The London School of Economics and Political Science

Infrastructure and Structural Transformation: Evidence from Ethiopia

Niclas Moneke

A thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of Doctor of Philosophy.

London, June 2020.

To two journeys.

Declaration

I certify that the thesis I have presented for examination for the degree of Doctor of Philosophy is solely my own work. The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. The thesis may not be reproduced without the prior written consent of the author. I warrant that this authorisation does not, to the best of my knowledge, infringe the rights of any third party. I declare that this thesis consists of approximately 24,000 words.

Acknowledgements

First and foremost, I owe Robin Burgess my gratitude for providing me with outstanding advice and guiding my path for almost a decade of my life. Ever since our first conversation, I kept discovering a whole new world.

For similarly excellent advice, comments, discussions and continuous support, I thank Oriana Bandiera, Gharad Bryan and Daniel Sturm. Rocco Macchiavello and Fabian Waldinger got me where I needed to be, for which I thank them wholeheartedly.

I benefited greatly from conversations with Tim Besley, Francesco Caselli, Doug Gollin, Michael Greenstone, Vernon Henderson, Gabriel Kreindler, David Lagakos, Rocco Machiavello, Ted Miguel, Bart Minten, David Nagy, Imran Rasul, Martin Rotemberg, Sahele Tamiru Fekede and Fabian Waldinger. This project would not have been possible without their insight.

Furthermore, Oriana Bandiera, Robin Burgess, Francesco Caselli, Steve Pischke and Thomas Sampson taught me how to teach and to teach well.

For being amazing colleagues and friends – while fighting early battles, crossing the desert, out in the field, or during the endgame – I thank: Karun Adusumilli, Shan Aman-Rana, Michel Azulai, Jan David Bakker, Clare Balboni, Andrès Barrios, Laura Castillo, Svetlana Chekmasova, Alexia Delfino, Dita Eckardt, Miguel Espinosa, Simon Franklin, Friedrich Geiecke, Tillman Hönig, Felix König, Eui Jung Lee, Clement Minaudier, Guiseppe Rossitti, Claudio Schilter, Chiara Sotis, Torsten Figueiredo Walter, Marc Witte and Céline Zipfel.

My family and friends know exactly what they mean to me and how much I owe to them. They should be proud of themselves for somehow managing to keep me sane, or so I hope.

Financial support from the Department of Economics, STICERD and the IGC is gratefully acknowledged.

Finally, and to end where I started, I would like to thank Robin Burgess and Gharad Bryan for the honour to teach the very course that motivated me to become a researcher myself. It was an inspiration, then and now.

Abstract

Roads are instrumental to market access. Electricity is a key technology for modern production. Both have been widely studied in isolation. In reality, infrastructure investments are commonly bundled. How such big push infrastructure investments interact in causing economic development, however, is not well understood. To this end, I first develop a spatial general equilibrium model to understand how big push infrastructure investments may differ from isolated investments. Second, I track the large-scale road and electricity network expansions in Ethiopia over the last two decades and present causal reduced-form evidence confirming markedly different patterns: access to an all-weather road alone increases services employment, at the expense of manufacturing. In contrast, additionally electrified locations see large reversals in the manufacturing employment shares. Third, I leverage the model to structurally estimate the implied welfare effects of big push infrastructure investments. I find welfare in Ethiopia increased by at least 11% compared to no investments, while isolated counterfactual road (electrification) investments would have increased welfare by only 2% (0.7%).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Economic development is strongly associated with structural transformation out of agriculture.[1] A long literature has studied specific infrastructure expansions as potential drivers of development and structural transformation.[2] In reality, infrastructure expansions are commonly bundled or tightly sequenced: famous examples include the New Deal, the Tennessee Valley Authority (TVA), the Soviet State Commission for Electrification (GOELRO) or the most recent Chinese Belt and Road Initiative (BRI).[3] How such

[1] Documented by Lewis (1954), Nurkse (1953), Schultz (1953) and Rostow (1960), this association was confirmed empirically by Kuznets (1973); cf. Figure (8.1) for contemporary descriptive evidence.

[2] Krugman (1991) and Krugman and Venables (1995) highlight transport infrastructure as driver of industrialisation. Contributions on its development effects include Michaels (2008), Banerjee, Duflo and Qian (2012), Faber (2014), Donaldson (2018) and Asher and Novosad (forthcoming). Other isolated infrastructure analyses study e.g. electrification (cf. Dinkelman (2011), Lipscomb, Mobarak and Barham (2013), Rud (2012), Burlig and Preonas (2016), Fried and Lagakos (2017), Kassem (2018)), schools (cf. Duflo (2001)) or dams (cf. Duflo and Pande (2007)).

[3] New Deal: interstate highways, public buildings, tunnels, bridges, airports, rural electrification; TVA: electrification, dams, roads, canals, libraries; Soviet GOELRO: power plants, roads, large-scale industrial complexes; Chinese BRI: roads, railroads, ports, electric supergrids, industrial zones.

combinations of infrastructure investments interact, however, is not well understood.[4]

This paper asks how the interaction of infrastructure investments affects economic development. I study the large-scale road and electricity network expansions in Ethiopia over the last two decades – a recent prime example of rapid big push infrastructure investments in a low income country. I provide new evidence that the interaction of two particular kinds of large-scale infrastructure investments matters for structural transformation and welfare in a low income country.[5] Such interactions may be crucial to understand the contested effects of electrification.

First, I develop a spatial general equilibrium model with many locations and multiple production sectors and expose the economy to two distinct, possibly interacted infrastructure investments: road construction (which decreases trade costs for all tradeable sectors) and electrification (which only benefits production of the 'modern' sectors, i.e. manufacturing and services). As shown elsewhere, previously remote locations that gain a new road lose manufacturing employment (Faber, 2014; Baum-Snow, Henderson, Turner, Zhang & Brandt, 2018). In contrast, I show how locations' road connection combined with electrification allows manufacturing employment to recover. Therefore, big push infrastructure can exhibit markedly different structural transformation patterns than isolated infrastructure investments.

[4] A notable exception is Kline and Moretti's (2014) study of the long-term implications of the TVA.

[5] In line with a long literature in macroeconomics (Herrendorf, Rogerson & Valentinyi, 2014), I define structural transformation as the reallocation of employment across sectors of the economy.

Second, in order to test these predictions empirically, I provide new, geo-identified data on the rapid, big push infrastructure expansion drive in Ethiopia: over the course of only two decades, the road network quadrupled, whereas the electric grid doubled in extent.[6] I track the roads and electricity network expansions over time and across space, and link this infrastructure data with information on local economic activity from country-wide household surveys. This allows me to analyse how locations' change in infrastructure access translates into structural transformation and welfare. I provide evidence on how roads alone and roads interacted with electrification give rise to opposing structural transformation patterns in newly connected locations.

Third, I take these reduced-form moments to the model and develop a structural estimation procedure to estimate the aggregate and welfare effects of big push infrastructure. I do so by estimating a new elasticity, i.e. the elasticity of manufacturing and services productivity with respect to electrification, which I then feed back to the baseline-calibrated model to estimate counterfactual road and/or electricity investment schemes and their effects on welfare.

Methodologically, to show how asymmetric infrastructure investments from roads and electrification can amplify heterogeneity in sectoral employment across space, my theoretical framework features: Ricardian interregional trade (Eaton & Kortum, 2002), to capture a rich geography of heterogeneous locations; general equilibrium implications of road investments

[6]The second-most populous country in Sub-Saharan Africa, Ethiopia currently has a population of approximately 105 million, covering an area approximately the same as France and Spain (or: California and Texas) combined. During the period of big push infrastructure investments, the landlocked country experienced dramatic economic development and poverty reductions: the share of the population living on less than \$1.90 per day (in 2011 PPP) fell from 55% in 1999 to 27% in 2016 (World Bank, 2016).

via changes in trade costs that lead labour to reallocate (Allen & Arkolakis, 2014; Redding, 2016); general equilibrium implications of electrification via its differential effect on productivity across sector-location pairs (Bustos, Caprettini & Ponticelli, 2016); and, finally, changes in sectoral employment as outcome of interest that captures the underlying infrastructure-induced effects (Michaels, Rauch & Redding, 2011). Intuitively, the combination of heterogeneous stochastic productivity draws, consumers' love of variety for tradeables and heterogeneous trade links creates a heterogeneous ('core–periphery') allocation of labour across space (Redding, 2016). The interaction of big push infrastructure investments amplifies this heterogeneity in previously understudied ways, although such combinations of infrastructure shocks are empirically common.

A key identification challenge is that infrastructure investments are likely endogenously allocated with respect to sectoral employment or growth. The extremely high cost of such investments in low income countries demand conscious allocation decisions, for example by targeting locations with the highest growth potential first.[7] Therefore, ordinary least squares estimation of the effects of infrastructure allocation are more likely than not biased.[8]

Facing two potentially endogenous infrastructure investments with respect to sectoral employment outcomes across time and space, I develop two instrumental variables to overcome these endogeneity concerns: for electrification, I

[7]For example, a single electric substation required to step down high transmission voltages to medium and low distribution voltages cost approximately $25m in Ethiopia in 2016. A single kilometre of 132kV transmission line cost approx. $200k and a single kilometre of two-lane asphalt road approx. $500k.

[8]Similarly, one would expect difference-in-differences estimators (where both parallel trends and stable unit of treatment value assumptions are not unlikely to be violated) to be biased.

exploit locations' proximity to straight transmission lines that connect newly opening hydropower dams with Addis Abeba.[9,10] Intuitively, the electrification instrumental variable makes use of the fact that the likelihood of a location getting electrified increases dramatically from the exogenous year of dam opening onwards if that location happens to lie along a straight line between major sources of supply and demand.[11] For the road network expansion, I exploit locations' orthogonal distance to Italian colonial road arteries. These historic trunk roads were drawn freely by Mussolini himself to conquer and occupy all of Ethiopia's ancient kingdoms, starting from Asmara (in today's Eritrea) and Mogadishu (in today's Somalia). Although most of these colonial roads from the 1930s deteriorated, hundreds of small bridges across streams and rivers remained, from which reconstruction of the Ethiopian all-weather road network re-started in the 1990s.[12] Temporal variation in the roads instrument can be generated after realising, firstly, that road construction falls under the authority of the eleven regional governments in Ethiopia and, secondly, that regions only had limited resources to build. Therefore, I construct an algorithm that determines, for each region, all locations' orthogonal distances to the Italian colonial straight line, calculates regions'

[9]Dams were historically selected for construction according to their geographic suitability, not according to which places lie along the, on average several hundred kilometre long, path to the capital.

[10]The dam openings employed for my identification strategy constitute approximately 75% of total generation capacity in 2016, while overall electricity generation in Ethiopia is 98% hydro-powered. Electricity demand is geographically focused in Addis, which demands in excess of 80% of electricity supply.

[11]Akin to many major infrastructure projects, dam commissioning time deviates widely from plan, with even experts from the managing utility, Ethiopian Electric Power, unable to predict delays.

[12]The large number of bridges and crossings was made necessary due to the arbitrary, several mountain ranges-crossing routing drawn up by Mussolini, which Italian construction followed remarkably closely despite its apparent disconnect with reality in terms of the adverse terrain.

annual budget and then proceeds in building out this budget until the annual mileage allocation has been reached. First stages are strong and robust throughout.[13]

While the reduced-form provides estimates on how changes in infrastructure relate to changes in structural transformation, the aggregate effects implied by these causal differences are likewise of interest. To this end, I inform the spatial general equilibrium model with the reduced-form moments and structurally estimate the aggregate and welfare effects of big push infrastructure against counterfactual investments. I develop a five-step structural estimation procedure: first, to link road investments to changes in trade costs, a model object, I measure effective, terrain-adjusted distances from each location to each other in my sample of 689 Ethiopian districts.[14] Alluding to spatial arbitrage, I then estimate trade costs from price gaps between origin-destination pairs of barcode-level goods, from which I can derive an elasticity of trade with respect to distance for all goods. Second, I calibrate the model on baseline observables to obtain baseline sectoral productivities. Third, I set up a moment condition based on the reduced-form estimates of how infrastructure investments affect employment across sectors and make a functional form assumption about how productivities in manufacturing and services are affected by electrification. Fourth, I numerically solve the baseline-calibrated model forward until the moment condition holds in terms of the model's endogenous variables, such as sectoral employment shares (Faber &

[13]First stages and 2SLS results using three instruments (roads IV, electricity IV and their interaction) for the two endogenous variables (roads and the roads and electricity interaction) are qualitatively similar.

[14]Districts cover, on average, an area of appox. 40 by 40km with a population of approx. 150,000.

Gaubert, 2016). This step allows me to estimate a new object: the elasticity of manufacturing and services productivity with respect to electrification. Fifth, I structurally estimate welfare under big push infrastructure. Given the electrification elasticity, I can also estimate roads-only and electrification-only counterfactuals.

In the reduced-form, I find starkly different patterns of big push infrastructure on sectoral employment compared to only road investments: roads alone cause services employment to increase at the expense of agriculture and, especially, manufacturing employment. In contrast, the interaction of roads and electrification causes a strong reversal in manufacturing employment. This big push infrastructure effect on sectoral employment appears material since only households in big push infrastructure locations report significantly increased household expenditure and higher real consumption, as proxies for income (Deaton, 2003) and economic growth (Young, 2012), respectively.

The structural estimation provides an additional result: that big push infrastructure investments appear to exhibit aggregate welfare effects that are approximately an order of magnitude larger than those arrived at by isolated counterfactual investments of only roads or only electrification. This finding is particularly interesting in light of recent puzzling evidence in the electrification literature: whereas studies aimed at estimating aggregate effects of electrification find large, transformative effects on economic development (cf. Lipscomb et al. (2013), Rud (2012) and Kassem (2018)), studies aimed at estimating its microeconomic effects find consistently very small or virtually zero effects (cf. Lee et al. (2014), Lee, Miguel and Wolfram (2016) and Burlig and Preonas (2016)). My paper adds a new insight: that interactions of infrastructure investments can give rise to potentially large effects on economic development. In my particular context, the combination of market access provided by roads infrastructure and the positive productivity effect of electrification on non-agricultural production is key.

I provide further reduced-form evidence on the underlying channels of effects with respect to: heterogeneity across space, occupation- and industry-level patterns of structural transformation, distinct demographic profile changes in the labour force in big push infrastructure locations, and further suggestive evidence on the potential underlying modernisation in such locations.

For example, the strong spatial heterogeneity in response to infrastructure shocks predicted by the model is directly confirmed in the reduced-form: districts close to larger towns see the largest adverse manufacturing employment effects, whereas more remote places appear relatively shielded due to transport cost remaining high (cf. Behrens, Gaigné, Ottaviano and Thisse (2006)). In line with the model, it is also the former locations that disproportionately benefit from electrification.

Closer inspection of the structural estimation results on welfare provides intuition on why big push infrastructure investments matter: in counterfactuals without electrification, road-receiving locations almost exclusively belong to the pool of previously peripheral locations with low manufacturing and services productivity vis-à-vis the core, such that welfare gains from integration are modest. Similarly, electrification alone, under a baseline road network of late 1990s extent, mostly increases productivity in remote locations with extremely high transport costs. Hence, although some positive welfare effects driven by local demand are predicted, electrified locations miss out on other regions' increased import demand for their newly electrified manufacturing varieties. Only the interaction of infrastructure investments reaps both sources of welfare gains.

The remainder of this thesis is organised as follows: Chapter 2 develops a simple spatial general equilibrium model. Chapter 3 introduces the empirical context in Ethiopia and describes the data. I then present my reduced-form empirical strategy (Chapter 4) and the reduced-form results (Chapter 5). Chapter 6 details the structural estimation strategy, provides welfare results and studies policy counterfactuals. Chapter 7 concludes.

Chapter 2

Spatial General Equilibrium Model

To guide the empirical analysis throughout this paper, I present a spatial general equilibrium model characterised by the following broad features: firstly, locations differ in their productivity, geography and trade links with each other, as in a multi-region Ricardian trade setup à la Eaton and Kortum (2002). Secondly, road investments are assumed to have general equilibrium effects via trade costs, the reallocation of labour across space and the resulting changes in trade across (many) locations as in Allen and Arkolakis (2014) and Redding (2016). Third, electrification investments are assumed to have general equilibrium effects via productivity, similar to models of differential productivity shocks across space such as Bustos et al. (2016). Lastly, I assume the economy to consist of multiple sectors of production such that changes in sectoral employment across locations (i.e. spatial structural transformation) capture an outcome of interest as in Michaels et al. (2011) and Eckert and Peters (2018).

2.1 Setup

My theoretical framework follows the spatial general equilibrium model of structural transformation proposed by Michaels et al. (2011), which combines the canonical Helpman (1998) model with an Eaton and Kortum (2002) structure of Ricardian inter-regional trade.[1] I extend this framework by adding non-tradeable services as a third sector of the economy, which, as I show below, captures both a theoretically and empirically relevant aspect of the economy.[2] Furthermore, I expose the economy to two distinct, spatially-varying, but potentially interacted shocks: a trade cost reduction from new roads and a productivity shock affecting non-agricultural sectors of the economy in newly electrified locations.

A geography in my setting consists of many locations, $n \in N$, of varying land size (H_n) and endogenous population (L_n). Consumers value consumption of agricultural sector final goods, C^T, manufacturing sector final goods, C^M, services, C^S, and land, h, (which one may call 'housing'). Utility of a representative household in location n is assumed to follow an upper tier Cobb-Douglas functional form over goods and land consumption, scaled by a location-specific amenity shock η_n:

$$U_n = \eta_n C_n^{\alpha} h_n^{1-\alpha} \tag{2.1}$$

I assume $0 < \alpha < 1$. The goods consumption index is defined over consumption of each tradeable sector's composite good and services:

[1] Uy, Yi and Zhang (2013) provide a related model of structural change in a setting of Ricardian international trade.

[2] Desmet and Rossi-Hansberg (2014), Coşar and Fajgelbaum (2016) and Nagy (2017) provide alternative two-sector models that likewise address questions of spatial development and structural change.

$$C_n = \left[\psi^T \left(C_n^T\right)^\rho + \psi^M \left(C_n^M\right)^\rho + \psi^S \left(C_n^S\right)^\rho\right]^{1/\rho} \tag{2.2}$$

I follow a long macroeconomic literature on structural transformation and assume consumption of sectoral composite goods to be complementary, i.e. $0 < \kappa = \frac{1}{1-\rho} < 1$. As highlighted by Michaels et al. (2011), the upper-tier Cobb-Douglas and middle-tier CES utility formulation admits both prominent sources of structural transformation proposed in the macroeconomic literature: differential productivity growth across sectors (cf. Baumol (1967) and Ngai and Pissarides (2007)) as well as non-homothetic preferences that embody Engel's law of an income elasticity of demand below one in food-producing sectors such as agriculture (cf. Matsuyama (1992), Kongsamut, Rebelo and Xie (2001) and Herrendorf, Rogerson and Valentinyi (2013)).

Consumers exhibit love of variety for both tradeable sectors' goods, C^T and C^M, which I model in the standard CES fashion, where n denotes the consumer's location and i the producer's location, whereas j is a measure of varieties. Consumption of each tradeable sector's good is defined over a fixed continuum of varieties $j \in [0, 1]$:

$$C_n^T = \left[\sum_{i \in N} \int_0^1 \left(c_{ni}^T(j)\right)^\nu dj\right]^{\frac{1}{\nu}} \tag{2.3}$$

where I assume an elasticity of substitution across varieties, ν, such that varieties within each sector are substitutes for each other, $\sigma = \frac{1}{1-\nu} > 1$. An equivalent formulation, integrated over a continuum of M-sector varieties

$c_{ni}^{M}(j)$, yields manufacturing sector goods consumption, C_n^M. Equation (2.4) provides the classic Dixit-Stiglitz price index over traditional sector goods:[3]

$$P_n^T = \left[\sum_{i \in N} \int_0^1 \left(p_{ni}^T(j) \right)^{1-\sigma} dj \right]^{\frac{1}{1-\sigma}} \tag{2.4}$$

On the production side, firms in a given location and tradeable sector produce varieties for consumption in (potentially) many other locations. Production of varieties in both tradeable sectors uses labour and land as inputs under constant returns to scale subject to stochastic location–sector specific productivity draws.

$$Y_n^T = z^T \left(\frac{L_n^T}{\mu^T} \right)^{\mu^T} \left(\frac{h_n^T}{1-\mu^T} \right)^{1-\mu^T} \tag{2.5}$$

$$Y_n^M = z^M \left(\frac{L_n^M}{\mu^M} \right)^{\mu^M} \left(\frac{h_n^M}{1-\mu^M} \right)^{1-\mu^M} \tag{2.6}$$

where $0 < \mu^T, \mu^M < 1$ and, z^K denotes the sector-location-specific realisation of productivity z for variety j in sector $K \in \{T, M\}$ and location n. Following Eaton and Kortum (2002), locations draw sector-specific idiosyncratic productivities for each variety j from a Fréchet distribution:

$$F_n^T(z^T) = e^{\left(-A_n^T z^T\right)^{-\theta}} \tag{2.7}$$

$$F_n^M(z^M) = e^{\left(-A_n^M z^M\right)^{-\theta}} \tag{2.8}$$

It follows from the properties of the Fréchet distribution that the scale parameters, A_n^T and A_n^M, govern the average sectoral productivity in location n

[3]The manufacturing sector's Dixit-Stiglitz price index, P_n^M, follows an equivalent formulation integrated over $p_{ni}^M(j)$

across all varieties, since, for example, larger values of A_n^K decrease $F_n^K(z^K)$ and thus increase the probability of higher productivity draws, z^K, for all tradeable sector varieties $K \in \{T, M\}$ in region n. The shape parameter, θ, determines the variability of productivity draws across varieties in a given location n, with lower θ values implying greater heterogeneity in a location's productivity across varieties. Since the empirical application focuses on sector-location specific average productivity shocks, I assume the shape parameter, θ, to be the same across sectors and locations.

