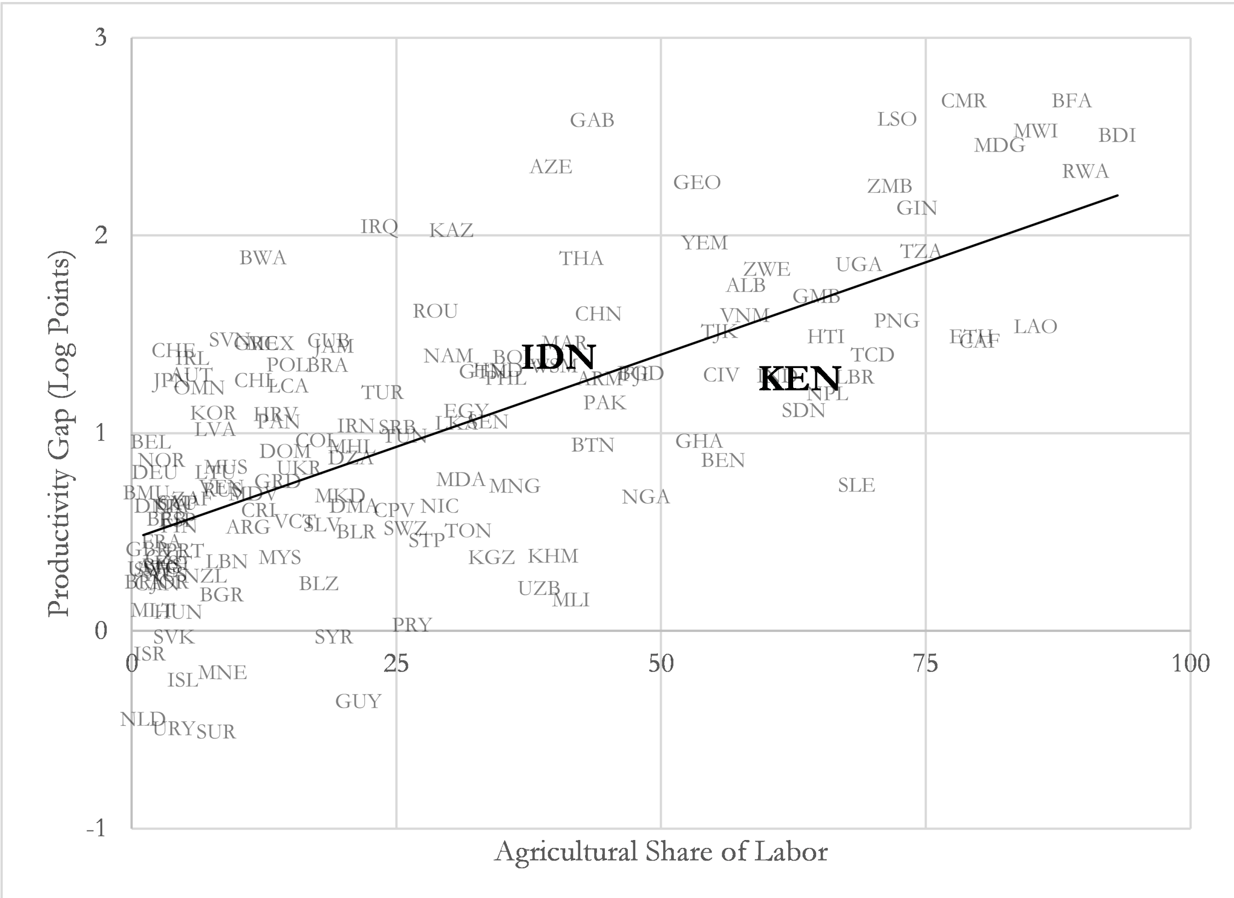


Table source data is from Gollin, Lagakos, and Waugh (2014), Online Appendix Table 4. Kenya (KEN) and Indonesia (IDN) are highlighted.

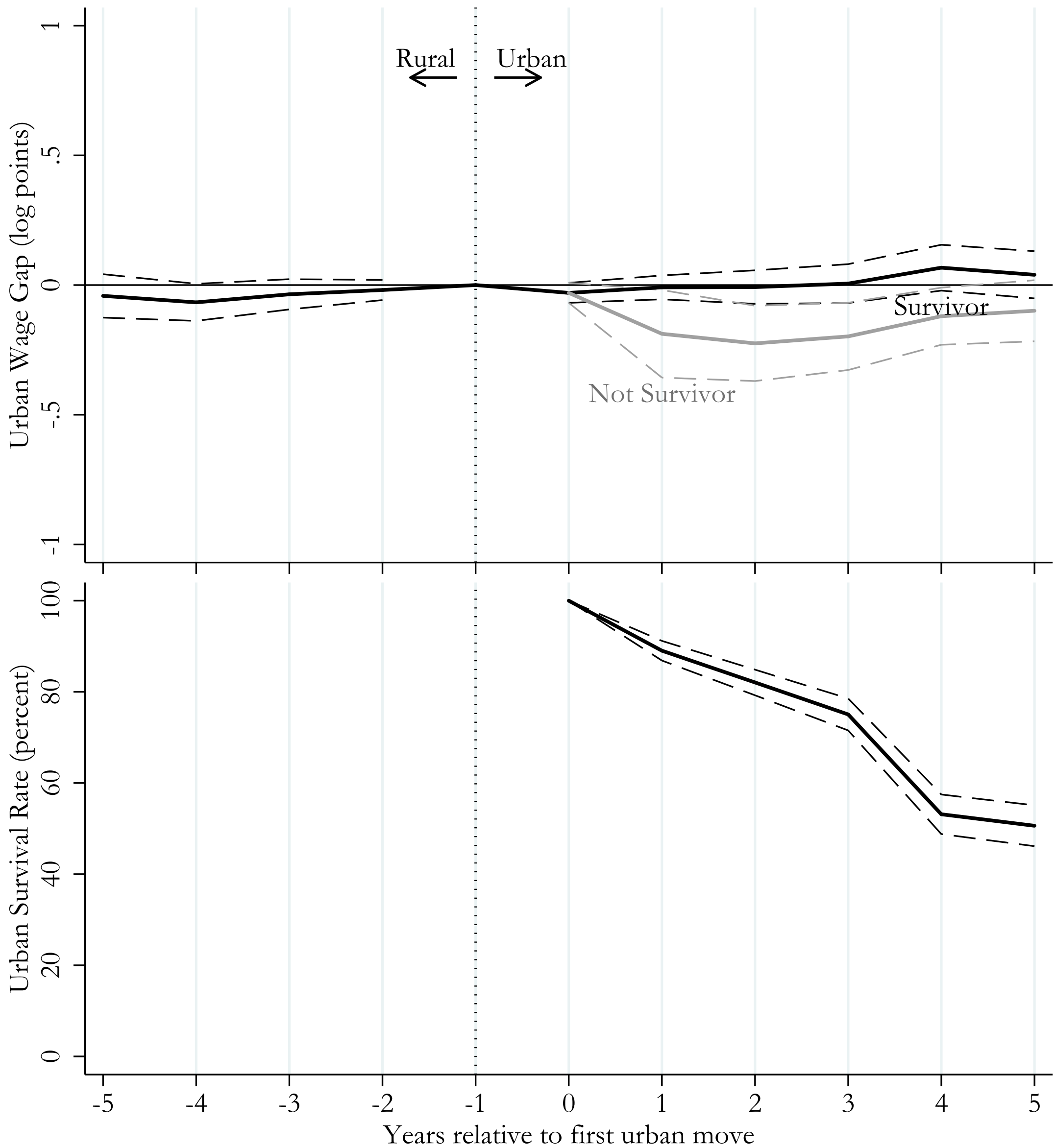
Figure A.3: Types of Individual Agricultural Productivity Data