Trade in both sectors' final goods is costly and trade costs are assumed to follow an iceberg structure: more goods have to be produced at origin since parts 'melt away' during transit to its intended destination location for consumption. I denote trade costs between locations n and i as d_{ni}, such that quantity $d_{ni} > 1$ has to be produced in i for one unit to arrive in n. By assumption, within-region consumption of locally produced goods does not incur trade costs, i.e. $d_{nn} = 1$. I also assume that trade costs are the same across sectors ($d_{ni}^T = d_{ni}^M$), symmetric ($d_{ni} = d_{in}$) and that a triangle inequality holds between any three regions i, n, o, $d_{ni} < d_{no} d_{oi}$.

Given perfect competition in both production sectors, the price of a given T-sector variety, $p_{ni}^T(j)$, equals marginal cost, weighted by factor shares, inverse productivity and trade costs:

$$p_{ni}^T(j) = \frac{w_i^{\mu^T} r_i^{1-\mu^T} d_{ni}}{z_i^T(j)} \tag{2.9}$$

Similarly standard, relative factor demand equals inverse, factor share-weighted, factor prices, where transport cost cancel out due to the symmetric overuse of factors:

$$\frac{h_i^T}{L_i^T} = \frac{(1-\mu^T)}{\mu^T}\frac{w_i}{r_i} \tag{2.10}$$

Given Fréchet-distributed productivity shocks per variety (and location), each location (n) will buy a given variety from its minimum-cost supplier location (i):

$$p_{ni}^T(j) = min\{p_i^T(j); i \in N\} \tag{2.11}$$

Eaton and Kortum (2002) show how such a characterisation of prices and origin-destination trade between locations i and n in varieties j gives rise to a formulation of the share of expenditure destination location n spends on agricultural sector (and equivalently manufacturing sector) final goods produced in origin i:

$$\pi_{ni}^T = \frac{A_i^T \left(w_i^{\mu^T} r_i^{1-\mu^T} d_{ni}\right)^{-\theta}}{\sum_{k\in N} A_k^T \left(w_k^{\mu^T} r_k^{1-\mu^T} d_{nk}\right)^{-\theta}} \tag{2.12}$$

In this gravity-style equation, the traditional sector's shape parameter, θ, which governs the heterogeneity of within-location productivities across varieties, determines the elasticity of trade with respect to production and trade costs.

Production of non-tradeable services also uses labour and land as inputs, but output is a single homogeneous 'services good':

$$Y_n^S = A_n^S \left(\frac{L_n^S}{\mu^S}\right)^{\mu^S} \left(\frac{h_n^S}{1-\mu^S}\right)^{1-\mu^S} \tag{2.13}$$

Throughout, I assume agriculture to be the most and services the least land-intensive sector, $\mu^T < \mu^M < \mu^S$. Without trade in services, the non-tradeable services good's price equals marginal cost:

$$P_n^S = \frac{w_n^{\mu^S} r_n^{1-\mu^S}}{A_n^S} \tag{2.14}$$

Within each location, the expenditure share on each tradeable sector's varieties and services depends on the relative (local) price of each sector's (composite) good:

$$\xi_n^K = \frac{\left(\psi^K\right)^\kappa \left(P_n^K\right)^{1-\kappa}}{\left(\psi^M\right)^\kappa \left(P_n^M\right)^{1-\kappa} + \left(\psi^T\right)^\kappa \left(P_n^T\right)^{1-\kappa} + \left(\psi^S\right)^\kappa \left(P_n^S\right)^{1-\kappa}}, K \in \{T, M, S\} \tag{2.15}$$

Since κ is assumed to lie between zero and one demand between sector goods is inelastic. Therefore, a sector's share of (goods) consumption expenditure is increasing in its relative price index.

Given the properties of the Fréchet distribution of productivities, tradeable sectoral price indices can be further simplified to arrive at expressions that only depend on factor prices, productivities and transport cost, as well as parameters. Equation (2.16) presents the simplified T-sector price index. An equivalent formulation holds for the M-sector.

$$P_n^T = \gamma \left[\sum_{k \in N} A_k^T \left(w_k^{\mu^T} r_k^{1-\mu^T} d_{nk} \right)^{-\theta} \right]^{-1/\theta} = \gamma \left(\Phi_n^T \right)^{-1/\theta} \tag{2.16}$$

where $\Phi_n^T = \sum_{k \in N} A_k^T (w_k^{\mu^T} r_k^{1-\mu^T} d_{nk})^{-\theta}$ and $\gamma = [\Gamma((\theta + 1 - \sigma)/\theta)]^{\frac{1}{1-\sigma}}$. $\Gamma(\cdot)$ denotes the Gamma function and I assume $\theta + 1 - \sigma > 0$ to ensure this function is defined. The above simplified tradeable sector price indices can in turn be used to express expenditure shares.

To arrive at a spatial equilibrium, I provide conditions for land market clearing, labour market clearing and a labour mobility condition. For an equilibrium in the land market, total income from land must equal total expenditure on land, where the latter summarises land expenditure by consumers, M-sector firms, T-sector firms and S-sector firms. I assume land is owned by goods-consuming landlords who do not otherwise supply labour. In the empirical setting of Ethiopia, where land is overwhelmingly owned by the state, one may think of landlords as equivalent to the local government which spends its income from land on goods and land consumption itself.

The land market clearing condition can be stated as follows:

$$\begin{aligned} r_n H_n = & (1-\alpha)\left[w_n L_n + r_n H_n\right] \\ & + \sum_{k \in N} \pi^T_{kn} \xi^T_k \left(1-\mu^T\right) \alpha \left[w_k L_k + r_k H_k\right] \\ & + \sum_{k \in N} \pi^M_{kn} \xi^M_k \left(1-\mu^M\right) \alpha \left[w_k L_k + r_k H_k\right] \\ & + \pi^S_{nn} \xi^S_n \left(1-\mu^S\right) \alpha \left[w_n L_n + r_n H_n\right] \end{aligned} \tag{2.17}$$

Similarly, labour market clearing requires that total labour income earned in one location must equal total labour payments across sectors on goods purchased from that location everywhere:

$$\begin{aligned} w_n L_n = & \sum_{k \in N} \pi^T_{kn} \xi^T_k \mu^T \alpha \left[w_k L_k + r_k H_k\right] \\ & + \sum_{k \in N} \pi^M_{kn} \xi^M_k \mu^M \alpha \left[w_k L_k + r_k H_k\right] \\ & + \pi^S_{nn} \xi^S_n \mu^S \alpha \left[w_n L_n + r_n H_n\right] \end{aligned} \tag{2.18}$$

Finally, and to close the model, free mobility of workers across locations implies that workers will arbitrage away any differences in real wages across locations, such that real wages across all locations must be equalised in equilibrium. In other words, the wage earned by workers in a given location after correcting for land and goods prices, as well as a location's amenity value, must be equalised:

$$V_n = \bar{V} = \frac{\alpha^{\alpha}(1-\alpha)^{(1-\alpha)}\eta_n w_n}{[P_n]^{\alpha/(1-\kappa)}\, r_n^{(1-\alpha)}}, \quad \forall n \tag{2.19}$$

where $P_n = \left(\psi^M\right)^{\kappa}\left(P_n^M\right)^{1-\kappa} + \left(\psi^T\right)^{\kappa}\left(P_n^T\right)^{1-\kappa} + \left(\psi^S\right)^{\kappa}\left(P_n^S\right)^{1-\kappa}$. Once sectoral price indices in the denominator are substituted for with equation (2.16), the equivalent M-sector formulation and equation (2.14), the labour mobility condition [eq. (2.19)] can also be expressed only in terms of productivities, trade costs and factor prices.

2.2 General equilibrium

For each location, and given parameter values $(\alpha, \kappa, \mu^T, \mu^M, \mu^S, \theta, \sigma)$, a matrix of trade costs (d_{ni}) and vectors of sectoral productivities (A_n^T, A_n^M, A_n^S), the model admits three equations for the three endogenous variables in each location: land market clearing [eq. (2.17)], labour market clearing [eq. (2.18)] and the labour mobility condition [eq. (2.19)] allow to solve for a general equilibrium of the model in terms of its core endogenous variables wages (w_n), land rental rates (r_n) and population (L_n). Michaels et al. (2011) prove existence and uniqueness for the two-sector version, which follows through to the three-sector version presented here.

The endogenous variables of interest for the empirical analysis, sectoral employment, L_n^T, L_n^M, L_n^S (or sectoral employment shares, $\lambda_n^K = L_n^K / L_n$ for each sector $K \in \{T, M, S\}$, respectively) can be derived from the unique solution for wages, rental rates and population with the help of sectoral labour market clearing. Analogous to the labour market clearing condition above, I assume that each sector's labour income has to likewise equal total sectoral labour payments on goods purchased from that location everywhere:

$$w_n L_n^T = \sum_{k \in N} \pi_{kn}^T \xi_k^T \mu^T \alpha \left[w_k L_k + r_k H_k \right] \tag{2.20}$$

$$w_n L_n^M = \sum_{k \in N} \pi_{kn}^M \xi_k^M \mu^M \alpha \left[w_k L_k + r_k H_k \right] \tag{2.21}$$

$$w_n L_n^S = \pi_{nn}^S \xi_n^S \mu^S \alpha \left[w_n L_n + r_n H_n \right] \tag{2.22}$$

As described in Section (6) below, the general equilibrium conditions may also be exploited to back out (empirically unobserved) sectoral productivities given (empirically observed) population and sectoral employment shares via calibration of the model. In contrast to Redding's (2016) hypothetical setting, I am unable to invert the model to solve for unobserved productivities (and amenities) since rental rates are generally not available in the Ethiopian context given the almost exclusively nationalised status of land ownership during my study period. Therefore, instead of inverting the general equilibrium system to determine productivities, I have to calibrate the model to back out the unique combination of sectoral productivities for each location such that the observable data constitutes a spatial equilibrium.

2.3 Numerical solution algorithm

To solve this highly non-linear system of equations, I develop an algorithm that numerically solves for the unique equilibrium values of workers, wages and rental rates. The algorithm follows an iterative procedure, consisting of an inner and outer envelope. First, for given initial guesses of workers, $L_n^{initial}$, in each location, I adjust an initial wage guess, $w_n^{initial}$ to ensure the labour market clears in each location, while simultaneously adjusting an initial rental rate guess, $r_n^{initial}$ to ensure the land market clears in each location. Once factor prices converge to clear factor markets in each location, I check for deviations from real wage equalisation (as predicted by the labour mobility condition under converged factor prices). I then adjust the initial guess of worker allocation across locations to arbitrage away any potential real wage deviations from its median until real wages are equalised everywhere.

The numerical solution provides further insights into the drivers of heterogeneity in the spatial general equilibrium system: for symmetric productivities and trade costs across sectors, sectoral employment shares converge to a constant, independent of location.

In contrast, to achieve a unique equilibrium with heterogeneous sectoral employment across locations, the above assumption of either heterogeneity in productivity across sectors within locations, or differential trade costs across sectors are sufficient. Since I aim to empirically estimate relevant effects of shocks which manifest themselves in sectoral heterogeneity across locations, I opt for the empirically more realistic assumption of heterogeneous sectoral productivities within locations, that is $A_n^M \neq A_n^T \neq A_n^S$, while trade costs faced by firms in a given location are the same across (tradeable) sectors.

2.4 Comparative statics and simulations

For the purposes of studying the effects of infrastructure investments on sectoral employment, I assume that investments in the all-weather road network decrease transport costs between locations and investments in electrification increase local manufacturing and services sector productivities in electrified locations.

Since I am interested in structural transformation as a proxy for economic development, the objects of interest are the partial derivatives of sectoral employment shares, say $\lambda_k^M = \frac{L_k^M}{L_k}$, with respect to changes in trade cost, productivity or both:

$$\frac{\partial \lambda_k}{\partial d_{ni}}, \quad \frac{\partial \lambda_k}{\partial A_n} \quad and \quad \frac{\partial^2 \lambda_k}{\partial d_{ni} \partial A_n}, \quad k \in \{i, .., N\}$$

In partial equilibrium, as previously autarkic regions gain access to market (a reduction in the iceberg trade cost d_{ni}), the pre-existing employment in the manufacturing sector (given autarky) suddenly competes with the manufacturing sector varieties from larger (and already electrified) agglomerations. Therefore, unless the initial manufacturing sector productivity draw was high, the sectoral employment share of the manufacturing sector in the newly road-connected location would be expected to fall.

However, as productivity in peripheral, road-connected locations improves following the roll-out of electrification, some manufacturing varieties become profitable for export, such that the manufacturing employment share may actually rise.

At least in partial equilibrium for a previously autarkic location, a drop in transport cost and a drop in transport cost coupled with a positive productivity shock have opposing predictions for structural transformation according to the theoretical framework, but amplify each other in already connected locations with respect to increases in the manufacturing employment share.

In general equilibrium, however, the above intuition is complicated by free worker mobility, the effects of transport improvements in one location on all other locations in the network and the changing nature of comparative advantage across varieties throughout the network following electrification in any single location. By means of simulating the numerical solution for various shocks, I provide further intuition into the general equilibrium predictions of the model regarding changes in the sectoral employment shares below.

Two graphical results present the core predictions guiding my empirical analysis below: Figure (8.9) depicts the changes in relative manufacturing employment shares resulting from a simulated change in transport cost from new roads built between 2000 and 2016 in Ethiopia, whereas Figure (8.10) depicts changes in relative manufacturing employment shares as a result of a simulated combined transport cost and electrification shock.

As highlighted in Figures (8.9) and (8.10), the sign of the change in relative sectoral employment due to either a road or a road and electrification shock depends in a highly non-linear fashion on transport-cost adjusted comparative advantage across locations. Transport-cost adjusted comparative advantage, though, changes naturally everywhere in response to either shock: if two locations, A and B, get connected via a new road, a far-away location C may lose its comparative advantage in supplying location B with a certain variety to location A. Likewise, electrifying far-away location C may reverse this situation at the expense of location A again.

Thus, a decrease in transport cost as simulated in Figure (8.9) affects the manufacturing sector's employment share both in districts of Ethiopia that are simulated to obtain a new road connection and those that are not (or already have access): the distribution of manufacturing employment changes is widely dispersed across both groups of locations, although newly connected locations see a disproportionately larger mass of manufacturing share reductions (at the expense of previously connected or unconnected locations).

Similarly, a positive productivity shock in addition to the decrease in transport cost as simulated in Figure (8.10) (akin, empirically, to a road-connected location also being electrified), also affects sectoral employment in all locations, not only newly electrified: again, sectoral employment changes in manufacturing are widely dispersed, but newly electrified locations with road access are more likely to see increases in their manufacturing employment share.

The simulation of interacted infrastructure investments in the above theoretical framework, under certain parameter settings (discussed in greater detail in Section (6)), delivers opposing results in terms of the average effects on sectoral employment shares across locations.

Such opposing simulation results mask three distinct theoretical channels at work: for a transport cost reduction $d'_{ni} < d_{ni}$ in previously remote location i, the first channel at play under heterogeneity in factor intensities across sectors (e.g. $\mu^M > \mu^T$) is Heckscher-Ohlin-type comparative advantage. Since the price index drop in the smaller location i is larger than the similar drop from integration to all other locations, location i will see in-migration, which will specialise in the more labour-intensive sector.

The second channel at play is a classical Baumol (1967) effect where labour moves out of the more productive sector everywhere after the trade cost reduction allowed a given total sectoral demand in the economy to be satisfied with less labour. Hence, if manufacturing productivity in newly connected locations is higher than that of agriculture, the manufacturing employment will decrease in all locations.[4]

A final channel at play is Ricardian comparative advantage, namely that a formerly remote location's relative sectoral productivity will determine if it will start exporting more varieties of the traditional or the manufacturing sector, with direct implications on the connected location's pattern of sectoral employment, at the expense of the location formerly exporting this variety. In general, the Heckscher-Ohlin channel will be diluted by greater trade cost across the geography, since the price index response of connection will be more muted accordingly. Which of the opposing forces of Baumol-style labour-saving and Ricardian comparative advantage prevails in determining the sectoral employment response in road-connected places, however, is a function of trade cost and productivity levels. The productivity shock of electrification has similar effects, although the direction of the Ricardian comparative advantage effect on sectoral employment depends on the magnitude of the manufacturing sector productivity increase.

[4]Given the empirically observed low employment shares of manufacturing in Ethiopia as highlighted in Figure (8.4), such a setting appears empirically likely.

Chapter 3

Empirical Context and Data

3.1 Why Ethiopia? An (almost) ideal study setting

I study the effects of big push infrastructure investments on structural transformation and economic development in the context of Ethiopia over the last two decades. Ethiopia represents a recent, prime example of rapid infrastructure expansions that, given their large scale and extent, appear worthy of designation as big push infrastructure investments. Ethiopia provides an (almost) ideal study setting for several reasons: first, the country experienced large-scale investments in two separate kinds of infrastructure, namely all-weather roads and the electricity network. Exploiting differences in the sequencing of these two infrastructure expansions allows me to study both the individual effect (of roads) and the interaction effect (of roads and electricity).

In particular, the all-weather road network expanded roughly fourfold between the late 1990s and today, from approx. 16,000km to approx. 70,000km. Figure 8.2 provides a graphical account of this expansion. My focus on all-weather roads, i.e. roads with either asphalt, bitumen or gravel surface, follows an underlying understanding that trade and market access rely on year-round accessibility (ideally by lorry) of a given location.

Over the same time period, although with a short time lag, the electricity network doubled in extent from 95 to 191 major electric substations. Figure 8.3 displays this expansion of the electricity network during my sample period. Electric substations are crucial for electrification since they are required to step down the high voltages used in long-distance overland transmission to low voltage levels used for local distribution networks. High voltage transmission lines efficiently conduct electricity over distances of several hundred kilometres between major sources of generation (such as hydropower dams) and concentrations of demand (such as cities). Local low voltage distribution networks supply individual firms, households and other end users with electricity.

Second, the almost complete lack of direct infrastructure substitutes in Ethiopia implies that the all-weather road and electricity network expansions I track capture genuine extensive margin effects of access to infrastructure. In particular, Ethiopia is a landlocked country without major navigable rivers or canals. During my study period, the single existing railway line (to neighbouring Djibouti and its port) was still out of order.[1] Another new railway project only began construction in 2015.[2]

With respect to access to energy and substitutes for grid electricity, only a handful of isolated diesel generators originating from the 1960s operated

[1] A recently completed, newly built replacement railway to Djibouti was inaugurated in October 2016. Due to equipment failures, however, commercial operations only started in January 2018.

[2] cf. International Rail Journal's news coverage in February 2015: https://www.railjournal.com/index.php/africa/work-starts-on-delayed-ethiopian-project.html

in selected major cities. All of these major cities were grid-electrified before my study period and, thus, do not feature as compliers in the instrumental variables strategy below. Self-generated energy from off-grid solar home systems generated approximately one megawatt of capacity midway through my sample (GTZ, 2009), compared to total installed grid capacity in 2018 of 4,256 MW (World Bank, 2018). An additional 25,000 solar home system panels (à 5-10 W each) were purchased by the Ethiopian government for decentralised installation by 2013.[3] Thus, due to both the low penetration and the low voltage and performance of the existing solar home systems in Ethiopia during my study period, off-grid solar cannot be regarded as a feasible substitute to grid electricity access. Other off-grid alternatives (such as mini-hydropower systems) are not known to have been present beyond isolated cases.

Third, a rich set of household surveys that cover the entire country in relatively regular intervals since the late 1990s were conducted by Ethiopia's Central Statistical Agency. At least two distinct sources of occupational choice data exist in the case of Ethiopia – both with reasonable spatial coverage, large survey sample sizes and mostly overlapping in time (see Subsection 3.2). In total, four rounds of survey data of the high quality and internationally standardised Demographic & Health Survey (DHS) are available [2000, 2005, 2011, 2016]. These repeated cross-sections of household-level (and invidiual-level) data are complemented by three rounds of the Ethiopian National Labour Force Survey (NLFS) [1999, 2005, 2013], which yields a decent coverage of my study period of interest from the late 1990s to the very recent past. An additional survey instrument covering household

[3] cf. All Africa's news coverage in August 2013: https://allafrica.com/stories/201308070099.html

expenditure and consumption, the Household Consumption and Expenditure Survey (HCES), is also available for at least four rounds [1999, 2005, 2011, 2016]. The broad spatial coverage and consistent temporal coverage across different surveys presents unusually rich non-experimental data in a low income context.[4]

Based on the above data, Figure 8.4 presents micro-founded descriptives on macroeconomic structural transformation patterns in Ethiopia that started at least during the mid-1990s, if not earlier. In particular, the share of employment in the agricultural sector declined from 89.3 per cent in 1994 to 56.6 per cent in 2016, despite population growth of approximately two per cent annually, mostly driven by rural, agrarian areas. Starting from very low levels of relative employment, services (manufacturing) employment increased from 7.6 (2.3) per cent in 1994 to 33.5 (9.9) per cent in 2016. Hence, most structural transformation in Ethiopia overall occurred from agriculture to the services sector. However, a comparison of sectoral employment to sectoral value-added trends (see Appendix Figure A1) over the same time period highlights a recent uptick in industry value-added between 2011 and 2016, which does not yet appear to result in markedly higher relative industry sector employment.

Especially if structural transformation is of a low-level nature, i.e. out of agriculture into mostly small-scale, informal retail services (see Section 5 below), positive income and welfare effects of such sectoral shifts are not obvious, neither at the individual level, nor in the aggregate. However, as shown in Figure 8.5, my study period displays an almost exploding time series of GDP per capita and a dramatic reduction in headcount poverty, using either

[4]Censi were conducted in Ethiopia in 1984, 1994 and 2007, with the planned 2017 census experiencing repeated delays. The crucial census round for my analysis, 2007, saw the occupational choice question dropped from the questionnaire.

national or international measures. Hence, even from a purely descriptive perspective, a study of large-scale infrastructure investments in relation to structural transformation seems warranted.[5] Finally, with a population of approximately 105 million people, an area roughly the size of France and Spain combined and the second largest low income country economy in Sub-Saharan Africa, the study of the Ethiopian big push infrastructure investments and the resulting structural transformation appears of interest in its own right, with potential external validity for other low income countries that plan similar transport and energy infrastructure investments.

3.2 Data

I provide new information on the electricity grid and the road network expansions in Ethiopia as the foundation of my analysis of big push infrastructure investments since the late 1990s.

Resulting from a close collaboration with Ethiopian Electric Power (EEP), the state utility charged with electricity generation and transmission, I obtained confidential information on the exact location, capacity, equipment and

[5]Common sense may deem the tracking of a large-scale infrastructure expansion in a country with a centrally located capital city (which also happens to be the country's largest, as well as its undisputed administrative, business and industry hub), as a potentially moot exercise: one could expect that economic activity, in line with population density, decreases radially from the centre, such that any reasonable least-cost infrastructure network expansion would also follow a radiating process outwards from the centre. Thus, the location and timing of expansion investments could be expressed as a function of distance to the centre. Fortunately, this hypothesis is without foundation in the case of Ethiopia: as highlighted in Appendix Figure A3, population density in Ethiopia is spread out irregularly, and also does not interact in a straightforward manner with either elevation (see Appendix Figure A4) or terrain ruggedness (see Appendix Figure A5). In short, large parts of the Ethiopian population live in highly rugged, elevated and remote locations, which do not necessarily align with either favourable natural endowments in terms of agricultural productivity, nor radial distance to the economic centre.

commissioning time of each of the electric grid's substations.[6] These records cover a total of 191 substations which came online before 2018, with the first isolated substations constructed in 1959. To reconstruct the expansion of the interconnected system ("grid"), I also obtained information on each transmission line, its location, connecting nodes, voltage and commissioning times, as well as further information regarding recent upgrades into stability-enhancing equipment (e.g. reactors and capacitators) associated with each line. To inform tentative cost-benefit calculations, I also collected construction cost estimates from the engineering team with respect to unit costs of transmission infrastructure and past records of selected actual project expenditures.

Finally, I also collected locations, capacity, operational status and commissioning time information on all power plants to track generation. The Ethiopian electricity supply is mostly provided by hydropower from nine major dams, as well as at least three wind farms, one geothermal power plant and by-generation from at least three sugar refineries. Dam openings since 2016 are currently ignored in my analysis due to the lack of outcome variables spanning this very recent past (see below).

Although the opening of a substation represents a *de facto* necessary precondition for electrification of a given location and its surrounding areas, it does, however, not perfectly capture distribution-level connections at the neighbourhood- or village-level. Therefore, I also obtained new, previously undisclosed information from Ethiopian Electric Utility (EEU) on the extent of distribution networks behind a given substation, for a large subset of substations. This information has exact geographical information on village- and

[6]Formerly a single state utility known as Ethiopian Electric Power Corporation (EEPCo), EEPCo was broken up into two separate entities in 2013: a generation and transmission utility, Ethiopian Electric Power (EEP), and a distribution utility, Ethiopian Electric Utility (EEU).

town-level electrification status. Although originally lacking exact information on the timing of distribution network expansion, I obtained complementary records on town- and village-level electrification status combined with the year of electrification. This comprehensive distribution network coverage and expansion dataset covering villages and towns inside districts is currently used for robustness of my district-level analysis (see Section 4), given limits to the spatial identification of the outcome variables.

I obtained information on the expansion of all-weather roads mostly from the Ethiopian Roads Authority (ERA). In particular, I employ several historical and present maps and geographic information system (GIS) data from various, partially undisclosed records. For the years 2006, 2012 and 2016, I have obtained GIS data and maps, which rely at least partially on actual road surveys in the period of up to one and a half years before the stated date. In particular, the final cross-section from 2016 relies on a several weeks-long on-the-ground data collection effort by ERA that verifiably mapped every road in the country, recording surface type, quality, width, current state and GPS markers at regular intervals.[7] Earlier maps were supposedly based on partial road surveys and/or records of road construction projects. However, I cannot independently verify this claim given the lack of centrally recorded road construction documentation at the project-level.

[7]The 2016 ERA road survey also contains estimates of the original year of each road's construction, which I use for cross-validation of earlier maps. In contrast, this information forms the foundation of related papers studying the expansion of the Ethiopian road network such as Adamopoulos (2018). For their more localised analysis of rural roads, Gebresilasse (2019) and Kebede (2019) also make use of earth feeder roads from the 2016 ERA survey. These latter roads are dropped from my analysis due to the explicit focus on 'all-weather' (i.e. gravel, asphalt or bitumen surface) roads.

In addition, I use various other sources for cross-validation and to obtain better visibility on the pre-sample road network: I use GIS data from OpenStreetMap for the year 2014 to cross-verify the earlier and later ERA records. I also use manually digitised historical CIA maps from 1969, 1972, 1976, 1990 to obtain the pre-sample period. The CIA's 1999 map is used as the first cross-section in the sample and the 2009 map for cross-validation of ERA records. Furthermore, I also make use of a biennial, district-level road density dataset (1996-2012) kindly provided by Shiferaw, Söderbom, Siba and Alemu (2015) for robustness checks. Changes in district road density correlate highly with the map-derived measures of district level all-weather road access I employ in the main analysis.

With respect to the outcome variables of interest, I am first and foremostly interested in structural transformation, which I interpret in line with the literature (Herrendorf et al., 2014) as changes in sectoral employment. Thus, I require information on relative employment, which I derive from two repeated household- and individual-level surveys: the Demographic & Health Survey (DHS) for Ethiopia with rounds 2000, 2005, 2011 and 2016, and the Ethiopian National Labour Force Survey (NLFS) with rounds 1999, 2005 and 2013. In particular, I use respondents' answer to questions about their 'current occupation', which I then group into three sectors, agriculture, manufacturing and services according to the International Standard Classification of Occupations (ISCO), in its ISCO-88 and the more recent ISCO-08 iterations.

Both the DHS and the NLFS are repeated cross-sections of enumeration areas (EA), with approximately 20 to 30 households enumerated per EA. Effective sample sizes for the DHS rounds amount to 12,751 individuals in

2000, 14,052 (2005), 21,080 (2011) and 19,157 (2016), from approximately 650 EAs, which differ per round. The NLFS sample sizes are on average ten times larger than the DHS, but contain greater measurement error and incomplete responses. Due to the repeated cross-section nature of the outcome variables, I aggregate individual responses to the enumeration area and then generate an (unbalanced) district panel from districts that contain at least two sampled EAs. Therefore, all of my below analyses using relative employment as dependent variable are run at the district-year level using only panel districts. Figure A2 provides an overview of the spatial and temporal coverage of DHS EAs throughout Ethiopia.

Neither the DHS, nor the NLFS samples are representative at the district level. However, both surveys' enumeration areas were sampled randomly proportional to population size (i.e. the number of households). Therefore, although results from individual districts cannot be considered representative of that district, average treatment effects of infrastructure investments are unbiased as long as deviations from random sampling of enumeration areas (conditional on population) do not systematically correlate with infrastructure allocations.[8] For example, this orthogonal sampling assumption would be violated if areas closer to a road were more likely to be drawn for enumeration due to ease of access for enumerators. Given the centralised process of drawing samples of enumeration areas by the Central Statistical Agency in Addis Abeba without any involvement of local enumerators (who would usually not yet be hired for the design phase), such a violation appears unlikely. In addition, Appendix Table A1 provides an auxiliary test if new entry into the sample

[8]Due to the nature of sampling proportional to population size, all of the below specifications, including first stages, include controls for initial population levels.

correlates with the build-out of the road and electric grid at the district level. Results suggest that as infrastructure is rolled out over time, districts that newly enter the sample in later years (cf. columns 2-4 of Appendix Table A1) do not statistically significantly correlate with treatments at the district level.[9,10]

With respect to the geo-identification of enumeration areas (and, thus, households), two qualifications are due: first, the enumeration area locations of NLFS EAs are provided in codified form, which is at times only imperfectly geographically traceable. Missing codebooks at the Ethiopian Central Statistical Agency in combination with missing old maps make cross-referencing of old codebooks to old district and enumeration area delineations for some cases close to impossible. Second, even the DHS-provided GPS coordinates for EAs locations are not perfectly reliable due to the common random displacement applied to GPS coordinates prior to publication. To ensure survey respondents' anonymity, DHS EA coordinates of rural (urban) EAs are randomly displaced within a 0-10km (0-5km) radius.[11] Therefore, although I have exact geo-

[9]Orthogonality between sample entry and treatments at the district level is a necessarily imperfect test given that the random sampling assumption is designed to hold only at the level of enumeration areas, not at the level of (significantly larger) districts. However, maps of census tracts and enumeration areas for samples enumerated before the Census in 2007 (e.g. DHS survey rounds 2000 and 2005, and NLFS survey rounds 1999 and 2005) could not be retrieved and original paper copies of these enumeration areas may not exist anymore. Therefore, the above orthogonal sampling test cannot be performed at a finer and higher resolution than the district.

[10]Column 1 of Appendix Table A1 highlights how the baseline cross-sectional sample of districts does correlate highly positively with the presence of district-level infrastructure. This positive correlation is natural when infrastructure is endogenously allocated to where economic activity (and therefore population) is concentrated. In other words, column 1 confirms the need for a quasi-experimental identification strategy, which is explained in detail in Chapter 4 below.

[11]In principle, these displacements supposedly neither cross zone borders (the second highest administrative level), nor country borders, although they may cross district borders (the third highest administrative level). In practice, however, a handful of displacement errors were corrected manually.

identified information on infrastructure placement, the finest spatial resolution the analysis can support is constrained by the available outcome variable data.

Finally, for reduced-form analyses of household expenditure responses to infrastructure investments, several of the Ethiopian Central Statistical Agency's Household Consumption and Expenditure Surveys (HCES) were obtained – in particular the 1999, 2005, 2011 and 2016 rounds. Of these, at the current point in time, I can geo-locate the 1999 and 2016 rounds. For 1999 (2016), 1,264 (2,106) enumeration areas were sampled and 17,332 (30,229) households surveyed, while, as is common for most household surveys in Ethiopia, the non-sedentary population in Afar and Somali regions were excluded. From these two geo-located survey rounds I obtain information on household expenditure per capita, household size and further household demographics. This information can then be aggregated to enumeration areas. Repeated draws of enumeration areas from the same district over time allow creation of a pseudo-panel at the district-survey-level similar to the DHS and NLFS pseudo-panels described above. In addition, the HCES contain extremely detailed information on item-by-item consumption quantities and (local market-verified) prices for all goods consumed by each household in a four day recall period for non-durables. For durables consumption, the information was enlisted using both three and twelve month recall periods. This additional information will further enrich the analysis in the future with respect to separating price from quantity effects of infrastructure investments on household consumption, as well as tracing changes in the number of varieties consumed.

Chapter 4

Empirical Strategy

4.1 Identification challenges

A key identification challenge in studying the effects of big push infrastructure on structural transformation is that infrastructure investments are likely endogenously allocated with respect to the outcomes of interest, such as changes in sectoral employment or growth. Given the extremely high cost of infrastructure investments, conscious allocation decisions are to be expected, for example by targeting high growth potential locations. Ordinary least squares estimation is more likely than not biased. Likewise, the identification assumptions underlying difference-in-differences or two-dimensional fixed effects research designs are most likely violated, i.e. parallel trends between treatment and control locations, and the stable unit treatment value assumption (SUTVA).

Several aspects of this identification challenge are relevant in the Ethiopian context: potential spatial targeting of investments across districts, temporal prioritisation of investments across districts, and natural sequencing of different investments arising from interdependence between road and electricity network investments.

A classic endogeneity concern in the allocation of infrastructure investments is that policymakers make active allocation decisions in which locations should obtain access and which not. The high cost of transmission infrastructure implies that spatial targeting of locations to receive access is an obvious feature of electrification. In particular, existing engineering guidelines highlight the primary cost driver to be minimised as the length of transmission lines. Engineers involved in the Ethiopian grid expansion have also reported privately that in order to minimise cost, during the early years of electrification only politically demanded locations would obtain a transmission line connection (and substation), apart from locations that accidentally happened to lie on a relatively straight line between supply (e.g. a hydropower plant) and demand (e.g. the major load centre(s)). Despite recent advances towards rural electrification in Ethiopia, many rural and remote locations throughout the country will probably remain without electricity access for the foreseeable future.

Another cause for endogeneity in the allocation of infrastructure investments across locations could be temporal prioritisation according to unobservables if all locations eventually obtain infrastructure access. In the particular case of roads in Ethiopia, for example, the government formulated an explicit policy to connect all of the more than 689 district capitals with an all-weather road by 2020 – an objective that was successfully achieved by 2016 already.[1] Hence, in my analysis of road network investments, a key endogeneity concern is the timing of a district's connection (in contrast to the above issue of endogenous district selection into treatment and control, since

[1] In the following, I use district capitals and district centroids interchangeably, where the latter replaces the former in districts for which information on the district capital is not available and no obvious administrative center exists.

all districts obtain road access treatment eventually). Appendix Figure A11 confirms that more densely populated districts were in fact connected to an all-weather road earlier than more rural, sparsely populated districts, pointing towards potentially endogenous connection timing.

To provide an overview of potential policy objectives that guided infrastructure allocations across districts in Ethiopia, Appendix Figure A6 provides a test of balance across treatment and control districts, which is soundly rejected: road-connected districts are significantly larger and more likely to lie in the rugged, low temperature and very fertile highlands. The right panel of Appendix Figure A6 also shows that electrification was substantially less targeted at the district level, but that mostly easier to access, less rugged districts received a connection.

Finally, at the level of transmission lines and substations, these major infrastructure items also crucially rely on at least some means of transport to be available for construction. Hence, sequencing of infrastructure investments appears natural in the context of transport and electrification. Therefore, translated to the Ethiopian context of big push infrastructure investments, a possible endogeneity concern arises from the fact that electrification only reaches previously road-connected places.

Further identification challenges arise from other sources of omitted variable bias that affect both the infrastructure expansions and structural transformation (such as natural resource windfalls, global economic cycles, capital flows, donor funding, etc.), which are entirely plausible in the Ethiopian context. Likewise, reverse causality in the form of sectoral shifts causing greater demand for infrastructure investments should also not be ruled out *ex ante*. Finally, measurement error in the right-hand side variable may lead to attenuation bias, for example due to inaccurate timing information of electric grid expansion or imprecise historic road maps.

4.2 Instrumental variables

Given the expected violation of identifying assumptions of differences-in-differences estimators, and in the absence of arbitrary policy rules generating sharp connection status discontinuities, I resort to employing an instrumental variables identification strategy.[2]

Regarding electrification, the instrumental variable (IV) is founded on the fact that electricity supply must be connected to demand, or in engineering terms: to the load centre. Translated to the Ethiopian context, electricity generation originates to 98 per cent of total installed capacity from hydropower dams in the Ethiopian highlands. The largest load centre, however, is Addis Abeba, which also hosts the load dispatch center of the interconnected system, in charge of operations management and system stability.

Therefore, I develop an IV which yields a hypothetical electrification status and timing for each location based on that location's proximity to a straight line corridor from a newly opened hydropower dam in mostly remote parts of Ethiopia to Addis Abeba. From the year of dam opening onwards, all districts lying along this straight line connecting the dam to Addis will be considered hypothetically electrified.

With respect to such an IV's identifying assumptions, the validity assumption reads that the hypothetical electrification status of districts along

[2]Unlike, for example, Asher and Novosad (2016) who exploit a dichotomous rural road targeting policy based on Indian villages' population size above some idiosyncratic threshold.

a straight line from a new dam to Addis does display a statistically significant relationship with these districts' actual electrification status and year of electrification. I draw straight line connection corridors of 25km diameter for nine dams and two large-scale wind farms.

The random assignment assumption of the IV would imply that a given district's exposure to a straight line corridor was spatially and temporally as good as randomly assigned. In other words, locations that lie between both of the straight line endpoints, which would usually span several hundred kilometres, are not systematically different from nearby locations off the straight line corridor. Appendix Figure A10 provides evidence that districts which happen to lie under a straight dam-to-Addis line are indeed not statistically significantly different from neighbouring districts (that lie just next to the instrumental variable line) across a wide array of observable covariates.[3]

Likewise, the timing of the high-voltage line coming online due to the opening of the hydropower dam should also be exogenous. Given frequent multi-year delays in these large dam construction projects, the assumption of exogenous final commissioning time appears to have merit in the Ethiopian context.[4]

Finally, the exclusion restriction requires that the straight line corridor from dam to Addis does not affect structural transformation in the years and

[3]The original instrumental variable line buffer has a width of 25km and districts which lie underneath this buffer are considered treated according to the instrument. To identify directly neighbouring districts, a counterfactual buffer of 75km width was drawn. From the resulting set of 75km buffer districts, the original (25km) districts were removed to arrive at a sample of neighbouring districts from both sides of the original line buffer. Covariate values are either time invariant or represent initial values at the beginning of the study period.

[4]In private conversations with current and former EEPCo and EEP senior engineers in charge of grid planning and expansion, i.e. local experts with decades of relevant experience, providing accurate predictions of dam construction delays was described as non-trivial.

locations exposed to the (now online) hypothetical transmission line, other than through actual electrification.

In sum, my electrification IV represents a classic 'inconsequential units' IV (cf. Redding and Turner (2014)) brought to the electrification context.[5] Figure 8.6 provides a graphical representation of the instrument and how differential proximity to straight line corridors (and their opening years) generate spatial and temporal variation in districts' hypothetical electrification status.