	Lower quality measures ←			→ Higher quality measures
(A) Indonesia				
Source of agricultural productivity and hours		**Self-employed profits (commercial and subsistence agriculture)[1]**		**Wage employment**
Individual-years in Agriculture		55,149		24,567
Individuals in Agriculture		6,351		4,765
Agriculture productivity gap (Standard error) [Individual-months]		0.084*** (0.032) [142,993]		0.021 (0.026) [149,939]
		0.078*** (0.021) 275,600		
(B) Kenya				
Source of agricultural productivity and hours	**Less reliable individual agricultural productivity data[2]**	**Self-employed profits (subsistence agriculture)**	**Self-employed profits (commencial agriculture)**	**Wage employment**
Individual-months in Agriculture	16,830	2,991	4,124	14,422
Individuals in Agriculture	1,424	253	105	556
Agriculture productivity gap (Standard error) [Individual-months]		0.085 (0.189) [36,536]		0.157 (0.121) [97,572]
		0.061 (0.106) 4,791		

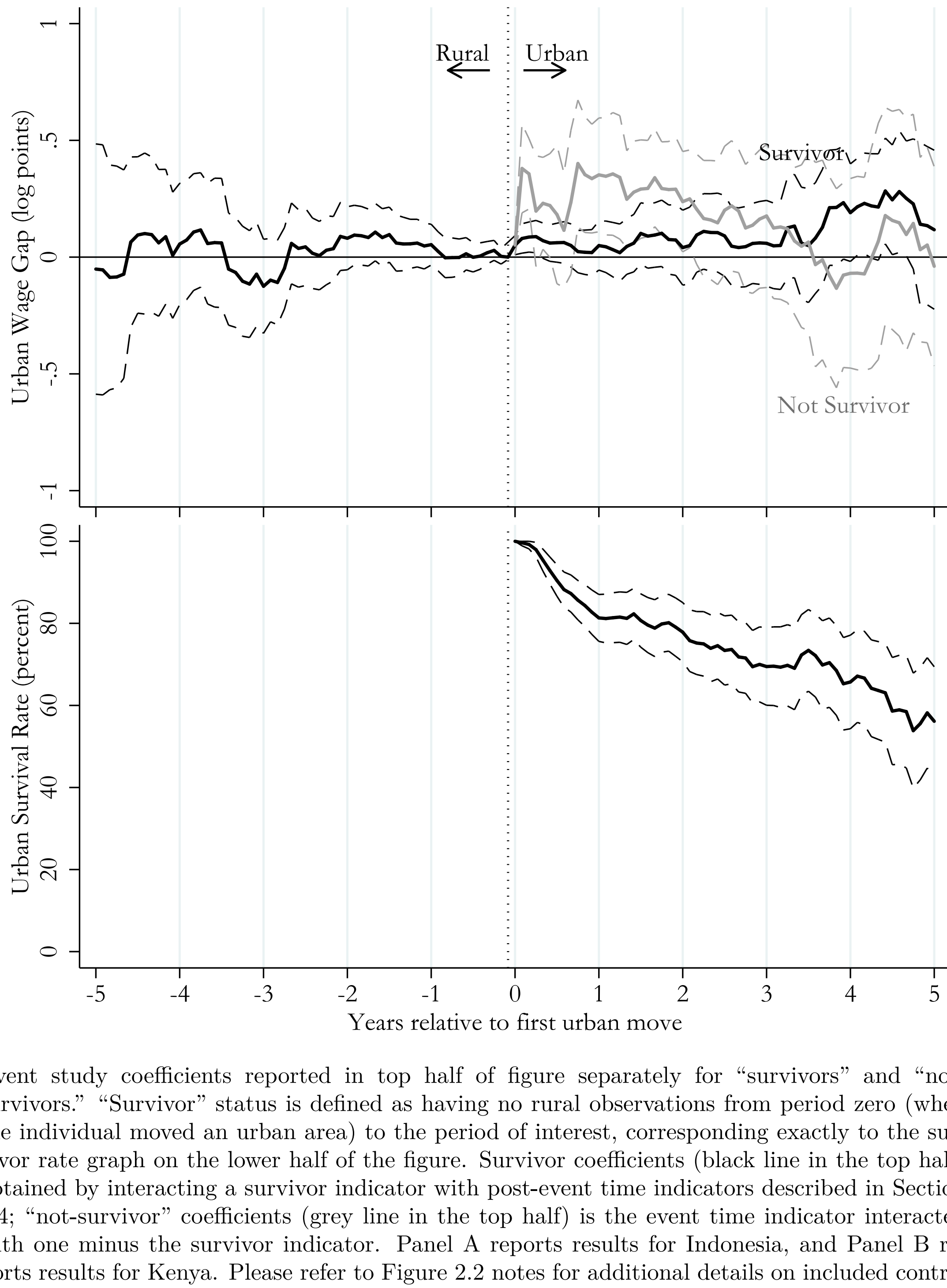
[1]The IFLS does not distinguish between profits in subsistence and commercial agriculture. [2]Less reliable agricultural productivity data encompasses individual-months where the only source of agricultural productivity data is from activities where the respondent is not the main decision maker and other household members contribute some hours. All estimates can be found in Table 2.8.

Figure A.4: Event Study of Urban Migration for Urban Survivors

(a) Indonesia



(b) Kenya



Event study coefficients reported in top half of figure separately for “survivors” and “not-survivors.” “Survivor” status is defined as having no rural observations from period zero (when the individual moved an urban area) to the period of interest, corresponding exactly to the survivor rate graph on the lower half of the figure. Survivor coefficients (black line in the top half) obtained by interacting a survivor indicator with post-event time indicators described in Section 2.4; “not-survivor” coefficients (grey line in the top half) is the event time indicator interacted with one minus the survivor indicator. Panel A reports results for Indonesia, and Panel B reports results for Kenya. Please refer to Figure 2.2 notes for additional details on included control variables and computation of survivor rates.

Figure A.5: Event Study of Rural Migration

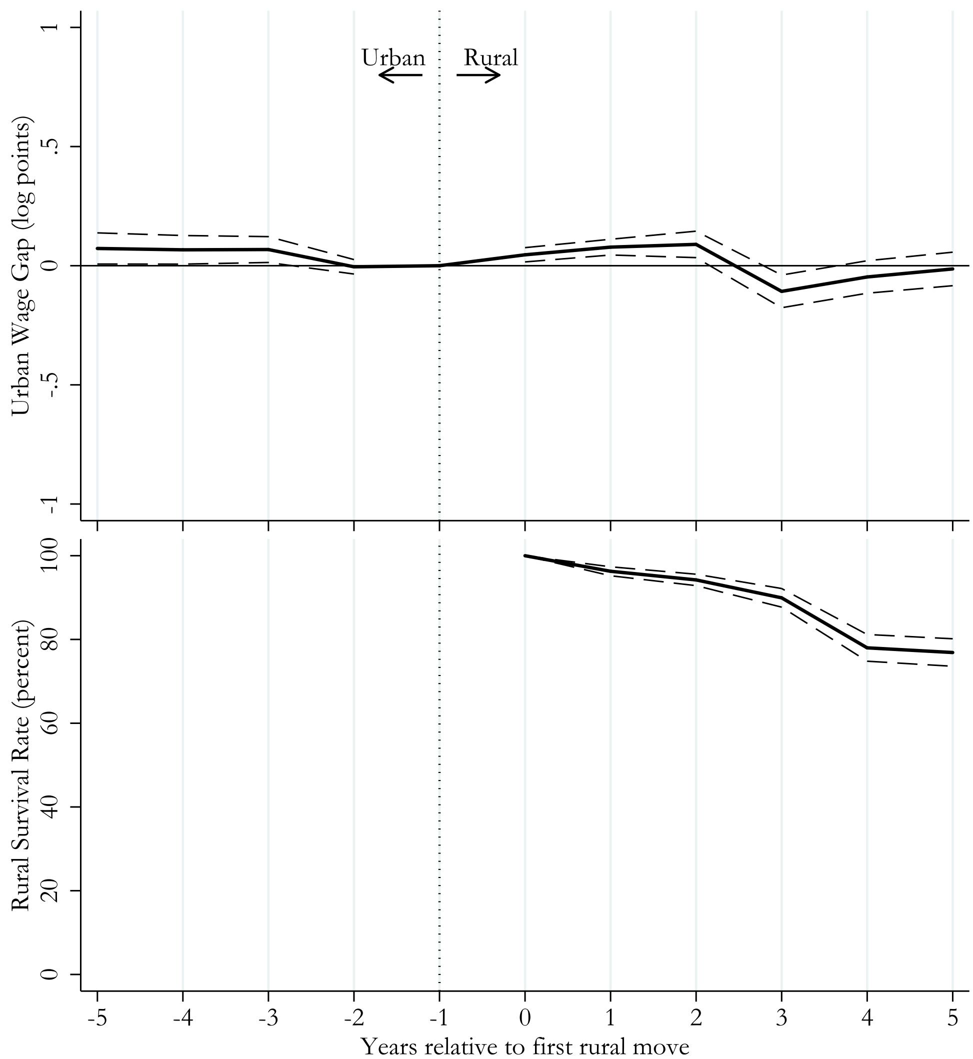


Figure uses data on individuals in the IFLS who are born in urban areas. Event time indicator variables defined analogously to Figure 2.2 except with respect to individuals' first observed rural move. Coefficients multiplied by negative 1 to interpret difference in earnings as an urban premium. Sample includes 1,296 movers with wage observations at the time of move and one period prior; 636 individuals report wages five years later. Please refer to Figure 2.2 notes for additional details on included control variables and computation of survivor rates.