Regarding my instrumental variable for the timing of a district's road connection, I construct a hypothetical road expansion based on distance to straight line arteries drawn up by 1930s Italian colonial plans for road construction: In order to conquer Eritrea, Ethiopia and Somalia, as well as to effectively occupy their territory, the Italian invaders initiated a large-scale road construction effort starting in 1936. Either lacking information about the local geography and terrain or actively ignoring it, Benito Mussolini himself appears to have designed at least five major road arteries to connect the capitals of former ancient kingdoms to each other and to major ports, allowing the Italian colonial forces in theory to penetrate the hinterland of the conquered territory.

In particular, straight line axes were drawn to connect Addis Abeba, the capital of the defeated Ethiopian Empire, to both Asmara (then capital of Italian Eritrea) and Mogadishu (then capital of Italian Somaliland). In addition, the ancient kingdom capitals (and centres of regional power) of Gonder (Begemder Kingdom), Dessie (Wollo province), Nekempte (Welega province), Jimma (Kaffa Kingdom), Yirga Alem (Sidamo Kingdom) and Harar/Jijiga (Emirat of Harar/Hararghe province), as well as the Red Sea port at Assab

[5] Examples of similar instrumental variables include Michaels (2008) and Kassem (2018), who also use exposure to artificial lines to instrument for infrastructure expansions.

were to be connected either directly to one of the major capitals or on the way. The resulting straight line arteries are depicted in Figure 8.7.

Starting in 1936, actual Italian road construction followed to a surprising degree Mussolini's grand design of unrealistically straight road arteries, irrespective of the adverse terrain covered. Before their defeat at the hands of British and allied forces in the Horn of Africa in 1941, Italian colonial authorities managed to construct at least 4,000 kilometres of paved and 4,400 kilometres of unpaved roads. Appendix Figure A7 provides a historic picture of the construction efforts during the late 1930s.

On the territory of today's Ethiopia, approximately 3,378 kilometres of paved 'highways' were constructed, of which at least 1,970 kilometres were finished including state-of-the-art asphalt surfacing. Importantly, a lasting feature for future Ethiopian road construction were the 4,448 small and 128 large bridges finished by the Italian colonial authorities, artefacts necessitated by the idiosyncratic routing through the Ethiopian Highlands mass and multiple mountain ranges.[6]

For the purposes of this paper's roads IV, I exploit the fact that Ethiopian road construction in the 1990s started reconstruction of its road network from the former Italian colonial trunk network and subsequently, during the period of my study from 1999 to 2016 fanned out road access to nearby cities, towns and settlements, closely following geographic features (i.e. mostly valleys and ridges). Appendix Figures A8 and A9 provide two exemplary cases of how Ethiopian road construction connected nearby settlements and districts almost

[6]Apart from the vast Ethiopian Highlands itself (the 'roof of Africa'), of the remaining eight major mountain ranges in Ethiopia, four were crossed: the Ahmar mountains, the Entoto Mountains, the Mount Afdem range and the Semien Mountains.

orthogonally starting from the (previously reconstructed) Italian colonial roads.

I therefore construct a roads instrumental variable in the following way. Starting from the seven straight-line arteries designed by Mussolini (and depicted in Figure 8.7), I calculate orthogonal, shortest distances to every district capital, as the crow flies.[7] One should also note that this distance is calculated from the plausibly exogenous straight lines schematically drawn by the Italians, not from the actual roads that were subsequently constructed (and re-constructed) under these designs. Since road construction in the Federal Democratic Republic of Ethiopia is a politically regional matter, I run the following algorithm separately and simultaneously for each of the eleven regions of Ethiopia.[8]

Given the total length of (straight line) road connections to be built in every region to connect every district in that region, I calculate the annual mileage per region of road construction to achieve this goal of universal district road access by the end of the sample period, i.e. over a seventeen year period (2000-2016).[9] With this annual mileage goal at hand, I allow each region to build the shortest stretches of (straight line, orthogonal) district connections first

[7]Districts which contain arteries are considered already 'treated' with road access by the roads IV.

[8]For the city-regions of Addis Abeba and Harar, in which districts always had access to at least one all-weather road in 1999, the instrument (inevitably) predicts road connections in 1999 already. For the city-region of Dire Dawa, the instrument predicts both of the city's districts to be connected by 2000. The algorithm procedure to generate temporal variation therefore only generates meaningful variation for the remaining eight major regions of Ethiopia: Afar, Amhara, Beneshangul Gumuz, Gambela, Oromia, SNNPR, Somali and Tigray. Given the status of the three city-regions as 'always takers' in the frequentist sense, this prediction is expected and resonates well with empirical reality.

[9]As confirmed by the maps of the Ethiopian road network in 1999, the reconstruction and rehabilitation or re-surfacing of the original Italian road network was finished by then. Therefore, I assume that new construction started from the year 2000 onwards.

until the goal for a given year has been reached. For every subsequent year, I then re-calculate the total distance to connect each non-connected district to its closest Italian artery, derive an annual mileage target and fill this target with the shortest remaining connections.[10]

One relevant peculiarity of the above algorithm is that road distances to connect a given district are never updated: the calculated distance is always taken as the distance to the nearest Italian artery, which does not vary over time. This deliberate choice against a continuous-updating algorithm, that would calculate the shortest distance to either the Italian artery or the nearest connected district capital, arises from a potential threat to the exclusion restriction, where short district connections from district capital to district capital pick up agglomerations of population (and thus smaller district sizes). Once the district closest to the artery of such an agglomeration would be connected, the succeeding districts would be connected relatively sooner compared to an algorithm without continuous distance-updating. Therefore, to guard against this potential violation of the exclusion restriction, I do not update distances and always have the algorithm build (relatively less realistic) connections to the closest Italian artery, irrespective of any districts already connected in between.

An obvious concern with the temporal variation created according to the above algorithm (starting from the Italian colonial arteries) would be that proximity to straight line arteries would also reflect persistent historic differences in district population density, for example. Hence, closer-to-artery

[10]This updating of the annual mileage target achieves a more realistic distribution of construction activity than keeping the initial annual mileage target for all remaining sixteen years, which leads to a runaway process of road connection that is considerably faster than the actual build-out observed on the ground.

districts may be both larger and connected first, violating the instrument's exclusion restriction. To defuse this concern, Appendix Figure A12 highlights that districts which are predicted to get connected according to the instrument around sample year thresholds (e.g. for the DHS: 2000, 2005, 2011, 2016) are not statistically significantly different from each other. For example, a district predicted to be road-connected according to the instrument in 2005 would still be added to the complier population for the period 2001 (the year after the first DHS survey) to 2005 (the year of the next survey). In contrast, a district predicted to be road-connected according to the instrument in 2006 would be counted as a complier for the subsequent sample period until 2011 (the year of the third DHS survey). Therefore, random assignment of the road connection year is especially important around these sample threshold years (i.e. 2000, 2005, 2011 and 2016 for the DHS, and 1999, 2005 and 2013 for the NLFS). Appendix Figure A12 confirms that although post-sample threshold year districts are in fact marginally smaller, further from zone capitals and hotter, no statistically significant differences (which may foreshadow violations of the exclusion restriction) can be detected.[11]

In sum, my Italian artery roads IV provides temporal and spatial variation in district road access derived from a plausibly exogenous source, namely orthogonal straight line distance to Italian straight line arteries. The resulting district-level variation in the predicted arrival year of an all-weather road connection is also depicted in Figure 8.7.[12]

[11]The result in Appendix Figure A12 is not sensitive to pooling across thresholds: it likewise holds for each individual pre-/post-threshold comparison (e.g. 2000 vs 2001, 2005 vs 2006, etc.). Results are available upon request.

[12]Accordingly, the instrument takes a value of one from the district-year in which a given district got connected (as determined by the regional budget split rule) onwards.

For additional robustness, I also provide an alternative instrumental variable for a district's road connection derived solely from a hypothetical least-cost network expansion. I construct the least-cost network in the following way: construction starts from the historic Italian colonial road network, which provides a plausibly exogenous baseline all-weather road network cross-section for Ethiopia (see above). I then extend this baseline in a least-cost fashion by employing common minimum spanning tree algorithms such as Kruskal's and Boruvka's algorithms, following the explicit policy objective to connect all district capitals by the end of the sample period in 2016.

The algorithms thus provide spatial variation in terms of how each district will get connected to the network (which vary slightly across Kruskal's and Boruvka's algorithms with respect to their order), but do not yet provide any temporal variation in districts' connection timing. Therefore, I additionally employ a simple budget split rule to the output of the minimum spanning tree algorithms, such that only a certain amount of new all-weather road mileage can be built per year, following the order dictated by the least-cost algorithm. I obtain a hypothetical road network that features both spatial and temporal variation with respect to which district will get connected to the all-weather road network in which year. Appendix Figure A13 provides a graphical representation of the alternative, least-cost road IV's variation exploited in two-stage least squares robustness estimation below.

4.3 First stages and reduced-form specification

My reduced-form empirical strategy builds on two separate instrumental variables for two potentially endogenous infrastructure investments: the electricity network expansion and the road network expansion. Due to the nature of big push infrastructure investments in Ethiopia, a third endogenous variable is also of key interest, that is the interaction of both roads and electrification investments.

As shown in Table 9.2, however, my sample does not feature three genuine, independently identifiable infrastructure investments, i.e. isolated roads investments, isolated electricity investments and combined roads and electrification investments. Instead, driven by the sequencing of electrification in Ethiopia to follow the roads expansion with a short time lag, I do not observe isolated electricity investments in districts without all-weather road access.[13] Therefore, I face a situation of effectively two endogenous variables for which at least two instrumental variables are required: a roads instrument, and an instrument for the interaction between roads and electricity investment, which I construct in a standard fashion by interacting the roads IV with the electricity IV. To account for this feature, I drop the level effect for electrification in all results below.

First stage results are presented in Tables 9.3 (for the NLFS sample) and 9.4 (for the DHS sample) and show a strong and statistically significant relationship between instrumental variables and endogenous regressors. Cragg-

[13]The roads data appear erroneous for seven isolated district-year observations, for which electrification supposedly arrived before the road. All of the results below are robust to excluding these seven district-year cases of measurement error.

Donald, Sanderson-Windmeijer and classic F-test statistics all indicate non-weak instruments.[14]

Both first stages across samples include year fixed effects and a battery of initial district-level controls. These include initial district temperature mean, initial district soil quality, log distance to the nearest administrative capital, log distance to the nearest major agricultural market town, initial district satellite-derived nightlights and a district's ruggedness. Since the first stages are in theory unconstrained by the repeated-cross section sampling of the outcome variables below, I test for first stages in the full sample of district-years. Instruments in these complementary first stages (available upon request) are strong and coefficients are qualitatively unchanged, even when both year and district fixed effects are included, for which the panelised DHS and NLFS samples lack power.

For the below reduced-form empirical evidence, the (likely biased, see above) OLS specification, run on data aggregated to the district-year level, would be:

$$\begin{aligned} Agriculture_{d,t} = \alpha + \beta_1 Roadind_{d,t} + \beta_2 Roadind_{d,t} * Elecind_{d,t} \\ + \delta_d + \lambda_t + \epsilon_{d,t} \end{aligned} \tag{4.1}$$

However, following the observations in Subsection 4.1 above regarding the endogeneity concerns associated with this OLS estimator, I instead run two-stage least squares (2SLS) on the following specifications, with year-fixed

[14]For multiple endogenous regressors, as in my case, Sanderson and Windmeijer (2016) provide the most relevant weak instrument F-test statistic. However, the authors explicitly recommend to report classical F-test and Cragg-Donald test statistics along their Sanderson-Windmeijer statistic for comparative purposes and robustness. I report all three statistics in 2SLS results throughout.

effects and district-level initial values as controls:[15]

$$
\begin{aligned}
Roadind_{d,t} &= \eta_0 + \eta_1 RoadIV_{d,t} + \eta_2 RoadIV_{d,t} * ElecIV_{d,t} \\
&\quad + X'_d \tau_1 + \rho_t + \nu_{d,t}
\end{aligned}
\tag{4.2}
$$

$$
\begin{aligned}
Roadind_{d,t} * Elecind_{d,t} &= \eta_3 + \eta_4 RoadIV_{d,t} + \eta_5 RoadIV_{d,t} * ElecIV_{d,t} \\
&\quad + X'_d \tau_2 + \rho_t + \nu_{d,t}
\end{aligned}
\tag{4.3}
$$

$$
\begin{aligned}
Agriculture_{d,t} &= \alpha + \beta_1^{2SLS} \widehat{Roadind}_{d,t} + \beta_2^{2SLS} Roadind\widehat{_{d,t} * E}lecind_{d,t} \\
&\quad + X'_d \gamma + \lambda_t + \epsilon_{d,t}
\end{aligned}
\tag{4.4}
$$

In the core estimation specification [eq. (4.4)], the possible outcome variables are $Agriculture_{d,t}$, $Services_{d,t}$, $Manufacturing_{d,t}$ or $Notwork_{d,t}$ and represent the share of people reporting an agricultural, services or manufacturing sector occupation, or no current employment in the last week, respectively, in district d, aggregated from all EA's (villages) i in that district, in the year of NLFS, DHS or HCES survey round t.[16]

$Roadind_{d,t}$ represents a dummy if district d contains an all-weather road in year t, while $Roadind_{d,t} * Elecind_{d,t}$ captures the interaction of dummies if district d was connected to both a road and substation in year t. X'_d denotes initial district-level controls that are either time-invariant (e.g. ruggedness, distance to market town, distance to administrative zone capital, soil quality) or would

[15]As mentioned above, the pseudo-panelisation of repeated cross-sectional surveys at the district level reduces the available sample to an extent that I lack power on the first stages to estimate 2SLS combined with full two-dimensional fixed effects.

[16]The sample restriction employed across NLFS, DHS and HCES datasets is that respondents must have worked at some point during the last twelve months, but are allowed to be not currently working to enter the sample. An additional age restriction ensures that only respondents older than eleven or younger than 70 enter the sample.

constitute bad controls if included as time-varying controls (e.g. nightlights, or even temperature anomalies).

The coefficients β_1^{2SLS} capture the effect of access to a road alone on the different sectoral employment shares, while the different β_2^{2SLS} coefficients capture the big push infrastructure interaction term at the heart of this paper.

Chapter 5

Reduced-form Results

I first estimate local average treatment effects of the two infrastructure investments outlined above: the effects of roads only and the big push interaction of roads with electricity access on sectoral employment at the district-year level. Table 9.5 provides the results from regression equation (4.4) run on sectoral employment shares from the NLFS repeated cross-sections. Throughout, standard errors are clustered at the district-level, which is the level of the treatment.[1]

I find a strong positive sectoral employment share response in services (+19.2%) and a strong negative employment share response in manufacturing (-11.5%) from road access alone. Agricultural employment is weakly negative, but insignificant due great heterogeneity in responses across districts, as explained in the results on spatial heterogeneity below.

Districts that obtain road access and electrification, in contrast, see a large reversal in the manufacturing employment share (+13.1%), such that the interaction effect at least fully overcomes the road-induced decrease.[2] Agricultural

[1]In order to also allow for spatial correlation structures beyond arbitrary district borders, I also test for robustness using Conley standard errors. Results (not reported) are unchanged.

[2]The p-value of both coeficients combined being indistinguishable from zero is 0.7581, as reported in Table 9.5.

employment in big push infrastructure locations decreases sharply (-20.2%), whereas services employment increases further, but insignificantly.[3]

The above results show sectoral employment splits according to 'major occupational divisions' of the International Standard Classification of Occupations (ISCO), which is my preferred definition. In theory, respondents' occupations may only imperfectly capture in which sector of the economy they actually work. Therefore, Table 9.6 confirms the above regression results based on equivalent sectoral classifications derived from first-digit International Standard Industrial Classification of All Economic Activities (ISIC) industry groups. Results are remarkedly similar, highlighting that respondents' occupations-derived sector of work and their industry-derived sector of work do not deviate systematically in the sample to explain the core results.

To understand better which occupational changes drive the above result of big push infrastructure investments to cause diverging sectoral employment patterns, I break down sectoral employment into employment by occupational subgroup. Accordingly, I re-run the two-stage least squares regressions using employment shares in these occupational subgroups as outcome variables: for the NLFS, the outcome variables are the first-digit ISCO (or ISIC) subgroups. For the DHS, the outcome variables are the 'major occupational groups'. Figure 8.8 presents this occupational breakdown for the NLFS (upper panel) and DHS (lower panel) samples, where the two regression coefficient bars for each

[3]Appendix Table A7 provides the same specification run on the NLFS sample excluding the Somali region in eastern Ethiopia, which was not sampled for the DHS survey. Sparsely populated and dominated by pastoral tribes, the Somali region is commonly understood as an outlier along cultural, economic and political lines. Results are statistically slightly stronger, but qualitatively unchanged.

occupational subgroup represent a separate regression. Each light orange bar relates to β_1^{2SLS}, the road infrastructure effect, whereas each dark bar relates to β_2^{2SLS}, the big push infrastructure interaction effect.[4]

I find that road access especially decreases relative employment for crafts, i.e. traditional manufacturing occupations, and increases relative employment for services/sales, i.e. mostly small-scale retail. Thus, the strong positive effect on services employment reported above is confined to employment in retail and sales, which is mostly informal in nature.[5] In contrast, road access alone causes manufacturing employment to drop mostly in skilled (i.e. artisanal, craft and/or handiwork) activities. Anecdotal evidence confirms such supposed adverse impacts on local manufacturing production in the face of sudden external competition, either from the economic centre or imports of Chinese manufactured goods.[6]

In contrast, when combined with electricity, big push infrastructure investments cause positive reversals in employment in both crafts and elementary occupations, i.e. construction and mining. While both coefficients are qualitatively similar across NLFS and DHS samples, they are insignificant in the NLFS but significant in the DHS. Albeit at much lower amplitude, big push infrastructure also causes employment increases in plant operations and assembly occupations (i.e. employment associated with modern factory

[4]In relation to Table 9.5, light orange bars represent results from row 1 and dark orange bars those from row 2, respectively, in regressions where the three sectoral employment shares are replaced by each of the nine occupational subgroups' employment shares.

[5]Given the aggregate share of small-scale retail and sales employment in overall services employment in Ethiopia, the infrastructure-induced direction of structural transformation into services is likewise reflected in the macroeconomic data. Figure 8.4 shows a marked increase in services employment in Ethiopia during my study period.

[6]Gunning, Krishnan and Mengistu (2018) provide evidence on the price, income and variety effects of market integration in the Ethiopian context.

work). In the services-sector occupational subgroups, sales employment is insignificantly reduced, whereas employment as clerical and professional services jobs (e.g. engineers, architects, etc.) increase.[7] Finally, (skilled) agricultural employment decreases sharply in electrified locations across samples, although only significantly so in the DHS.

To provide additional insights into which industry subgroups are mostly affected by infrastructure investments, the lower panel of Appendix Figure A14 provides equivalent results for the industry subgroups (ISIC first-digit): roads cause employment decreases in manufacturing, mining, construction and administrative industries, while wholesale-retail and education industries see employment increase. In contrast, big push infrastructure investments lead to significantly more employment in manufacturing, construction, accommodation and food, as well as mining, exactly replicating the occupational subgroup results from above.

One potential concern for the interpretation of results would be if the construction industry represented exclusively non-tradeable activities, although almost one third of workers in the construction industry are classified as manufacturing workers (according to their occupations). However, in the theoretical framework, all manufacturing employment is considered to produce tradeable varieties. Therefore, Appendix Table A8 reports results under the most extreme possible re-classification of construction industry workers: results show that even if all construction industry workers were hypothetically considered to be part of the non-tradeable services sector, results remain qualitatively unchanged.

[7]Due to the extremely low total employment numbers in these occupational subgroups, results are noisy and only marginally statistically significant.

Overall, these findings point towards an overall compelling argument that the market integration from road access leads to reductions in occupational groups and industries one would expect to face greater import competition from trade, whereas occupational groups and industries expected to benefit from cheap imported manufactured goods, i.e. retail sales, increase their relative employment shares. With electrification, this pattern is reversed, where additional increases in services occupations mostly arise from jobs usually associated with white-collar activities and office work. The manufacturing reversal is likewise driven by increased construction and manufacturing, where the latter, albeit at low levels, includes newly arising modern plant employment.

Against the background of the model predictions from Section 2 above, I can thus confirm the diverging patterns of structural transformation from isolated against big push infrastructure investments. In addition, the model does also explicitly predict spatial heterogeneity, such that districts less shielded by transport costs should experience larger adverse effects from import competition (and, thus, greater reductions in manufacturing employment). Likewise, if the interaction of infrastructure matters, those districts with the larger market access should see greater benefits from electrification than those with less.