Figure A.6: Event Study of Rural Migration for Survivors

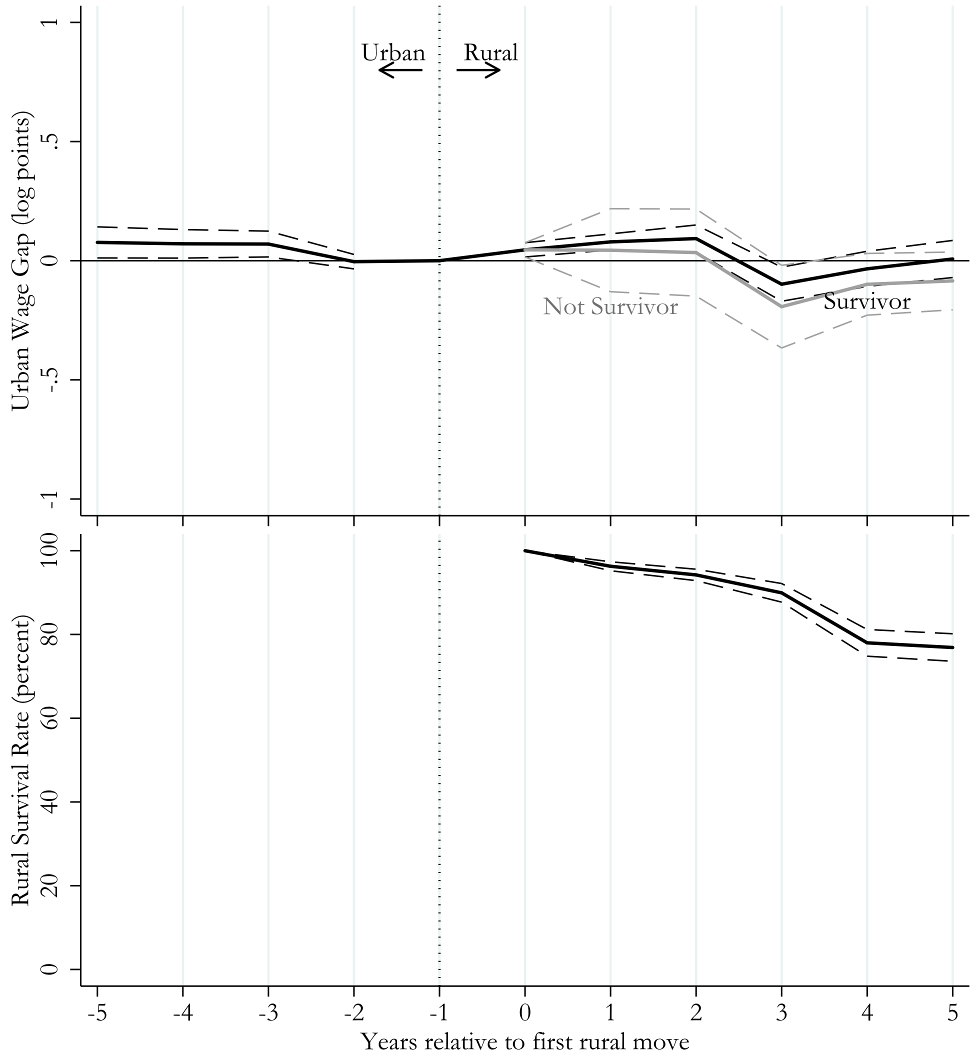
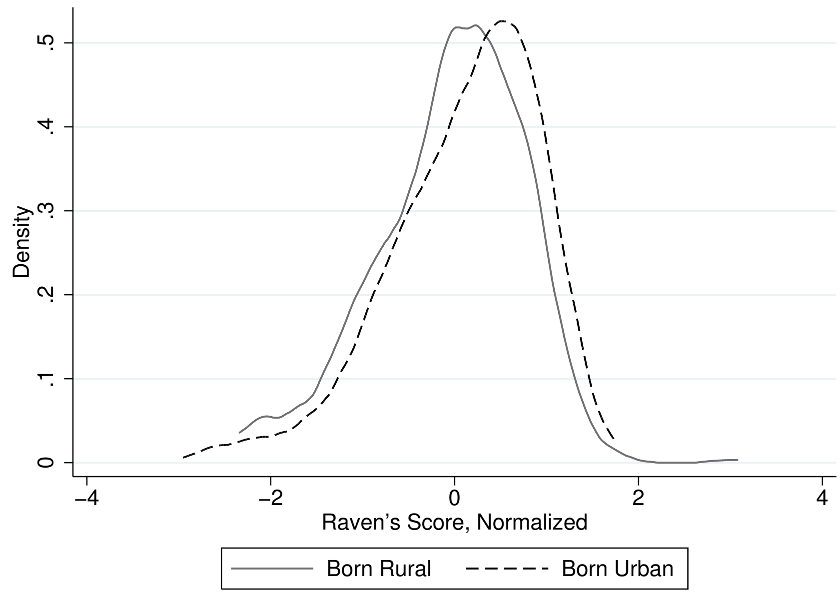


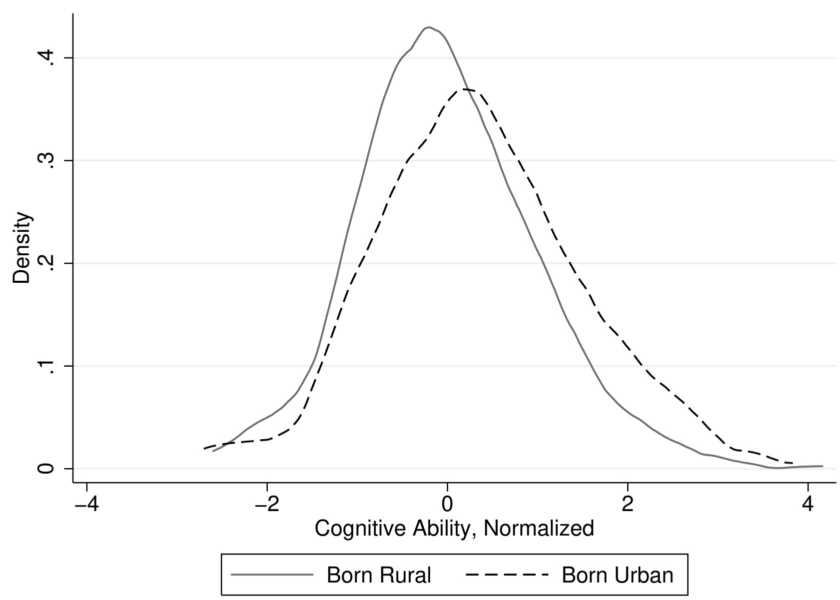
Figure uses data on individuals in the IFLS who are born in urban. Event study coefficients reported in top half of figure separately for "survivors" and "not-survivors." "Survivor" status is defined as having no urban observations from period zero (when the individual moved a rural area) to the period of interest, corresponding exactly to the survivor rate graph on the lower half of the figure. Survivor coefficients (black line in the top half) obtained by interacting a survivor indicator with post-event time indicators described in Section 2.4; "not-survivor" coefficients (grey line in the top half) is the event time indicator interacted with one minus the survivor indicator. Panel A reports results for Indonesia, and Panel B reports results for Kenya. Please refer to Figure 2.2 notes for additional details on included control variables and computation of survivor rates.

Figure A.7: Marginal Distributions of Cognitive Ability

(a) Indonesia—Normalized Ravens Matrices



(b) Kenya—Normalized Cognitive Ability Index



Panel A uses data from the IFLS, and Panel B uses data from the KLPS Kids sample of children of the KLPS sample, age 4–6. Panel A shows the marginal distributions of Raven matrix scores, normalized by one-year age bins, and Panel B shows the marginal distribution of a constructed Cognitive Ability Index, normalized by six-month age bins. See Table 2.11 for additional details about the cognitive ability index.