Table 9.8 confirms, firstly, the substantial spatial heterogeneity in structural transformation outcomes across space. Secondly, the pattern of heterogeneity matches model predictions: districts closer than the median distance to the nearest administrative zone capital (which is usually the nearest larger town or city), suffer heavier employment losses in manufacturing from road access, but also reverse this larger effect equally once electrification arrives. The latter reversal is heavily driven by agricultural employment decreases (which may have even increased from roads before), which closely confirms model predictions of locations' changes in Ricardian comparative advantage following a road-induced trade cost reduction (for tradeable sectors) and an additional, electrification-induced productivity shock (for non-agricultural sectors).[8]

Quite in contrast, for remote districts (with above median distance to the nearest city), the employment effects on manufacturing in column (3) of Table 9.8 are more muted and statistically insignificant. The overall services employment increase, however, appears predominantly driven by far-from zone capital districts (+25.8%), which makes sense given the extremely low services employment share in such locations without roads. In sum, these results appear to point to structural transformation patterns along districts' likely comparative advantage, and to shifting agglomerations of employment across space from big push infrastructure investments.

In line with the above findings, I can also confirm that the demographic make-up of employment changes significantly in big push infrastructure locations. Figure 8.11 highlights that these locations experience a concentration of the labour force around prime-working age, which may point towards a formalisation of employment. On average, workers are 2.2 years older

[8]Classic economic geography models (Krugman, 1991; Krugman & Venables, 1995) predict qualitatively similar spatial heterogeneity to arise from trade cost reductions in a multi-sector geography, although driven by an alternative mechanism of increasing returns to scale. Allen and Arkolakis (2014) provide conditions under which different classes of spatial economic models such as Krugman (1991) or the foundations of my framework, Helpman (1998) and Eaton and Kortum (2002), become isomorphic.

than in road-only connected districts (see Appendix Table A12).[9] As Figure 8.11 highlights, this effect is mostly driven by a considerable narrowing of the age distribution in road- and electricity-connected districts around the prime-working ages from 20-40 years, at the expense of especially teenagers participating in the labour force in districts with or without roads. This narrowing of the age pyramid can be confirmed in quantile regression estimates, as shown in Appendix Figure A15, where electrified districts add especially workers between the second and sixth decile of the age distribution, i.e. between ages 18 and 33.[10]

In order to interpret the shift in big push infrastructure locations' employment towards prime working age labour as potential formalisation of employment, further evidence on the nature of employment relationships is required: Table 9.9 highlights how big push infrastructure locations experience statistically significant increases in the share of employed workers, at the expense of self-employment. Appendix Table A11 breaks down employee and self-employed categories further and confirms that self-employment and especially unpaid household work decrease, whereas private sector employment increases markedly. Government employment reacts positively, although insignificantly, to both kinds of infrastructure investments.[11]

[9]Column 4 of Table A12 also shows how the share of divorced workers increases in interacted infrastructure districts, which one may interpret as a proxy for the arrival of greater economic opportunities that decrease the economic value of marriage.

[10]While the narrowing of the age pyramid appears symmetric across genders, the overall structural transformation results from infrastructure display a marked heterogeneity across gender: as I highlight in Appendix Figure A16 and Appendix Table A10, relative decreases in manufacturing employment at the expense of services due to road access are mostly driven by females, while the big push infrastructure effect out of agriculture into manufacturing is mostly driven by males.

[11]Interestingly, significantly more workers are employed in government parastatal entities (which dominated the Ethiopian economy before 2000) in road-connected locations, at the expense of private employment. Parastatal employees mostly work in (in decreasing

One of the core adjustment mechanisms in the spatial general equilibrium model is the movement of labour across places to equalise real wages. Without full migration flows or at least a Census sample with origins, I can instead only present in-migration results derived from the selective NLFS and DHS samples. Out-migration to unsampled districts cannot be detected by construction. Interestingly, no differential migration responses can be detected for road-connected places, whereas additional electricity connections do lead to economically meaningful, but only marginally significant positive immigration responses (see Table 9.10).

To check the robustness of the core reduced-form results discussed above, I briefly highlight further results regarding education, the selection of migrants and the labour force composition. With respect to education, the overall education results (cf. Appendix Table A14 are ambiguous since only road-connected districts show increases in literacy (column 1), whereas years of educational attainment only increase insignificantly. A placebo test if these education effects are indeed driven by infrastructure is provided in Appendix Table A15: as expected, the educational attainment of only those groups (teenagers [column 1] and young adults [column 2]) increases, who were young enough at the time of infrastructure arrival to still increase their education, either by staying in school or by opting for higher education. Interestingly, educational attainment by migrants is higher than that of non-migrants (columns 5 vs 6), pointing towards positive selection of migrants in road-connected districts.[12] Finally, overall labour force participation shows

order): farming, manufacturing, construction, wholesale/retail and food and accommodation industries, which together represent almost 60% of parastatal employment.

[12]Taken at face value, the negative coefficient in Table A15, column 5, row 2 would thus indicate negative selection of migrants into road- and electricity-connected districts. This result

an insignificant positive effect from roads, and a strong negative effect from electrification, in line with the age pyramid narrowing highlighted above: many teenagers opt out of the labour force, either due to lower fertility ('missing youths') or young people staying (longer) in education.

Without accurate measures of sectoral productivity differences, the above reduced-form results of substantial infrastructure-induced structural transformation may be irrelevant for economic development and growth. Therefore, and to motivate the structural estimation in Section 6 below, I briefly highlight reduced-form results on three distinct welfare proxies: household expenditure, household real consumption of durables (both from micro raw data) and remotely-sensed proxies for economic development such as nighlights.

Regarding the former, Table 9.11 reports 2SLS results of infrastructure investments on household consumption. Deaton (2003) recommends the use of household expenditure as proxy for income in settings with low-quality data on household incomes or wages, as in the case of Ethiopia.[13] Column 1 indicates that only the interaction of infrastructure has significantly positive effects on household expenditure per capita, and cannot simply be explained by smaller household sizes. If one takes the above sectoral employment results in combination with the model seriously, then the insignificant effect of roads access alone may confirm that while all consumers benefit from lower prices of

may or may not be counterintuitive, depending on the skill requirements of newly arising plant operations and construction jobs.

[13]The income schedule was recently removed from the HCES (formerly HICES, including 'Income') questionnaires due to the admittedly low quality of income responses. Similarly, despite its name, the National Labour Force Survey (NLFS) did not collect wage or income information from respondents.

imported manufactured goods, some previously local manufacturers may be harmed substantially.[14]

Following Young's (2012) approach, I present results on households' real expenditure as proxy for economic growth in Tables 9.12 and 9.13. Results confirm my interpretation of the structural transformation results: I do not find statistically significant improvements in real expenditure on either durables (Table 9.12) or housing (Table 9.13) from road access alone, whereas real consumption of eight out of twelve categories does markedly increase in big push infrastructure locations. If higher relative sectoral employment in manufacturing is in fact indicative of higher aggregate productivity, the above pattern of real consumption would be expected.

Finally, I attempt to shed light on the likely growth and welfare implications of the reduced-form results when measured more objectively from remotely-sensed sources, such as nightlights (Henderson, Storeygard & Weil, 2012). Table 9.14 reports two-stage least-squares estimates of satellite-derived outcomes on treatments: I find roads to lead to greater overall population density in districts, whereas electricity reduces this effect again, most likely due to fertility responses as districts develop economically. These results are also in line with the narrowing of the age distribution discussed above, which indicates that the prime-working age population in treated districts increases, whereas the overall population (mostly driven by infants and youth) would fall. Results of satellite-derived nightlights and built-up areas confirm that

[14]This trade-off resembles the opposing results on consumer and producer surplus in Atkin, Faber and Gonzalez-Navarro's (2018) context of foreign retail entry in Mexico, that led small-scale domestic retailers to exit.

big push infrastructure investments appear to result in noticeable increases in economic development.[15]

[15] One caveat with the remotely-sensed results, however, is that satellite-derived data products such as the DMSP-OLS nighlights or GHSL built-up area use nighlights either as direct signal or as an input to image processing algorithms, such that the resulting outcome rasters suffer from detection bias: economic growth in unelectrified areas may go entirely unnoticed.

Chapter 6

Structural Estimation

Equipped with the reduced-form, causal local average treatment effects of infrastructure investments on structural transformation in Ethiopia over the last two decades, as well as a theoretical model to characterise a spatial general equilibrium, I turn to structurally estimating the aggregate general equilibrium effects of infrastructure investments on structural transformation and welfare.

I proceed in five steps: first, I measure effective, terrain-adjusted distances between locations in Ethiopia and, using origin-destination price comparisons, estimate trade costs according to spatial arbitrage. Second, I calibrate the model on baseline observables to obtain baseline sectoral productivities. Third, I leverage the reduced-form results to set up a moment condition of changes in structural transformation the model should replicate following changes in infrastructure. This step necessarily involves taking a stance on the functional form of the relationship between productivities and electrification. Fourth, I numerically solve the baseline-calibrated model until the moment condition holds in terms of the model's endogenous variables. Finally, I structurally estimate aggregate structural transformation and welfare effects of big push infrastructure investments and compare them to counterfactual, isolated infrastructure investments.

6.1 Step 1: Roads and trade costs

In order to relate changes in roads infrastructure to changes in trade costs, a model object of central interest (cf. Subsection 2.4), I employ the following methodology. To start, I assume trade costs between locations to be a (log-linear) function of distances between locations: $d_{ni} = (distance_{ni})^{\tau}$. I develop an algorithm to measure the full matrix of effective distances from each district capital to each other. Then, I estimate trade costs directly from origin-destination price gaps for a subset of goods (for which origin locations are unique and cleanly identifiable), under an assumption of spatial arbitrage. The latter allows me to back out the elasticity of trade costs with respect to distance, τ. Finally, I use this elasticity to translate changes in the distance matrix due to new road construction into changes in trade costs for all locations and goods.[1]

Regarding the matrix of effective distances between locations, I compute an exhaustive matrix of district capital to distance capital least-cost distances by means of an Dijkstra algorithm employed on a tailored cost surface.[2] The algorithm then determines the least-costly route to connect each district capital to each other, separately. I generate the underlying cost surface from a terrain raster image overlaid with the year-specific rasterised all-weather road vector

[1]This procedure builds on the underlying assumption that the relationship between distance and trade costs is not systematically different across goods. In other words, the subset of goods with cleanly identifiable origin locations are assumed to be representative of a universal mapping of distance to trade costs for all traded goods.

[2]Wherever information on the district capital is not available, I instead use the district centroid, under the additional geometric constraint that this centroid has to lie inside the district.

layers. I then run the algorithm separately for the four years for which distinct road vector layers are available (i.e. layers for 1999/2000, 2005/2006, 2011/2012/2013 and 2016).

Terrain, or the difficulty in crossing a given pixel (representing a given stretch of land), is expressed as a terrain ruggedness index value with scores ranging from zero to 45. The full geography of Ethiopia is represented by a graph of approximately 12,000 quadratic pixels, each representing an area of approximately 185 times 185 metres (when measured at the equator). For pixels without an all-weather road in it, I measure the cost to cross the pixel as the distance (in kilometres) to traverse the pixel in North-South or East-West direction times one plus the terrain ruggedness index. For district centroid or capital pixels and all-weather road pixels, I set the cost to traverse the pixel as simply the distance covered (i.e. a terrain ruggedness index of zero plus the normalisation of one). Intuitively, my approach is equivalent to understanding a given all-weather road in a pixel to virtually level the terrain in trade cost terms.[3]

As a sanity check on the modelling choice that road infrastructure investments decrease trade costs, I test whether the reduced-form measure of access to an all-weather road indeed feeds through to lower effective distances. I can

[3]This procedure yields effective district capital to district capital distances that appear sensible: when benchmarked against state-of-the-art map routing engines such as OpenRouteService, random distance pairs from my distance matrix are very close to the OpenRouteService predicted distances travelled between the same locations. This manual robustness check also works in cases where no all-weather road connects the district capital to the remaining Ethiopian road network, mostly due to my terrain-avoidance algorithm yielding similar results to software engines with information on non-gravel, earth roads. Unfortunately, the underlying map source for OpenRouteService, OpenStreetMap, does not have information on Ethiopian roads and settlements beyond 2014. In addition to the lack of panel information in publicly available map engines, my algorithm is also more robust in the sense of employing a true Dijkstra-frontier recognition procedure.

confirm this link in the data: Table 9.15 highlights that there is a highly positive and significant association between the district-level all-weather road indicator (from the reduced-form above) and the transport cost matrix between district centroids or capitals from the Dijkstra least-cost algorithm.[4] In other words, the connections between district capitals deemed least-cost by the algorithm turn out to rely heavily on all-weather roads. An expansion in the road network is thus directly associated with statistically significant reductions in newly-connected districts' transport costs (and thus, up to a transformation described below, trade costs).[5]

Regarding the transport cost shock amplitudes, Appendix Figure A17 shows that the long difference in relative changes in Dijkstra algorithm least-cost distances from the earliest (1999) to the latest point in my sample (2016) conveys substantial heterogeneity in terms of shock amplitudes across space. The relative changes in the per district sums of least-cost distances to all other locations range from -35.16% to -7.18%. Hence, in the structural estimation, I feed empirically relevant variation in transport cost shocks across districts over time to the spatial general equilibrium model, where especially remote, but moderate to heavily populated locations are affected most.[6]

[4]Since I am not constrained by gaps in the coverage of outcome variables for either districts or years in this descriptive exercise, I run both OLS and fixed-effects specifications on the full panel of all 689 Ethiopian districts at four different points in time, i.e. for each of the four years for which I have distinct information on the extent of the all-weather road network as described in Section 3 above. This implies, necessarily, that I also run the Dijkstra least-cost algorithm four times on distinct cost surfaces.

[5]As columns (3) and (4) in Table 9.15 show, a district getting connected to the all-weather road network is associated with a 4% reduction in the sum of that district's least-cost distances to all other districts.

[6]At least seven distinct zones affected by large transport cost shocks emerge from Appendix Figure A17, in particular in central Amhara (South Wollo, circa 200 km north of Addis), northern Amhara (Wag Himru, circa 400 km north of Addis), northwestern Oromia (Horo Guduru, circa 200 km north-west of Addis), western Oromia (Ilubabor, circa 350 km west of Addis), practically the whole south and south-west of SNNPR (e.g. Kaffa and South Omo,

To make use of this empirically relevant, full matrix of district capital distances, I also require information on the elasticity of trade costs with respect to distance. To this end, I estimate trade costs from price gaps across locations between selected products from the Ethiopian Central Statistical Authority's Retail Price Index (RPI) raw data. This raw price data is available at a monthly level for 119 markets across Ethiopia from 1998 until at least 2017. Under the assumption of spatial arbitrage, price gaps across origin and destination locations reveal trade costs if the origin location of a traded good is unique and as such identifiable. Similar to Atkin and Donaldson's (2015) methodology, unique origin locations for up to 23 distinct goods can be identified at the barcode-level in the Ethiopian context using RPI raw data.[7]

Since I know the effective distances between origin-destination market pairs, the estimated trade costs from price gaps for the above 23 distinct goods allows to back out the implied elasticity of trade costs with respect to distance, τ, for this subsample. Finally, I assume this elasticity to be representative for the relationship between effective distance and trade costs of all traded goods and locations in Ethiopia. Therefore, I can translate changes in effective distances due to road construction over time into changes in trade costs, which are the relevant shock to the general equilibrium system from road construction.

circa 350-500 km south-southwest of Addis), as well as central Oromia (Arsi, circa 150-250 km south of Addis) and eastern Oromia (Hararge, circa 300 km west of Addis). For comparison, Appendix Figure A3 provides the distribution of population in 2015.

[7]The RPI raw price data employed here have greater temporal coverage and exploit several more confirmed product origin locations to construct destination-origin price gaps compared to Atkin and Donaldson's (2015) data. For simplicity, and in contrast to Atkin and Donaldson (2015), I do not allow mark-ups to vary across space. Allowing for potentially oligopolistic intermediaries would further amplify my core results of road investments.

6.2 Step 2: Calibration of baseline sectoral productivities

For the calibration of the model, I follow a similar procedure as Michaels et al. (2011) of using observable data, a matrix of trade costs and central parameter values either from the literature or from Ethiopian micro data to calibrate baseline sectoral productivities. In line with the reduced-form analysis above, the baseline year refers to the earliest point in time for which survey data is available, 1999/2000.

In particular, I use observable data on districts' initial sectoral employment (derived from shares) $\{L_n^M, L_n^T, L_n^S\}$, district land area $\{H_n\}$, as well as the trade cost matrix $\{d_{ni}\}$ from Step 1 in Subsection 6.1 above.[8] In addition, Table 9.16 provides parameter values for $\{\alpha, \kappa, \sigma, \theta, \mu^M, \mu^T, \tau\}$ obtained either from micro data or from the relevant literature. Together, these model inputs allow me to pin down baseline sectoral productivities $\{A_n^M, A_n^T, A_n^S\}$ and (normalised) wages $\{w_n\}$.

The calibration procedure solves a system of four equations for each location for four unknowns for each location. In particular, the three sectoral labour market clearing conditions [eq. (2.20-2.22)] and one labour mobility condition [eq. (2.19)] provide the four equations for the three unknown sectoral productivities and normalised wages.[9]

[8]Data sources for the structural estimation are provided in the data appendix, Appendix Section A.

[9]As Michaels et al. (2011) show, given the symmetry in sectoral labour market and land market clearing conditions, rental rates in equilibrium can be expressed as a function of only observables and wages, such that rental rates are pinned down via wages. In my case, the rental rate can be expressed as:

$$r_n = w_n \frac{1}{\alpha} \frac{L_n}{H_n} \left[(1-\alpha) + \left(\frac{1-\mu^T}{\mu^T} \right) \frac{L_n^T}{L_n} + \left(\frac{1-\mu^M}{\mu^M} \right) \frac{L_n^M}{L_n} + \left(\frac{1-\mu^S}{\mu^S} \right) \frac{L_n^S}{L_n} \right]$$

I solve this calibration system by solving several enveloped problems: first, for given labour-market clearing initial productivity guesses and wage guesses, I adjust relative productivities in the services sector to ensure sectoral labour market clearing.[10] Given the non-tradeable nature of this sector, relative services sector productivity has to be increased whenever labour expenditure on services exceeds labour income from services. Second, for both tradeable sectors, relative productivities have to be likewise adjusted to ensure sectoral labour market clearing. However, given the tradeable nature of both sectors, relative productivity adjustments have to occur in the opposite direction since the manufacturing productivity in location n will also affect consumption and production in location i. Finally, I employ the labour mobility condition to check for real wage equalisation. Adjustments to normalised wages straightforwardly ensure the allocation of population observed in the data at baseline, given calibrated, sectoral labour market-clearing productivities, constitutes a spatial equilibrium.

Uniqueness of the calibrated productivities and normalised wages follows from the proof for the two-sector model in Michaels et al. (2011). Given the non-tradeable nature of services and the irrelevance of interacted infrastructure shocks for the baseline calibration, this proof follows through to the three-sector version employed in this paper. Reassuringly, the calibration consistently arrives at a unique combination of productivities (under the same wage normalisation), independent of widely varying initial guesses.

.

[10]Before calibration of the different envelopes is initiated, I ensure that each location's overall labour market clears under the incoming productivity guesses. This bounding exercise is straightforwardly achieved by adjusting two out of three sectoral productivities and guarantees stability in the convergence towards the unique calibration solution.

As a basic robustness check, I feed the four calibrated vectors (i.e. sectoral productivities and wages) into the numerical solution algorithm described in Subsection 2.3 under random initial guesses for sectoral employment, wages and population, to confirm that the numerical solution algorithm will predict unique values for the endogenous variables that exactly match the observed data.