A.2 Additional Tables

Table A.1: Correlates of Rural Migration—Indonesia (Born Urban)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	-0.309***					-0.224***	-0.221***
	(0.017)					(0.025)	(0.018)
Secondary Ed.		-0.198***				-0.144***	-0.155***
		(0.009)				(0.012)	(0.010)
College			-0.131***			-0.00407	-0.0238*
			(0.011)			(0.014)	(0.012)
Female				-0.0386***		-0.0513***	-0.0506***
				(0.009)		(0.010)	(0.009)
Raven's Z-score					-0.0294***	0.00229	
					(0.006)	(0.006)	
Constant	0.642***	0.463***	0.379***	0.374***	0.359***	0.675***	0.671***
	(0.016)	(0.007)	(0.005)	(0.006)	(0.005)	(0.024)	(0.017)
Observations	11812	11812	11812	11812	8341	8341	11812

This table is a rural migration analog of Table 2.3. Each cell represents a regression coefficient with an indicator for being a rural migrant as the dependent variable. The sample is restricted to individuals born in urban areas. Please see notes from Table 2.3.

Table A.2: Correlates of Employment in Non-Agriculture—Indonesia (Born Urban)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Primary Ed.	0.207***					0.133***	0.160***
	(0.015)					(0.022)	(0.015)
Secondary Ed.		0.115***				0.0800***	0.0862***
		(0.005)				(0.006)	(0.005)
College			0.0722***			0.00479	0.00914*
			(0.004)			(0.004)	(0.004)
Female				0.0277***		0.0363***	0.0358***
				(0.005)		(0.005)	(0.005)
Raven's Z-score					0.0346***	0.0168***	
					(0.003)	(0.004)	
Constant	0.733***	0.863***	0.913***	0.912***	0.928***	0.741***	0.713***
	(0.015)	(0.005)	(0.003)	(0.003)	(0.003)	(0.021)	(0.015)
Observations	11812	11812	11812	11812	8341	8341	11812

This table is a analogous to Table 2.4 but is estimated on individuals born in urban areas. Please see notes from Table 2.4.

Table A.3: Correlates of Meals Eaten—Kenya

	(1) Log Consumption	(2) Log Earnings	(3) Log Wage
Log(Meals)	0.194*	0.278***	0.228***
	(0.090)	(0.065)	(0.066)
Number of observations	1062	4693	4315

Each cell reports a regression coefficient with the log of meals as the independent variable; dependent variables listed in the header of the table. These regressions do not have the sample restrictions found in Table 2.2. Log of household per capita consumption in column 1 available only for a subset of individuals from KLPS 3. Robust standard errors clustered by individual reported below in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.4: Kenya Urban Towns

	Population	Percentage of Urban Individual-Months
Nairobi	3,133,518	43.5
Mombasa	938,131	14.3
Busia	61,715	6.7
Nakuru	307,990	4.5
Kisumu	409,928	4.5
Eldoret	289,380	2.6
Kakamega	91,768	1.3
Kitale	106,187	1.1
Bungoma	81,151	1.1
Naivasha	181,966	0.9
Gilgil	35,293	0.5
Other	.	18.9

This table presents a list of reported towns from urban individual-month observations. Urban status is defined based on respondent answering that they live in a large town or city. Column 3 lists the fraction of individual months in analysis from a particular town. The source for town populations is the 2009 Kenya Census.

Table A.5: Non-Agricultural/Agricultural Gap in Earnings using Alternative Definition of Agriculture

(a) Indonesia

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Only non-agricultural employment	0.619***	0.302***	0.088***	0.013				
	(0.012)	(0.011)	(0.016)	(0.018)				
Any non-agricultural employment					0.728***	0.357***	0.304***	0.123***
					(0.013)	(0.012)	(0.017)	(0.020)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	275600	275600	275600	275600	275600	275600	275600	275600
Number of individuals	31843	31843	31843	31843	31843	31843	31843	31843

(b) Kenya

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Only non-agricultural employment	0.747***	0.521***	0.210**	0.096				
	(0.061)	(0.056)	(0.084)	(0.097)				
Any non-agricultural employment					0.814***	0.571***	0.369***	0.064
					(0.064)	(0.059)	(0.091)	(0.108)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	134221	134221	134221	134221	134221	134221	134221	134221
Number of individuals	4791	4791	4791	4791	4791	4791	4791	4791

Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Tables 2.5 and 2.6 with alternate definitions of non-agriculture. In the first "Only non-agricultural employment," an individual-time is considered *agricultural* if any of their jobs are *agricultural*, and non-agricultural otherwise. In the second, "Any non-agricultural employment," an individual-time is considered *non-agricultural* if any of their jobs are *non-agricultural*, and agricultural otherwise. For columns 4 and 8, the dependent variable is the log of earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.6: Non-Agricultural/Agricultural Gap in Earnings Within Rural Areas

(a) Indonesia

	Dependent variable: Log Earnings			
	(1)	(2)	(3)	(4) Log Wage
Non-agricultural employment	0.563***	0.312***	0.245***	0.065***
	(0.015)	(0.013)	(0.020)	(0.024)
Individual fixed effects	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y
Number of observations	179756	179756	179756	179756
Number of individuals	21434	21434	21434	21434

(b) Kenya

	Dependent variable: Log Earnings			
	(1)	(2)	(3)	(4) Log Wage
Non-agricultural employment	0.340***	0.206***	-0.048	-0.272*
	(0.072)	(0.066)	(0.119)	(0.143)
Individual fixed effects	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y
Number of observations	63545	63545	63545	63545
Number of individuals	2953	2953	2953	2953

Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Table 2.5, but restricts the sample to observations where the individual resides in rural areas. For column 4, the dependent variable is the log of earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 4, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Gap in Earnings for those Aged 30 or Younger, Indonesia

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.527***	0.293***	0.160***	-0.008				
	(0.019)	(0.018)	(0.031)	(0.038)				
Urban					0.431***	0.257***	0.082***	0.017
					(0.014)	(0.012)	(0.017)	(0.020)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	83349	83349	83349	83349	83349	83349	83349	83349
Number of individuals	19814	19814	19814	19814	19814	19814	19814	19814

This table uses data from the IFLS. The table repeats some of the analyses shown in Tables 2.5 and 2.6 but restricts the sample to observations where the individual is aged 30 years or fewer to allow better comparability to the KLPS sample. For columns 4 and 8, the dependent variable is the log of earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.8: Gap in Wage Earnings

(a) Indonesia

	Dependent variable: Log Wage Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	1.073***	0.366***	0.219***	0.021				
	(0.018)	(0.015)	(0.022)	(0.026)				
Urban					0.581***	0.210***	0.049***	0.013
					(0.015)	(0.011)	(0.013)	(0.016)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	149939	149939	149939	149939	149939	149939	149939	149939
Number of individuals	23033	23033	23033	23033	23033	23033	23033	23033

(b) Kenya

	Dependent variable: Log Wage Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.872***	0.651***	0.404***	0.157				
	(0.070)	(0.065)	(0.096)	(0.121)				
Urban					0.732***	0.629***	0.198***	0.132**
					(0.040)	(0.036)	(0.049)	(0.053)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	97572	97572	97572	97572	97572	97572	97572	97572
Number of individuals	4079	4079	4079	4079	4079	4079	4079	4079

Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Tables 2.5 and 2.6, but instead of using all available earnings as the dependent variable, this table only includes earnings from wage employment. For columns 4 and 8, the dependent variable is earnings from wage employment divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.9: Gap in Self-Employment Earnings

(a) Indonesia

	Dependent variable: Log Self-Employment Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.380***	0.264***	0.068**	0.084***				
	(0.017)	(0.016)	(0.028)	(0.032)				
Urban					0.420***	0.242***	0.008	0.012
					(0.018)	(0.017)	(0.023)	(0.027)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	142993	142993	142993	142993	142993	142993	142993	142993
Number of individuals	17268	17268	17268	17268	17268	17268	17268	17268

(b) Kenya

	Dependent variable: Log Self-Employment Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.082	0.048	0.064	0.085				
	(0.156)	(0.140)	(0.160)	(0.189)				
Urban					0.671***	0.486***	0.077	0.072
					(0.096)	(0.089)	(0.125)	(0.151)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	36536	36536	36536	36536	36536	36536	36536	36536
Number of individuals	1263	1263	1263	1263	1263	1263	1263	1263

Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Tables 2.5 and 2.6, but instead of using all available earnings as the dependent variable, this table only includes earnings from self-employment. For columns 4 and 8, the dependent variable is earnings from self-employment divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.10: Alternative Samples Kenya

(a) Subsistence agriculture included also if not main decision maker

	Dependent variable: Log Wage Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.785***	0.544***	0.273***	0.063				
	(0.062)	(0.058)	(0.089)	(0.105)				
Urban					0.856***	0.675***	0.261***	0.165***
					(0.038)	(0.035)	(0.047)	(0.050)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	134210	134210	134210	134210	134210	134210	134210	134210
Number of individuals	4790	4790	4790	4790	4790	4790	4790	4790

(b) Subsistence agriculture not included

	Dependent variable: Log Wage Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.646***	0.447***	0.178*	0.007				
	(0.068)	(0.063)	(0.096)	(0.120)				
Urban					0.833***	0.666***	0.243***	0.145***
					(0.039)	(0.035)	(0.047)	(0.050)
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	132085	132085	132085	132085	132236	132085	132085	132085
Number of individuals	4678	4678	4678	4678	4691	4678	4678	4678

Panels A and B use data from the KLPS, described in Section 2.3. Panel A also includes productivity from subsistence agriculture if the individual is not the main decision maker for the agricultural activity. Panel B excludes all data from subsistence agriculture. For columns 4 and 8, the dependent variable is total earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.11: Gap in Food and Non-Food Consumption, Indonesia

(a) Food Consumption

	Dependent variable: Log Food Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.459***	0.156***	0.104***			
	(0.010)	(0.007)	(0.011)			
Urban				0.274***	0.083***	0.014
				(0.010)	(0.007)	(0.011)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	82272	82272	82272	82272	82272	82272
Number of individuals	34820	34820	34820	34820	34820	34820

(b) Non-Food Consumption

	Dependent variable: Log Non-Food Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.942***	0.433***	0.164***			
	(0.013)	(0.011)	(0.017)			
Urban				0.613***	0.242***	0.042***
				(0.013)	(0.010)	(0.016)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	82272	82272	82272	82272	82272	82272
Number of individuals	34820	34820	34820	34820	34820	34820

Both panels use data from the IFLS. Panels A and B repeat the consumption analyses shown in Table 2.9, broken down by food and non-food consumption respectively. Please refer to Table 2.9 for further details. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.12: Gap in Consumption (Main Analysis Sample), Indonesia

(a) Total Consumption

	Dependent variable: Log Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.649***	0.238***	0.084***			
	(0.012)	(0.009)	(0.014)			
Urban				0.379***	0.129***	0.012
				(0.012)	(0.008)	(0.013)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	68440	68440	68440	68440	68440	68440
Number of individuals	30751	30751	30751	30751	30751	30751

(b) Food Consumption

	Dependent variable: Log Food Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.474***	0.146***	0.066***			
	(0.012)	(0.008)	(0.013)			
Urban				0.257***	0.077***	0.003
				(0.011)	(0.008)	(0.012)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	68440	68440	68440	68440	68440	68440
Number of individuals	30751	30751	30751	30751	30751	30751

(c) Non-Food Consumption

	Dependent variable: Log Non-Food Consumption					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-agricultural employment	0.951***	0.413***	0.113***			
	(0.015)	(0.012)	(0.019)			
Urban				0.575***	0.225***	0.023
				(0.014)	(0.010)	(0.017)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Number of observations	68440	68440	68440	68440	68440	68440
Number of individuals	30751	30751	30751	30751	30751	30751

All regressions use data from the IFLS. This table repeats the analyses shown in Table 2.9 and Appendix Table A.11 using the main analysis sample, which excludes individual-year observations without earnings measures. Thus, the sample size is smaller than in Table 2.9. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table A.13: Unemployment and Job Search Behavior, Kenya

(a) Unemployment

	Dependent Variable: Unemployment or Subsistence Agriculture			Dependent Variable: Unemployment		
	(1)	(2)	(3)	(4)	(5)	(6)
Urban	-0.033***	-0.027**	-0.001	0.155***	0.141***	0.116***
	(0.012)	(0.012)	(0.021)	(0.008)	(0.008)	(0.012)
Individual fixed effects	N	N	Y	N	N	Y
Control variables and time FE	N	Y	Y	N	Y	Y
Mean dependent variable	0.297	0.297	0.297	0.080	0.080	0.080
Number of observations	10917	10917	10917	10917	10917	10917
Number of individuals	6794	6794	6794	6794	6794	6794