6.3 Step 3: Moment conditions from reduced-form

The reduced-form results presented in Section 5 provide estimates of the differences in structural transformation and growth proxies caused by differences in locations' infrastructure access. However, these results do not provide insights regarding the implications of big push infrastructure on either aggregate structural transformation or welfare. In contrast, the spatial general equilibrium model developed in Section 2 provides a structure that can be leveraged to learn about these aggregate effects. In particular, given the knowledge of causally identified differences in outcomes, I can inform the model to replicate these differences in places receiving roads or big push infrastructure investments, and then derive welfare and counterfactuals from structural estimation.[11]

In my context, the local average treatment effect of infrastructure investments on structural transformation provides a 'net' effect after reallocation

[11] Faber and Gaubert (2016) follow a related approach which they call 'model-based indirect inference', whereby they leverage the exclusion restriction on their instrumental variable to identify agglomeration externalities. Ahlfeldt, Redding, Sturm and Wolf (2015) likewise identify agglomeration effects from reduced-form evidence, although their moment conditions exploit exogeneous variation from a natural experiment. Finally, Adão, Arkolakis and Esposito (2019) showcase how quasi-experimental variation can be exploited to derive model-implied optimal instrumental variable estimators to identify spatial linkages of geography-wide shocks.

of factors of production and trade has ensued. Therefore, I can constrain the spatial general equilibrium model to exactly replicate these changes in infrastructure-receiving locations. Based on this replication of empirical reality, I can then leverage the model to learn more about effects not observable in my reduced-form research design: aggregate sectoral employment responses, implied welfare and hypothetical welfare results from simulated counterfacutal infrastructure policies (e.g. isolated roads investments or isolated electrification). Regarding the latter, however, one crucial link between the reduced-form empirics and the model is still missing, namely how electrification access in a location translates into manufacturing and services sector productivity changes. Therefore, I also take a stance on the functional form of this relationship and estimate it structurally below.

The moment conditions below impose that the model-predicted sectoral employment shares in the manufacturing sector [eq. (6.1)] and in the services sector [eq. (6.2)] for each location match, in expectation, the predicted values for locations' sectoral employment shares from the causal reduced-form results in Section 5. I therefore effectively constrain the model to replicate the estimated differences. Minimising this moment condition employs evidence on the causal effect of road and electricity investments on sectoral employment patterns across locations, after controlling for time-invariant district and location-invariant time effects.

$$E\left[\left(\lambda_n^{M,model} - \widehat{\lambda_n^{M,2SLS}}\right)\right] = 0 \tag{6.1}$$

$$E\left[\left(\lambda_n^{S,model} - \widehat{\lambda_n^{S,2SLS}}\right)\right] = 0 \tag{6.2}$$

I use only sectoral employment shares predicted by the two-stage least squares procedure explained in Subsection 4.3, i.e. the three sectoral variants of equation (4.4), to isolate the structural estimation from the endogenous components in the relationship between infrastructure investments and structural transformation.[12]

Before I can structurally estimate the effects of big push infrastructure on structural estimation, I have to take a stance on the functional form of how electrification is assumed to affect productivity.[13] I assume that changes in sectoral productivities in manufacturing and services are log-linear in changes in electrification:

$$\Delta lnA_n^M = ln\frac{A_n^{M,2016}}{\overline{\overline{A^{M,2016}}}} - ln\frac{A_n^{M,1999}}{\overline{\overline{A^{M,1999}}}} = \phi log(1 + \widehat{\Delta elec_n}) + \nu_n \tag{6.3}$$

$$\Delta lnA_n^S = ln\frac{A_n^{S,2016}}{\overline{\overline{A^{S,2016}}}} - ln\frac{A_n^{S,1999}}{\overline{\overline{A^{S,1999}}}} = \phi log(1 + \widehat{\Delta elec_n}) + \nu_n \tag{6.4}$$

This functional form separates three elements that may drive sectoral productivity: country-wide productivity changes over time (corrected for via scaling by the geometric mean), district-specific innate productivity (taken out by first-differencing) and (causal) electrification (i.e. predicted electrification status at the district-level according to the electrification IV described in Sub-

[12]Since instrumental variable and two-stages least squares estimators are conceptually equivalent, the above conditions can be regarded as representing the two-stage least squares estimator interpretation of these moment conditions. In contrast, Faber and Gaubert (2016) employ the equivalent instrumental variable estimator interpretation of their moment conditions, namely that their instrument, say z, and some innate sectoral productivity are orthogonal. Both moment condition interpretations have the same objective of ensuring that the exclusion restriction holds in expectation.

[13]The equivalent relationship between roads and trade costs is described in Subsection 6.1 above.

section 4.2).[14] Finally, ϕ captures the elasticity of modern sector productivities with respect to electrification, which represents the core parameter to be estimated to minimise the (sample equivalent) of the moment conditins.

The above functional form is directly informed by the empirical context: in the Ethiopian setting since the late 1990s, the arrival of the electric grid provides the first constant source of energy supply for the overwhelming majority of producers. Hence, it is reasonable to understand electrification in Ethiopia as an extensive margin change with respect to the application of modern, power-driven means of production.[15] It is against this empirical background that I model electrification as a direct productivity shock for production in sectors most likely to benefit from such a sharp increase in the availability of modern production technology.[16]

Current work in progress expands the above functional form to allow for electrification-specific agglomeration effects, which makes any potential complementarity between roads and electrification for welfare more explicit and estimable. In particular, this extension aims at allowing deviations from the log-linear relationship between electrification and productivity, such that road-induced heterogeneity in agglomeration differentially scales an electrification elasticity otherwise constant across space.

[14]This formulation is similar in spirit to Ahlfeldt, Redding, Sturm and Wolf's (2015) functional form for their 'adjusted production fundamentals'.

[15]cf. Fried and Lagakos (2017) in the same empirical context, where the decrease in energy prices from electrification leads to a discrete jump towards adoption of modern production technology.

[16]Since firm competition and trade in the theoretical framework are determined by Ricardian comparative advantage, it would be qualitatively equivalent to assume that productivities in agriculture also increase from electrification, but less so than productivities in the modern sectors.

6.4 Step 4: Numerically solving for moment conditions

To ensure that the moment conditions hold given the model-predicted endogenous variables (such as sectoral employment shares), I iteratively solve the numerical solution forward to determine sectoral productivities under big push infrastructure investments. In particular, for given sectoral productivity guesses (e.g. baseline calibrated levels $\{A_n^{M,1999}, A_n^{S,1999}\}$ from Subsection 6.2 above), I solve the model numerically to obtain $\lambda_n^{K,model}\left(d'_{ni}, A_n^{M,initial}, A_n^{S,initial}\right)$, i.e. model-predicted employment shares for sectors $K \in \{M, S\}$ as a function of trade costs (adjusted for new roads construction) and initial productivity guesses.

I then check how closely the moment conditions hold under the model-predicted employment shares and employ an algorithm to adjust productivity guesses until the solution converges in satisfying the moment conditions at an accuracy of 10^{-6}.

From this iteration through the numerical solution, I obtain model-consistent sectoral productivities $\{A_n^{M,2016}, A_n^{S,2016}\}$ that minimise the distance to the reduced-form moments. Once the moment conditions are minimised, all objects that determine electricity-affected sectoral productivities [eq. (6.3) and eq. (6.4)] are observed with the exception of ϕ. I can therefore estimate this elasticity of productivity with respect to electrification via standard 2SLS and find an elasticity of approximately $\hat{\phi} \approx 0.284$.

Current work in progress benchmarks the above procedure against a standard generalised method of moments (GMM) procedure to estimate the moment condition-minimising elasticity of productivity with respect to electrification.

6.5 Step 5: Welfare and counterfactuals

Do big push infrastructure investments matter for welfare? Does the specific combination of infrastructure investments generate different results than hypothetical, isolated investment counterfactuals? Based on the endogenous variables that satisfy the moment condition estimated in Subsection 6.4 above, Figure 8.12 plots baseline welfare in terms of (equalised) real wages against different infrastructure scenarios. Expressed in terms of multiples of this baseline real wage level, the big push infrastructure investments generate more than 11% higher welfare than the baseline without investments. Interestingly, a roads only counterfactual achieves only 2% higher welfare, whereas an electrification only counterfactual achieves a meagre 0.7% welfare increase from baseline.

Closer inspection of the identities of the counterfactual investment locations provides further intuition on the result why big push infrastructure investments matter for welfare, too: in counterfactuals without electrification, road-receiving locations almost exclusively belong to the pool of previously peripheral locations with low manufacturing and services productivity vis-à-vis the core, such that welfare gains from integration are modest. These places lose manufacturing, with negative agglomeration implications, without gaining additional agglomeration from additional services or agriculture. Similarly, electrification alone, under a baseline road network of late 1990s extent, increases productivity in mostly remote locations with extremely high transport costs. Hence, although some positive welfare effects driven by local demand for now more productive manufacturing varieties are predicted, electrified locations miss out on the major lever of increased import demand for these new manufacturing varieties from other regions. Therefore, only the interaction of infrastructure investments reaps both sources of welfare gains and amplifies the agglomeration effects of electrification considerably.

A logical next step is to ask if the Ethiopian infrastructure investment strategy was optimal in the sense of the spatial general equilibrium model. In light of recent contributions to the literature (Fajgelbaum & Schaal, 2017; Balboni, 2019) that discuss this notion of optimal infrastructure allocation, the Ethiopian context gives rise to an additional notion of optimality, namely optimal sequencing. Given a constrained budget, which combination of isolated road, isolated electricity and interacted road and electricity investments would have maximised aggregate welfare? How about the distributional implications of such a policy experiment? For example, how many districts that received at least a road would have not received any infrastructure investments at all under welfare-optimal sequencing? Given the theoretical framework and structural estimation procedure developed above at hand, current work in progress answers these interesting questions to further inform infrastructure policy in low income countries.

Chapter 7

Conclusion

This paper presents causal evidence of big push infrastructure investments and their effects on structural transformation in a low income country, especially regarding the effect of combining road access and electrification on manufacturing and services employment.

In line with the predictions from a simple spatial general equilibrium model, I find that road access alone causes retail services employment to emerge, at the expense of traditional manufacturing. This adverse effect on manufacturing employment reverses, however, once locations gain additional electricity access. I argue that this reversal is driven by improved productivity via electricity-powered production processes.

As highlighted in the model, this latter finding confirms that big push infrastructure investments cause qualitatively different patterns of structural transformation than isolated infrastructure investments. Combining the reduced-form causal evidence with the structure of the model, results from a structural estimation procedure confirm that the welfare effects of big push infrastructure investments are considerably larger than the sum of its isolated infrastructure parts. I conclude that big push infrastructure investments appear to be in fact material to growth and welfare in low income country settings. Therefore, potential interaction effects of empirically common bundling or sequencing of infrastructure investments should be taken seriously, and interaction effects taken into consideration in the planning of infrastructure investments to maximise their impact.

Bibliography

Adamopoulos, T. (2018). Spatial integration, agricultural productivity, and development: A quantitative analysis of Ethiopia's road expansion program. *York University mimeo*.

Adão, R., Arkolakis, C. & Esposito, F. (2019). Spatial linkages, global shocks, and local labor markets: Theory and evidence. *Chicago Booth mimeo*.

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M. & Wolf, N. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, *83*(6), 2127–2189.

Allen, T. & Arkolakis, C. (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, *129*(3), 1085–1139.

Asher, S. & Novosad, P. (forthcoming). Rural roads and local economic development. *American Economic Review*.

Asher, S. & Novosad, P. (2016). Market access and structural transformation: Evidence from rural roads in India. *Dartmouth College mimeo*.

Atkin, D. & Donaldson, D. (2015). Who's getting globalized? The size and implications of intra-national trade costs. *NBER Working Paper Series*, (21439).

Atkin, D., Faber, B. & Gonzalez-Navarro, M. (2018). Retail globalization and household welfare: Evidence from Mexico. *Journal of Political Economy*, *126*(1), 1–73.

Balboni, C. (2019). In harm's way? infrastructure investments and the persistence of coastal cities. *London School of Economics mimeo*.

Banerjee, A., Duflo, E. & Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in China. *NBER Working Paper Series*, (17897).

Baumol, W. J. (1967). Macroeconomics of unbalanced growth: The anatomy of urban crisis. *American Economic Review*, *57*(3), 415–426.

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q. & Brandt, L. (2018). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*.

Behrens, K., Gaigné, C., Ottaviano, G. I. & Thisse, J.-F. (2006). Is remoteness a locational disadvantage? *Journal of Economic Geography*, *6*(3), 347–368.

Burlig, F. & Preonas, L. (2016). Out of the darkness and into the light? Development effects of rural electrification. *Energy Institute Working Paper No. 268*.

Bustos, P., Caprettini, B. & Ponticelli, J. (2016). Agricultural productivity and structural transformation: Evidence from Brazil. *American Economic Review*, *106*(6), 1320–1365.

Coşar, A. & Fajgelbaum, P. (2016). Internal geography, international trade, and regional specialization. *American Economic Journal. Microeconomics*, *8*(1), 24–56.

Deaton, A. (2003). Household surveys, consumption, and the measurement of poverty. *Economic Systems Research*, *15*(2), 135–159.

Desmet, K. & Rossi-Hansberg, E. (2014). Spatial development. *American Economic Review*, *104*(4), 1211–1243.

Dinkelman, T. (2011). The effects of rural electrification on employment: New evidence from South Africa. *American Economic Review*, *101*(7), 3078–3108.

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4-5), 899–934.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, *91*(4), 795–813.

Duflo, E. & Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, *122*(2), 601–646.

Eaton, J. & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779.

Eckert, F. & Peters, M. (2018). Spatial structural change. *Yale University mimeo*.

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway system. *The Review of Economic Studies*, *81*(3).

Faber, B. & Gaubert, C. (2016). Tourism and economic development: Evidence from Mexico's coastline. *NBER Working Paper Series*, (22300).

Fajgelbaum, P. & Schaal, E. (2017). Optimal transport networks in spatial equilibrium. *NBER Working Paper Series*, (23200).

Fried, S. & Lagakos, D. (2017). Rural electrification, migration and structural transformation: Evidence from Ethiopia. *Carleton College mimeo*.

Gebresilasse, M. (2019). Rural roads, agricultural extension, and productivity. *Amherst College mimeo*.

GTZ. (2009). Ethiopia's solar energy market: Target market analysis. *Gesellschaft für Technische Zusammenarbeit: Technical Report*.

Gunning, J. W., Krishnan, P. & Mengistu, A. T. (2018). Fading choice: Transport costs and variety in consumer goods. *CSAE Working Papers*, (WPS/2018-05-2).

Helpman, E. (1998). The size of regions. In D. Pines, E. Sadka & I. Zilcha (Eds.), *Topics in public economics: Theoretical and applied analysis* (pp. 33–54). Cambridge: Cambridge University Press.

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

Herrendorf, B., Rogerson, R. & Valentinyi, Á. (2013). Two perspectives on preferences and structural transformation. *American Economic Review*, *7*(103), 2752–2789.

Herrendorf, B., Rogerson, R. & Valentinyi, Á. (2014). Growth and structural transformation. chapter 6. *Handbook of Economic Growth*, *2*, 855–941.

Kassem, D. (2018). Does electrification cause industrial development? Grid expansion and firm turnover in Indonesia. *London School of Economics mimeo*.

Kebede, H. (2019). The gains from market integration: The welfare effects of rural roads in Ethiopia. *University of Virginia mimeo*.

Kline, P. & Moretti, E. (2014). Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority. *The Quarterly Journal of Economics*, *129*(1), 275–331.

Kongsamut, P., Rebelo, S. & Xie, D. (2001). Beyond balanced growth. *The Review of Economic Studies*, *68*(237), 869–882.

Krugman, P. (1991). Increasing returns and economic geography. *The Journal of Political Economy*, *99*(3).

Krugman, P. & Venables, A. J. (1995). Globalization and the inequality of nations. *The Quarterly Journal of Economics*, *110*(4), 857–880.

Kuznets, S. (1973). Modern economic growth: Findings and reflections. *American Economic Review*, *63*(3), 247–258.

Lee, K., Brewer, E., Meyo, F., Miguel, E., Podolsky, M., Rosa, J. & Wolfram, C. (2014). Barriers to electrification for "under grid" households in rural kenya. *NBER Working Paper Series*.

Lee, K., Miguel, E. & Wolfram, C. (2016). Experimental evidence on the demand for and costs of rural electrification. *NBER Working Paper Series*.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School*, *22*(2), 139–191.

Lipscomb, M., Mobarak, A. M. & Barham, T. (2013). Development effects of electrification: Evidence from the topographic placement of hydropower plants in Brazil. *American Economic Journal: Applied Economics*, *5*(2), 200–231.

Matsuyama, K. (1992). Agricultural productivity, comparative advantage, and economic growth. *Journal of Economic Theory*, *58*(2), 317–334.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the Interstate Highway System. *The Review of Economics and Statistics*, *90*(4), 683–701.

Michaels, G., Rauch, F. & Redding, S. J. (2011). Technical note: an Eaton and Kortum (2002) model of urbanization and structural transformation. *Princeton University mimeo*.

Nagy, D. K. (2017). City location and economic development. *mimeo*.

Ngai, L. R. & Pissarides, C. A. (2007). Structural change in a multisector model of growth. *American Economic Review*, *97*(1), 429–443.

Nurkse, R. (1953). *Problems of capital formation in underdeveloped countries.* Oxford: Blackwell.

Redding, S. J. (2016). Goods trade, factor mobility and welfare. *Princeton University mimeo*.

Redding, S. J. & Turner, M. (2014). Transportation costs and the spatial organization of economic activity. *NBER Working Paper Series*, (20235).

Rostow, W. W. (1960). The problem of achieving and maintaining a high rate of economic growth: A historian's view. *American Economic Review*, *50*(2), 106–118.

Rud, J. P. (2012). Electricity provision and industrial development: Evidence from India. *Journal of Development Economics*, *97*(2).

Sanderson, E. & Windmeijer, F. (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of Econometrics*, *190*(2), 212–221.

Schultz, T. W. (1953). *The economic organization of agriculture.* New York: McGraw-Hill.

Shiferaw, A., Söderbom, M., Siba, E. & Alemu, G. (2015). Road infrastructure and enterprise dynamics in Ethiopia. *The Journal of Development Studies*, *51*(11), 1–18.

Uy, T., Yi, K.-M. & Zhang, J. (2013). Structural change in an open economy. *Journal of Monetary Economics*, *60*(6), 667–682.

World Bank. (2016). World Development Indicators. *Version: 17 November, 2016*.

World Bank. (2018). Ethiopia Electrification Program: Proposed regular credit and a proposed scale-up facility credit. *Program Appraisal Document*.

Young, A. (2012). The african growth miracle. *Journal of Political Economy*, *120*(4), 696–739.