(b) Search Behavior

	Dependent variable: Total Hours Job Search		
	(1)	(2)	(3)
Urban	1.242***	1.216***	1.792***
	(0.144)	(0.150)	(0.266)
Individual fixed effects	N	N	Y
Control variables and time FE	N	Y	Y
Mean dependent variable	1.845	1.845	1.845
Number of observations	10917	10917	10917
Number of individuals	6794	6794	6794

Panel A reports urban gaps in unemployment. The first three columns define an individual as being unemployed if they are searching for work and have no income from wage or salary employment. The second three columns define an individual as being unemployed if they are searching for work and have no income from wage, salary, or proceeds from subsistence agriculture reported in the agricultural module. Sample sizes differ from analysis of wage gaps because questions about job search are contemporaneous to the time of the survey and are not retrospective. The dependent variable in Panel B is the number of hours a person reports to be searching for work; this variable equals 0 if the person is not searching for work. Like Panel A, data was only collected on search behavior contemporaneous to the time of the survey and thus sample sizes are smaller. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.14: Alternative Coefficient Standard Error Estimation

(a) Indonesia

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.703***	0.351***	0.252***	0.078***				
	(0.013)	(0.012)	(0.018)	(0.021)				
	[0.013]	[0.012]	[0.018]	[0.021]				
	⟦0.013⟧	⟦0.012⟧	⟦0.018⟧	⟦0.021⟧				
	{0.012}	{0.012}	{0.018}	{0.021}				
	⟨0.013⟩	⟨0.012⟩	⟨0.018⟩	⟨0.021⟩				
Urban					0.537***	0.227***	0.028**	0.002
					(0.012)	(0.010)	(0.013)	(0.014)
					[0.012]	[0.010]	[0.013]	[0.014]
					⟦0.012⟧	⟦0.010⟧	⟦0.013⟧	⟦0.014⟧
					{0.012}	{0.010}	{0.013}	{0.015}
					⟨0.012⟩	⟨0.010⟩	⟨0.013⟩	⟨0.014⟩
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	275600	275600	275600	275600	275600	275600	275600	275600
Number of individuals	31843	31843	31843	31843	31843	31843	31843	31843

(b) Kenya

	Dependent variable: Log Earnings							
	(1)	(2)	(3)	(4) Log Wage	(5)	(6)	(7)	(8) Log Wage
Non-agricultural employment	0.784***	0.551***	0.284***	0.061				
	(0.063)	(0.058)	(0.090)	(0.106)				
	[0.063]	[0.058]	[0.091]	[0.107]				
	⟦0.063⟧	⟦0.058⟧	⟦0.091⟧	⟦0.107⟧				
	{0.062}	{0.060}	{0.093}	{0.105}				
	⟨0.063⟩	⟨0.058⟩	⟨0.090⟩	⟨0.106⟩				
Urban					0.862***	0.683***	0.263***	0.165***
					(0.039)	(0.035)	(0.047)	(0.050)
					[0.039]	[0.035]	[0.047]	[0.050]
					⟦0.039⟧	⟦0.035⟧	⟦0.047⟧	⟦0.050⟧
					{0.038}	{0.035}	{0.046}	{0.052}
					⟨0.039⟩	⟨0.035⟩	⟨0.047⟩	⟨0.050⟩
Individual fixed effects	N	N	Y	Y	N	N	Y	Y
Control variables and time FE	N	Y	Y	Y	N	Y	Y	Y
Number of observations	134221	134221	134221	134221	134221	134221	134221	134221
Number of individuals	4791	4791	4791	4791	4791	4791	4791	4791

Panel A uses data from the IFLS, and Panel B uses data from the KLPS. The table repeats some of the analyses shown in Tables 2.5 and 2.6 and presents cluster robust standard errors computed several ways. For each coefficient, standard errors in parentheses in row 2 are the default standard errors reported by Stata. Rows 3 and 4 in single and double square brackets, respectively, report cluster robust standard errors CR2 and CR3 (Bell and McCaffrey 2002) that correct variance matrix bias by transforming residuals (see also Cameron and Miller, 2015). Row 5 in curly braces reports block bootstrapped errors for 1,000 bootstrap samples between stars. And, Row 6 in triangular brackets reports standard errors with Young (2016) effective degrees of freedom corrections. For columns 4 and 8, the dependent variable is the log of earnings divided by hours worked. Control variables include log hours, log hours squared, age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3, 4, 7 and 8, the control variables are reduced to only age squared. Significance stars reported reflect hypothesis tests using t-statistics computed from default standard errors, *** p<0.01, ** p<0.05, * p<0.1.

Table A.15: Gap in Consumption for those Born in Rural and Urban Areas, Indonesia

(a) Indonesian individuals born in rural areas

	Dependent variable: Log Consumption		
	(1)	(2)	(3)
Urban	0.580***	0.307***	0.133***
	(0.018)	(0.013)	(0.020)
Individual fixed effects	N	N	Y
Control variables and time FE	N	Y	Y
Number of observations	53840	53840	53840
Number of individuals	22240	22240	22240

(b) Indonesian individuals born in urban areas

	Dependent variable: Log Consumption		
	(1)	(2)	(3)
Urban	0.206***	0.034***	-0.047***
	(0.020)	(0.012)	(0.015)
Individual fixed effects	N	N	Y
Control variables and time FE	N	Y	Y
Number of observations	28363	28363	28363
Number of individuals	12524	12524	12524

Both panels use data from the IFLS. Panels A and B repeat the consumption analyses shown in Table 2.9, broken down by those born in rural and urban areas respectively. Please refer to Table 2.9 for further details. Control variables include age, age squared, years of education, years of education squared and an indicator for being female. When also including individual fixed effects in columns 3 and 6, the control variables are reduced to only age squared. All regressions are clustered at the individual level. Robust standard errors are in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix B

Appendix to Chapter 3

B.1 Additional Tables

Table B.1: Covariate Balance Across ALATs

	Small ALATs		Large ALATs	
	(1) Control	(2) Difference (T-C)	(3) Control	(4) Difference (T-C)
# of rural hhds	2298	320	3218	879
	[1978]	{0.633}	[1696]	{0.407}
Share of hhds that are rural	0.809	-0.094*	0.708	-0.039
	[0.104]	{0.082}	[0.217]	{0.597}
Average hhd size	4.71	0.01	4.70	0.14
	[0.34]	{0.950}	[0.35]	{0.302}
Land per farm (hectares)	41.35	7.63	43.40	0.68
	[43.49]	{0.711}	[42.48]	{0.973}
Cropland per farm (hectares)	12.58	-3.23	6.88	2.30
	[21.46]	{0.608}	[8.15]	{0.797}
Share of farmers working with DEAg AEAs	0.082	-0.020	0.096	0.014
	[0.109]	{0.562}	[0.072]	{0.878}
Corn yield (metric tons) per hectare	2.14	0.16	1.97	0.26
	[1.49]	{0.748}	[0.90]	{0.601}
Share of farms with running water	0.466	0.035	0.476	-0.051
	[0.247]	{0.673}	[0.210]	{0.490}
Colorado (winner) vote share	0.455	-0.007	0.472	-0.023
	[0.093]	{0.802}	[0.083]	{0.544}
Number of ALATs	17	21	22	11
p-value from Joint Test		0.299		0.178

"Control" and "Treatment" for small ALATs refer to ALATs that received cell phones in round 3 and rounds 1 and 2, respectively. The first three variables come from the 2002 census, the next five come from the 2008 agricultural census, and the final variable comes from the 2013 presidential elections. The number of AEAs and ALATs in columns (1) and (3) correspond to the respective numbers in the control group. Those in columns (2) and (4) correspond to the respective numbers in the treatment group. The joint test in the bottom row runs a regression of treatment assignment on all listed covariates. The joint test p-value is from a wild bootstrapped F-test imposing the null hypothesis that all coefficients equal zero. Standard deviations reported in square brackets and p-values from a Wu (1986) wild bootstrap procedure with 100,000 replication draws reported in curly braces. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table B.2: Average Effects of Receiving a Cell Phone on Other Measures of Performance

	(1) Satisfied	(2) Received Training	(3) Useful	(4) PCA
Treated	0.075	0.051	0.034	0.145*
	(0.047)	(0.030)	(0.022)	(0.079)
	{0.128}	{0.144}	{0.148}	{0.096}
Selected	-0.073	-0.015	-0.017	-0.078
	(0.051)	(0.023)	(0.021)	(0.071)
	{0.217}	{0.554}	{0.444}	{0.329}
Male	0.060	-0.002	0.011	0.051
	(0.057)	(0.030)	(0.025)	(0.088)
	{0.378}	{0.969}	{0.672}	{0.613}
Age	0.003	0.002	0.003*	0.008
	(0.002)	(0.002)	(0.001)	(0.004)
	{0.215}	{0.175}	{0.071}	{0.109}
Married	-0.063	-0.022	-0.024	-0.094
	(0.037)	(0.029)	(0.021)	(0.074)
	{0.107}	{0.494}	{0.295}	{0.245}
Distance to Farmers (log)	0.005	-0.027	-0.024	-0.059
	(0.053)	(0.027)	(0.023)	(0.081)
	{0.937}	{0.373}	{0.326}	{0.518}
Incumbent Party	-0.153**	-0.038	-0.032	-0.173**
	(0.051)	(0.024)	(0.025)	(0.074)
	{0.023}	{0.124}	{0.232}	{0.039}
Digit Span	0.027	0.002	0.006	0.026
	(0.015)	(0.011)	(0.010)	(0.030)
	{0.114}	{0.884}	{0.577}	{0.442}
Big 5 — Stability	-0.007	0.000	0.010	0.008
	(0.019)	(0.012)	(0.008)	(0.031)
	{0.690}	{0.967}	{0.221}	{0.796}

Continued on next page

Table B.2 – *Continued from previous page*

	(1) Satisfied	(2) Received Training	(3) Useful	(4) PCA
Big 5 — Plasticity	-0.034	-0.013	-0.012	-0.054
	(0.024)	(0.012)	(0.012)	(0.038)
	{0.211}	{0.282}	{0.342}	{0.198}
Servicer	AEA	AEA	AEA	AEA
Mean of Control Dep. Var	2.542	0.765	0.817	-0.437
R^2	0.025	0.017	0.025	0.017
Number of Phone Surveys	1838	1841	1841	1838
Number of AEAs	136	136	136	136
Number of ALATs	71	71	71	71
RI p-value	.183	.199	.181	.166
Includes Small ALATs	✓	✓	✓	✓

The outcome measure in the fourth column is the first principle component from a polychoric PCA of the outcome variables in the first three columns. These regressions include unreported indicators for small ALATs, survey wave, and an interaction of the two. Cluster robust standard errors and p-values from a Wu, 1986 wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.

Table B.3: Correlation Matrix of Performance Measures

	(1)				
	Share visited	Satisfied	Recieved Training	Useful	PCA
Share Visited	1				
Satisfied	0.311**	1			
Received Training	0.375***	0.522***	1		
Useful	0.372***	0.495***	0.783***	1	
PCA	0.417***	0.762***	0.913***	0.889***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The sample includes only AEAs in the control group.

Table B.4: Predictors of Productivity in the Control Group

	(1)
Male	0.040
	(0.040)
	{0.581}
Age	0.002
	(0.002)
	{0.539}
Married	0.070
	(0.037)
	{0.136}
Average Distance to Farmers (log)	0.019
	(0.034)
	{0.638}
Incumbent Party	0.014
	(0.031)
	{0.661}
Digit Span	0.032**
	(0.013)
	{0.010}
Big5 — Stability	0.008
	(0.016)
	{0.726}

Continued on next page

Table B.4 – *Continued from previous page*

	(1)
Big5 — Plasticity	-0.010
	(0.010)
	{0.343}
R^2	0.018
p-value for Model	0.165
p-value for Model (Not Wild)	0.000
Number of Phone Surveys	1091
Number of AEAs	68
Number of ALATs	22

The sample is all AEAs in the control group in large ALATs. Regressions also include survey wave indicators. The p-value for the model reports the implied p-value from a wild bootstrapped F-test for the null that all reported coefficients are equal to zero. Cluster robust standard errors and p-values from a Wu, 1986 wild bootstrap procedure with 100,000 replication draws reported in parentheses and curly braces, respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, corresponding to the bootstrapped p-values.