Chapter 8

Figures

Figure 8.1: Kuznets' Growth Fact: Structural Transformation out of Agriculture

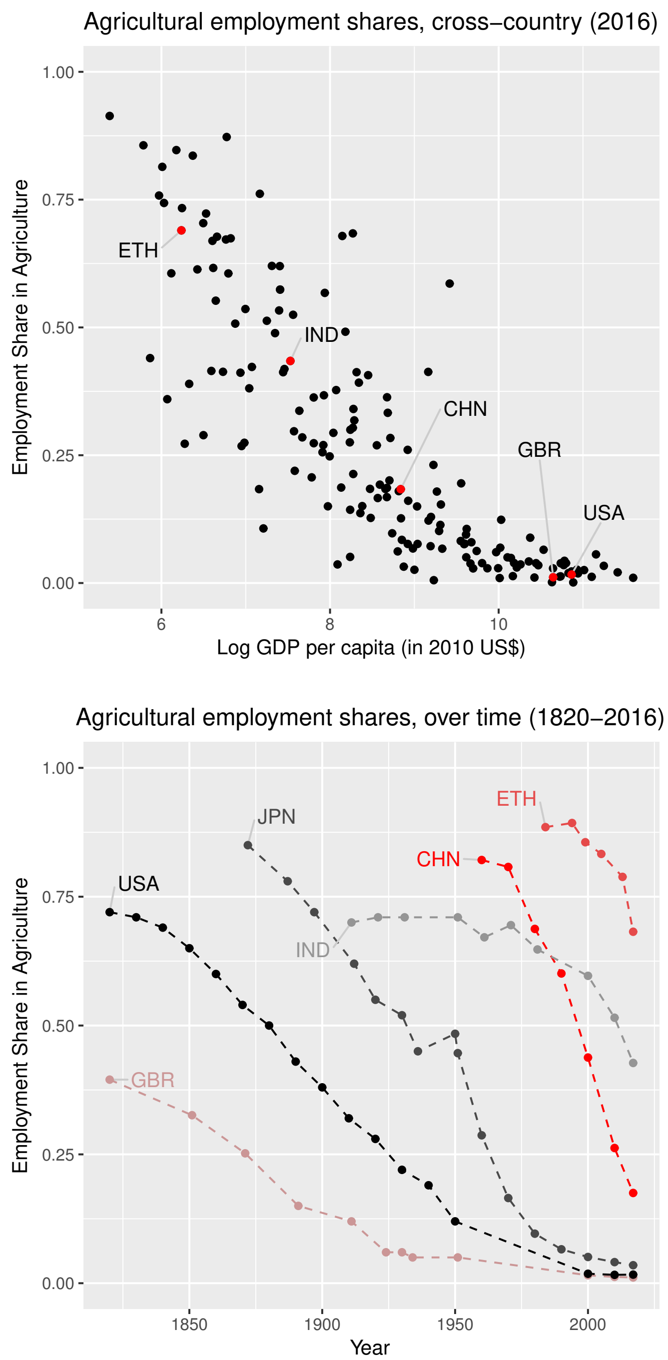
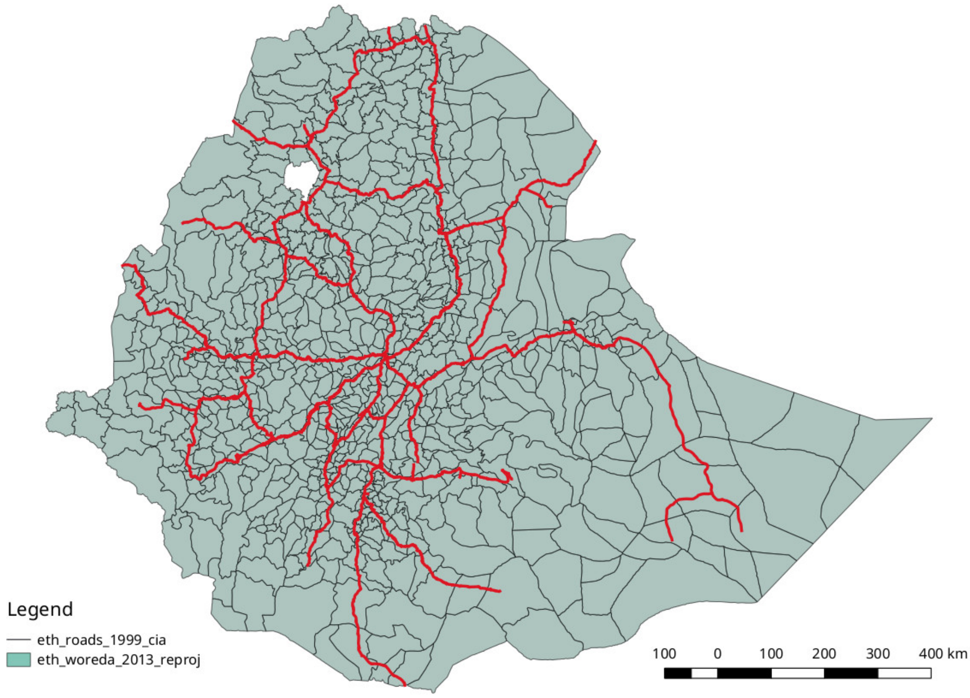


Figure 8.2: Large-scale Road Network Expansion (2000-2016)



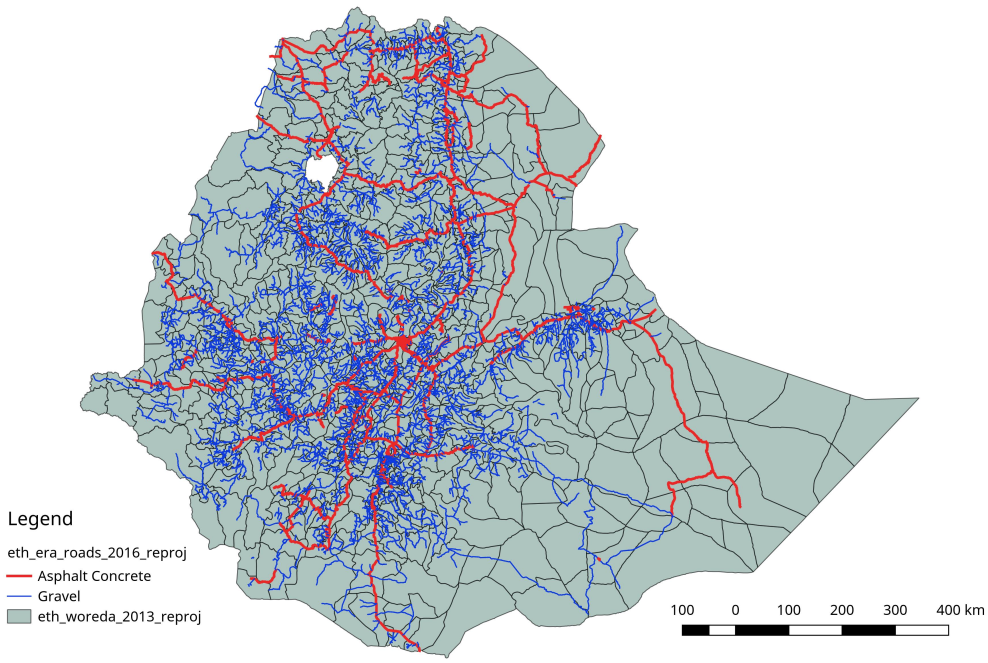
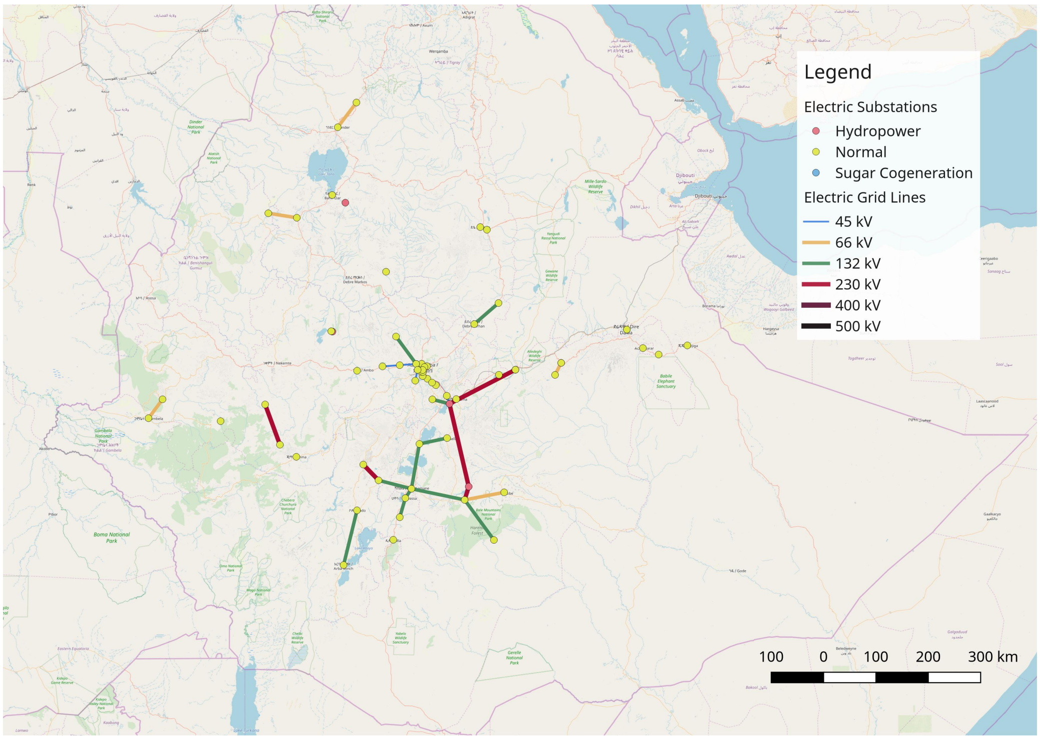


Figure 8.3: Large-scale Electricity Network Expansion (1990-2016)



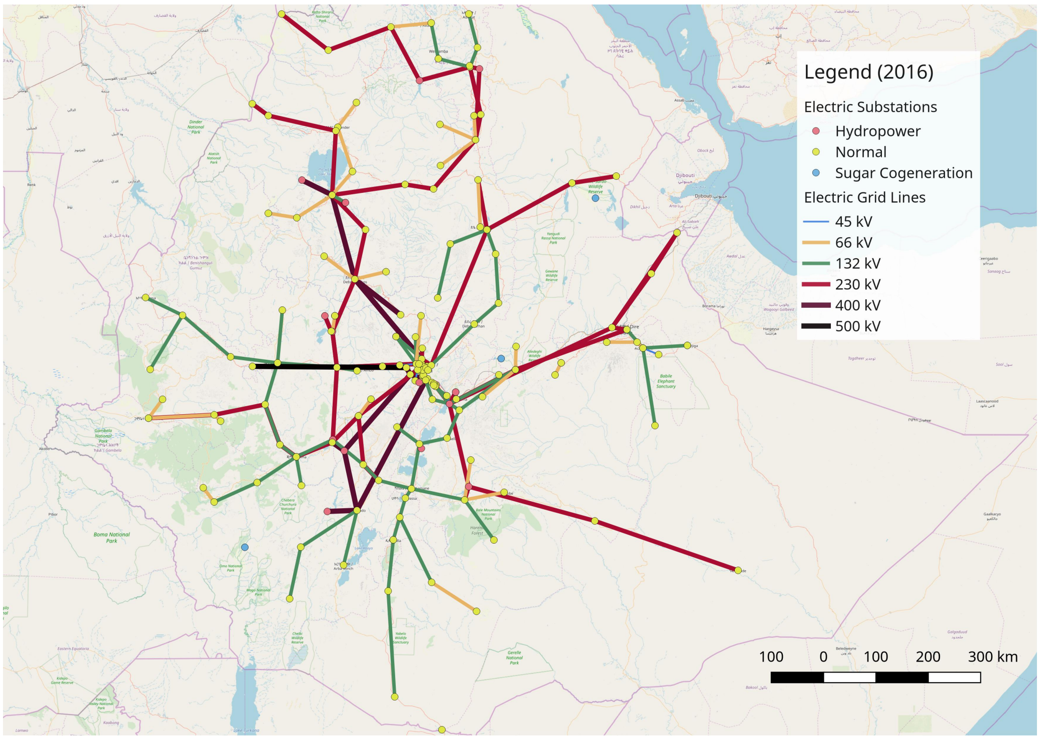


Figure 8.4: Sectoral Employment in Ethiopia (1994-2016)

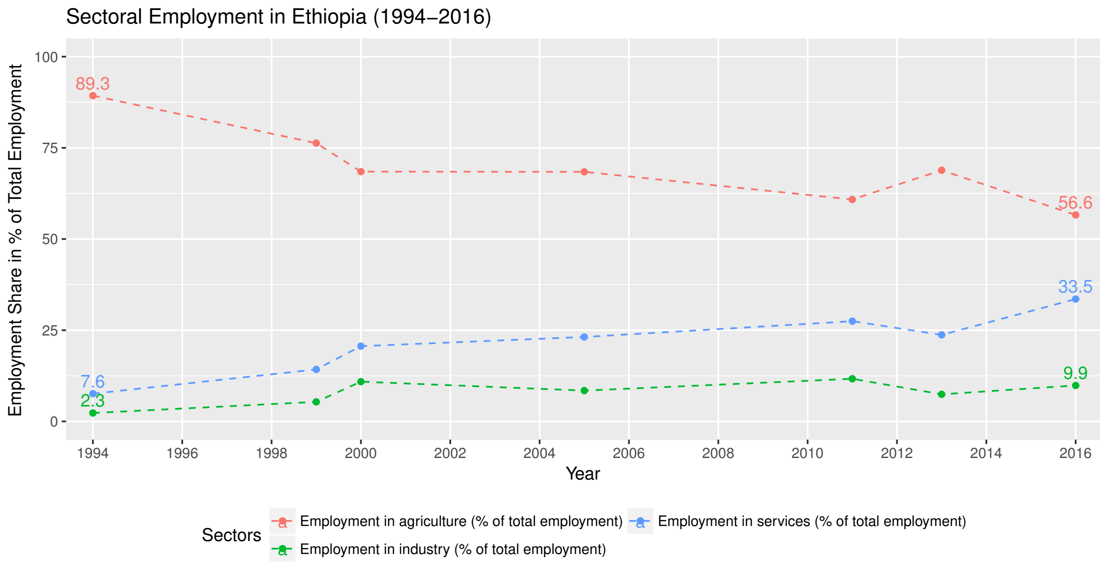


Figure 8.5: Poverty Headcounts and GDP per Capita in Ethiopia (1994-2016)

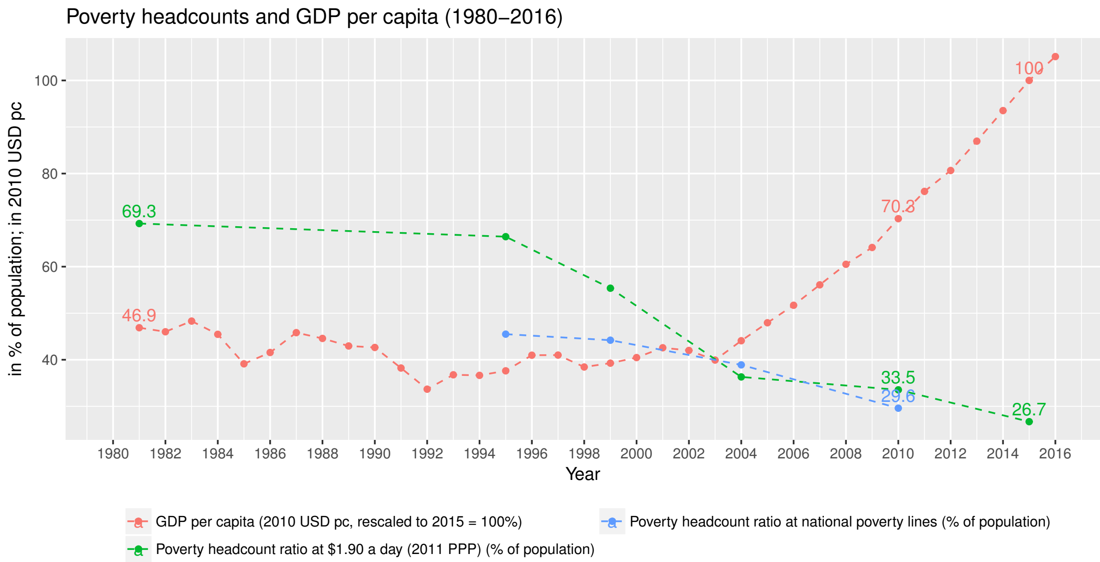


Figure 8.6: Electrification IV Corridors and Times, Connecting Dams with Addis Abeba

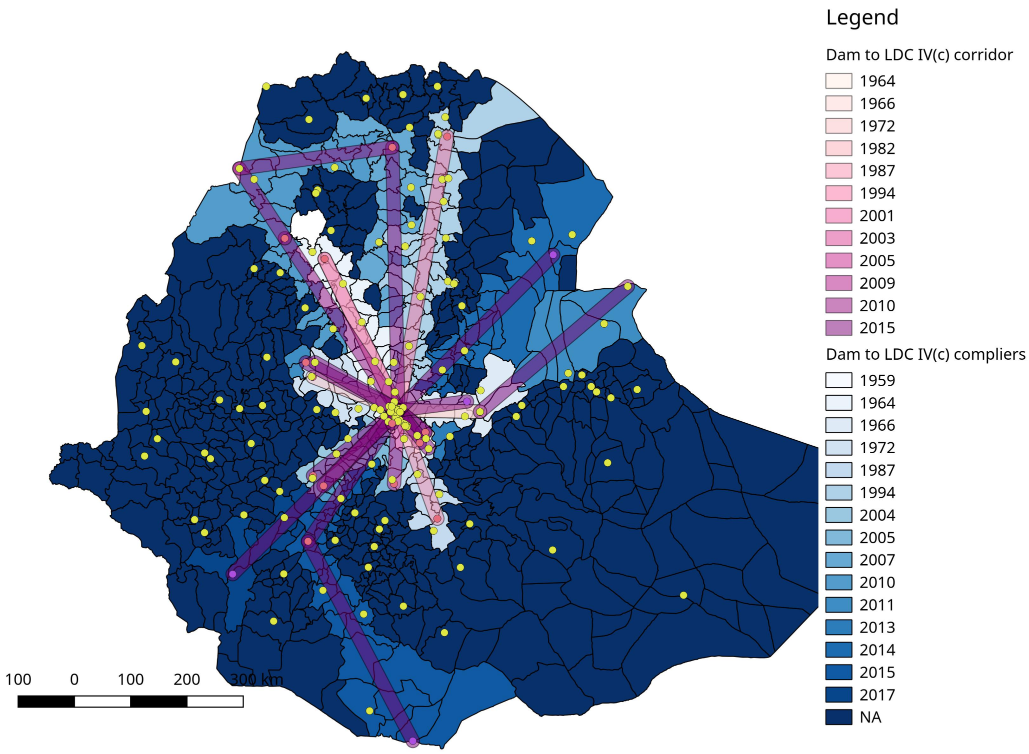


Figure 8.7: Road IV (Italian) District Connection Year to All-weather Road

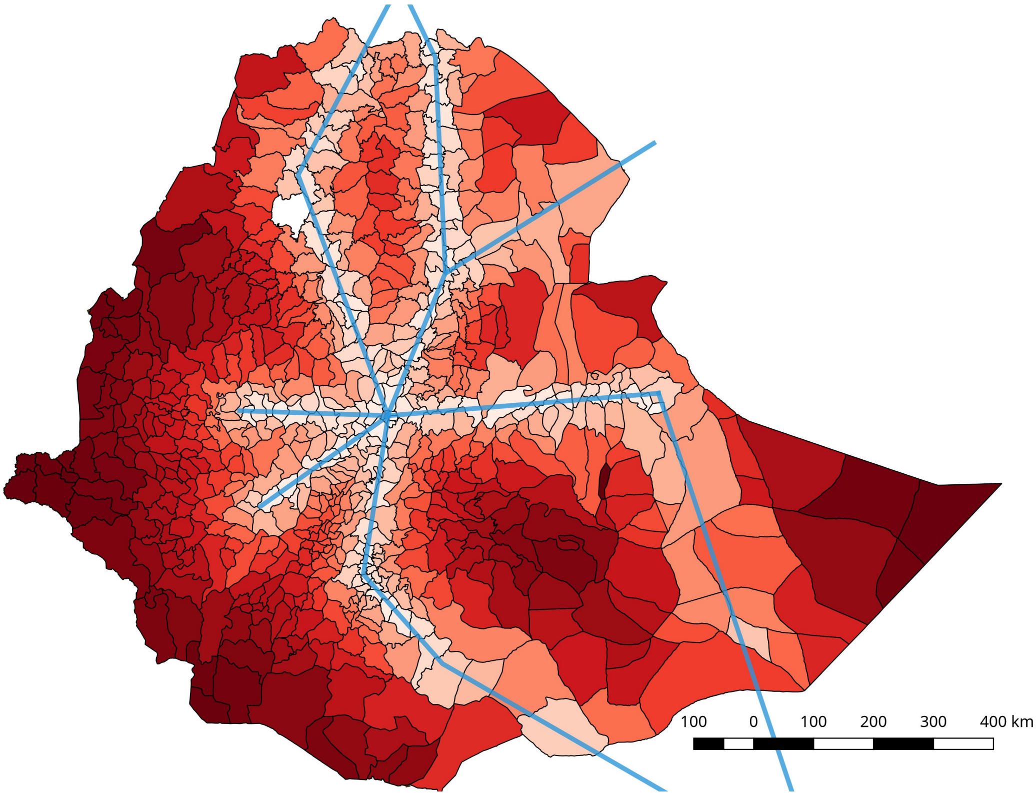


Figure 8.8: Roads and Roads & Electricity Interaction Coefficients by Occupational Groups (in NLFS or DHS-R dataset)

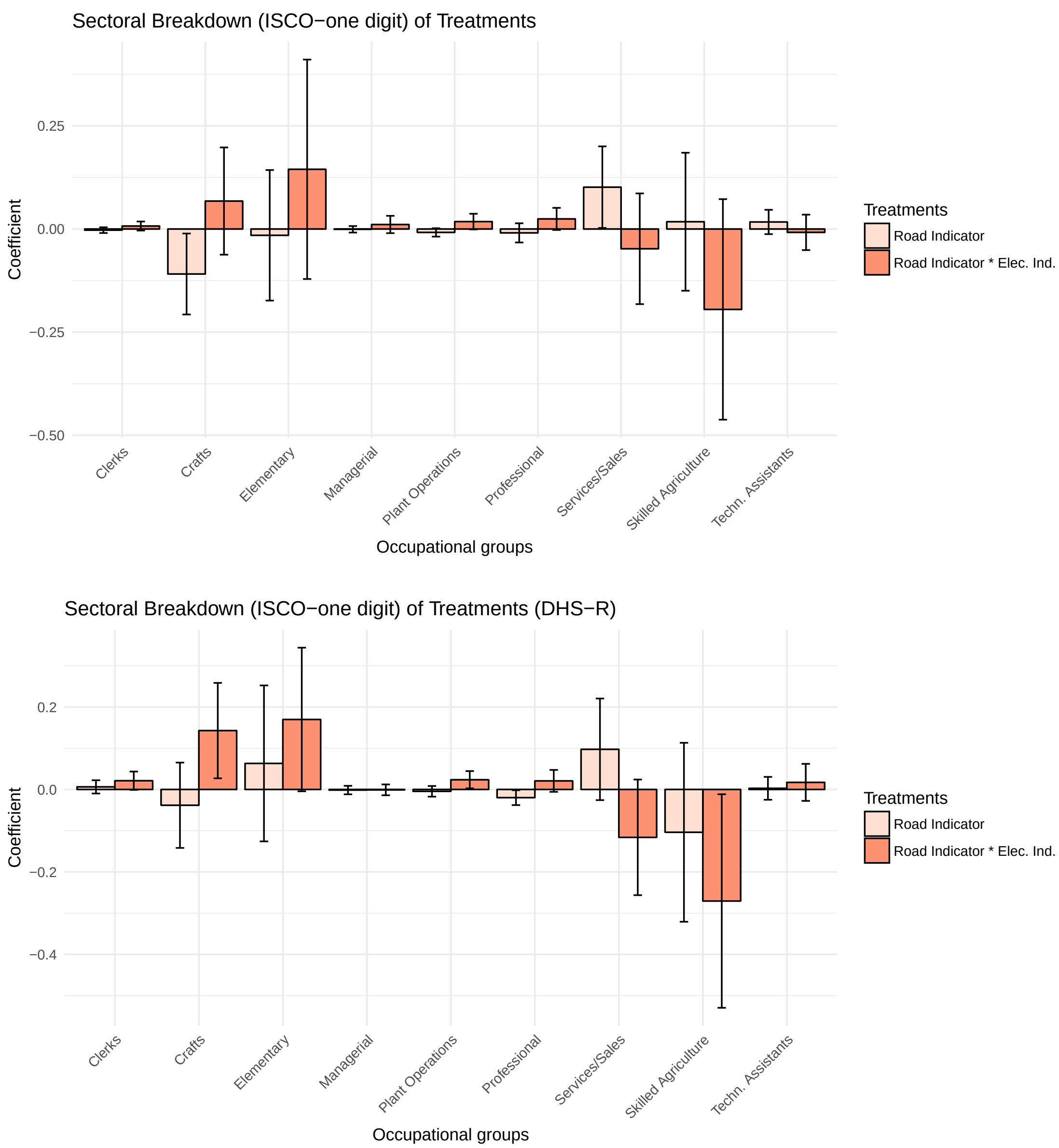


Figure 8.9: Simulated Change in Manufacturing Shares from Trade Cost Shock

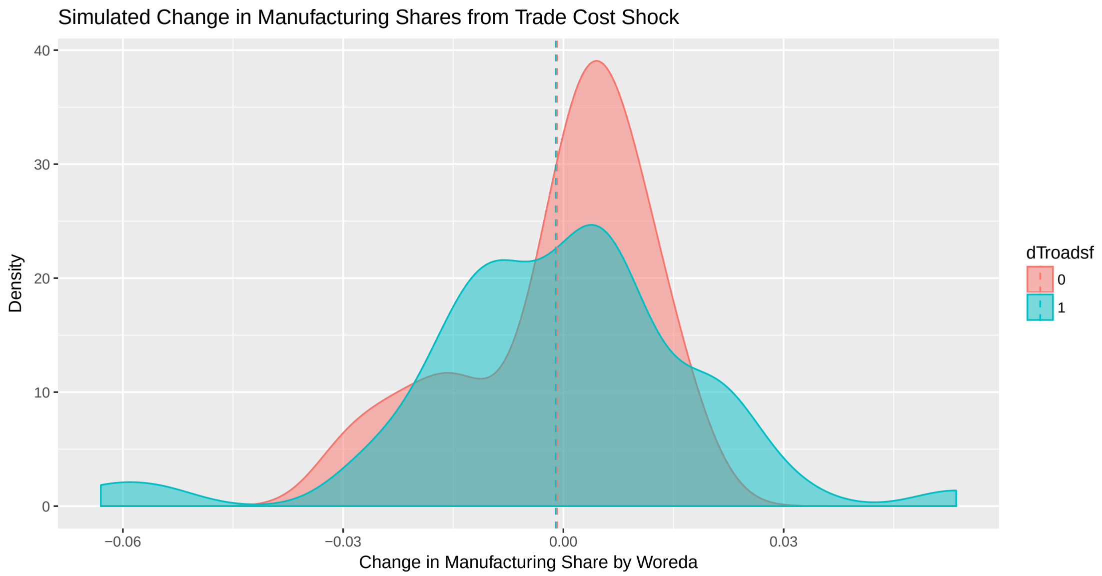


Figure 8.10: Simulated Change in Manufacturing Shares from Combined Trade Cost and Electrification Shock

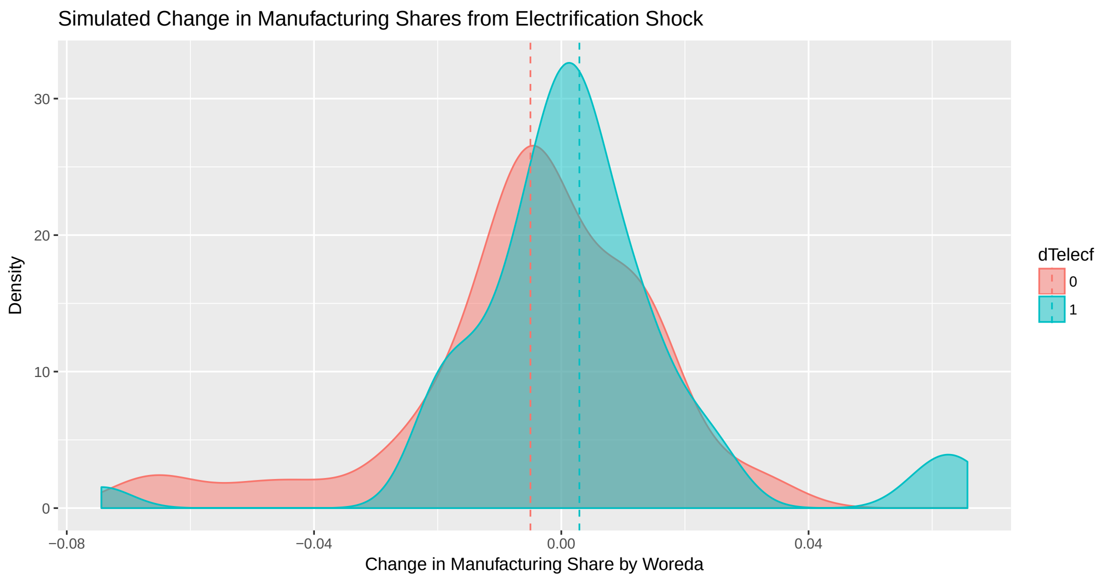


Figure 8.11: Age Distributions by Treatment Complier Status

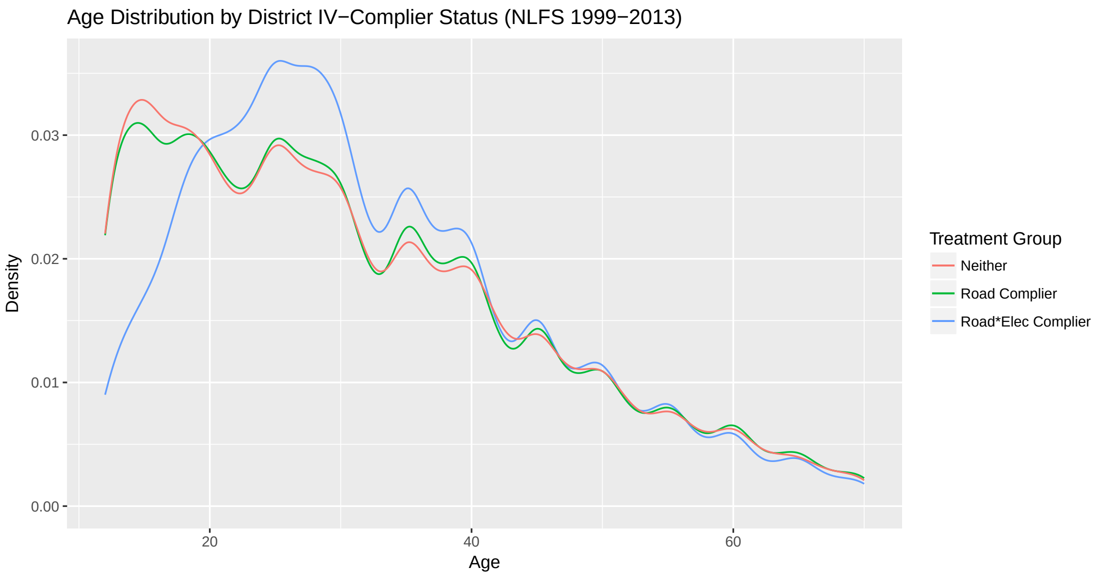
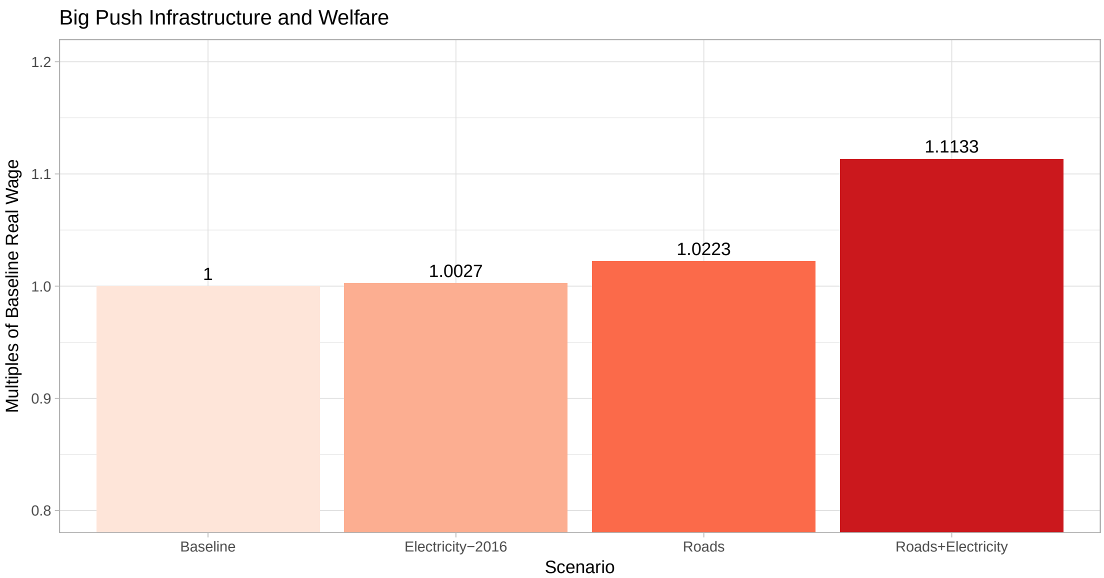


Figure 8.12: Welfare Estimates of Big Push Infrastructure



Chapter 9

Tables

Table 9.1: Roads and Electrification Indicators in NLFS sample (1999-2013)

		Road Ind.		
		0	1	Total
Elec. Ind.	0	328	630	958
	1	7	243	250
	Total	335	873	1208

Table 9.2: Roads and Electrification Indicators in DHS-R sample (2000-2016)

		Road Ind.		
		0	1	Total
Elec. Ind.	0	239	549	788
	1	8	243	251
	Total	247	792	1039

Table 9.3: First Stage: Roads-IV and Elec.-IV int., controls (1999-2013)

	Dependent variable:	
	Roads Ind. NLFS	Roads*Elec Ind. NLFS
	(1)	(2)
Road IV	0.169***	0.002
	(0.040)	(0.034)
Road IV*Elec IV	0.086***	0.197***
	(0.031)	(0.047)
Year FE	✓	✓
Controls	✓	✓
Cragg-Donald F.	9.993	9.993
Windmeijer cond. F.	16.747	13.143
F-test statistic	35.826	36.303
Observations	1,208	1,208
R^2	0.248	0.250
Adjusted R^2	0.241	0.243
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 9.4: First Stage: Roads-IV and Elec.-IV int., controls (2000-2016)

	Dependent variable:	
	Roads Ind. DHS	Roads*Elec Ind. DHS
	(1)	(2)
Road IV	0.188***	-0.037
	(0.048)	(0.035)
Road IV*Elec IV	0.097***	0.243***
	(0.025)	(0.050)
Year FE	✓	✓
Controls	✓	✓
Cragg-Donald F.	13.94	13.94
Windmeijer cond. F.	19.613	16.27
F-test statistic	40.998	42.623
Observations	1,039	1,039
R^2	0.264	0.272
Adjusted R^2	0.258	0.265
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 9.5: Occup. Change (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:		
	Agriculture	Services	Manufacturing
	(1)	(2)	(3)
Road Indicator	-0.072	0.192^{**}	-0.115^{*}
	(0.111)	(0.088)	(0.060)
Road*Elec Ind.	-0.202^{*}	0.070	0.131^{**}
	(0.113)	(0.087)	(0.059)
Model	2SLS	2SLS	2SLS
Year FE	✓	✓	✓
Controls	✓	✓	✓
Cragg-Donald F.		9.993	
Windmeijer cond. F.	16.747	13.143	
p-val $\beta_1 + \beta_2 = 0$	0.007	7e-04	0.7581
Observations	1,208	1,208	1,208
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 9.6: Occup. Change (NLFS-ISIC, excl. Somali), Roads and Elec. (1999-2013)

	Dependent variable:		
	Agr. [isic]	Ser. [isic]	Man. [isic]
	(1)	(2)	(3)
Road Indicator	-0.085	0.216^{**}	-0.125^{**}
	(0.110)	(0.099)	(0.062)
Road*Elec Ind.	-0.196	-0.041	0.232^{***}
	(0.144)	(0.126)	(0.089)
Model	2SLS	2SLS	2SLS
Year FE	✓	✓	✓
Controls	✓	✓	✓
Cragg-Donald F.		6.052	
Windmeijer cond. F.	15.261	10.377	
p-val $\beta_1 + \beta_2 = 0$	0.0066	0.0483	0.1181
Observations	1,188	1,188	1,188
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 9.7: Occup. Change (DHS-R), Roads and Elec. (2000-2016)

	Dependent variable:		
	Agriculture	Services	Manufacturing
	(1)	(2)	(3)
Road Indicator	-0.192	0.228^{**}	-0.035
	(0.128)	(0.105)	(0.065)
Road*Elec Ind.	-0.223	0.014	0.216^{***}
	(0.143)	(0.105)	(0.078)
Model	2SLS	2SLS	2SLS
Year FE	✓	✓	✓
Controls	✓	✓	✓
Cragg-Donald F.		22.328	
Windmeijer cond. F.	19.613	16.27	
p-val $\beta_1 + \beta_2 = 0$	0.0051	0.0291	0.0201
Observations	1,039	1,039	1,039

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9.8: Occup. Change (NLFS, zone cap. dist.), Roads and Elec. (1999-2013)

	$\geq$ med(zone capital dist.)			$<$ med(zone capital dist.)		
	Agr.	Ser.	Man.	Agr.	Ser.	Man.
	(1)	(2)	(3)	(4)	(5)	(6)
Road Indicator	-0.147	0.255^{*}	-0.106	0.022	0.134	-0.144
	(0.154)	(0.131)	(0.084)	(0.183)	(0.129)	(0.103)
Road*Elec Ind.	0.154	-0.208	0.045	-0.313^{**}	0.142	0.169^{**}
	(0.290)	(0.268)	(0.161)	(0.138)	(0.098)	(0.073)
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
Cragg-Donald F.		3.581			8.805	
Windmeijer cond. F.	8.25	2.444		6.17	7.437	
p-val $\beta_1 + \beta_2 = 0$	0.9753	0.837	0.644	0.0571	0.0093	0.7486
Observations	604	604	604	604	604	604

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9.9: Employment Relations (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:	
	Employed	Self-Employed
	(1)	(2)
Road Indicator	0.028	-0.024
	(0.067)	(0.067)
Road*Elec Ind.	0.241^{**}	-0.240^{**}
	(0.097)	(0.097)
Model	2SLS	2SLS
Year FE	✓	✓
Controls	✓	✓
Cragg-Donald F.		6.272
Windmeijer cond. F.	16.334	11.003
p-val $\beta_1 + \beta_2 = 0$	4e-04	5e-04
Observations	1,208	1,208
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 9.10: Migration (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:			
	Mig.<1yr	Mig.<2yr	Mig.<6yr	Mig. ever
	(1)	(2)	(3)	(4)
Road Indicator	0.001	−0.010	−0.036	−0.078
	(0.015)	(0.023)	(0.051)	(0.102)
Road*Elec Ind.	0.026*	0.043*	0.111**	0.197*
	(0.015)	(0.024)	(0.052)	(0.105)
Model	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Cragg-Donald F.		9.993		
Windmeijer cond. F.	16.747	13.143		
p-val $\beta_1 + \beta_2 = 0$	0.0425	0.1092	0.0857	0.1948
Observations	1,208	1,208	1,208	1,208

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9.11: Consumption (HCES), Roads and Elec. (2000-2016)

	Dependent variable:		
	HH Exp. (pc)	HH Size	HH Age
	(1)	(2)	(3)
Road Indicator	-641.58	0.51	0.03
	(1,448.17)	(0.42)	(1.60)
Road*Elec Ind.	$4{,}854.19^{*}$	-1.34^{**}	4.80^{*}
	(2,790.13)	(0.61)	(2.67)
Model	2SLS	2SLS	2SLS
Year FE	✓	✓	✓
Controls	✓	✓	✓
Cragg-Donald F.		8.241	
Windmeijer cond. F.	26.111	10.042	
p-val $\beta_1 + \beta_2 = 0$	0.1116	0.157	0.0713
Observations	572	572	572
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 9.12: Durables Exp. (DHS-HR), Roads and Elec. (2000-2016)

	Dependent variable:						
	Radio	TV	Refrig.	Bike	Scooter	Car	Phone
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Road Indicator	0.068	-0.113^{**}	-0.047	0.002	0.002	-0.006	0.005
	(0.074)	(0.046)	(0.058)	(0.016)	(0.004)	(0.006)	(0.019)
Road*Elec Ind.	0.171^{*}	0.175^{**}	0.098^{**}	0.005	-0.013	0.024^{*}	0.083^{**}
	(0.099)	(0.079)	(0.048)	(0.019)	(0.010)	(0.012)	(0.038)
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Cragg-Donald F.	15.257	15.257	5.632	15.257	15.257	15.257	15.257
Windmeijer cond. F. (I)	19.613	19.613	19.613	19.613	19.613	19.613	19.613
Windmeijer cond. F. (II)	16.27	16.27	16.27	16.27	16.27	16.27	16.27
p-val $\beta_1 + \beta_2 = 0$	0.3832	0.3659	0.714	0.2637	0.1458	0.0163	0.0197
Observations	1,039	1,039	788	1,039	1,039	1,039	1,039

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9.13: Housing Exp. (DHS-HR), Roads and Elec. (2000-2016)

	Dependent variable:				
	Elec.	Tap Water	Flush Toilet	Floor	Ln(Rooms pp.)
	(1)	(2)	(3)	(4)	(5)
Road Indicator	0.042	-0.023	-0.036^{*}	0.093	0.051
	(0.084)	(0.132)	(0.021)	(0.068)	(0.151)
Road*Elec Ind.	0.308^{**}	0.533^{***}	0.054^{*}	0.124	0.087
	(0.145)	(0.178)	(0.030)	(0.110)	(0.133)
Model	2SLS	2SLS	2SLS	2SLS	2SLS
Cragg-Donald F.	15.257	15.257	15.257	15.257	11.258
Windmeijer cond. F. (I)	19.613	19.613	19.613	19.613	4.813
Windmeijer cond. F. (II)	16.27	16.27	16.27	16.27	12.963
p-val $\beta_1 + \beta_2 = 0$	0.0197	0.0072	0.4915	0.0438	0.5127
Observations	1,039	1,039	1,039	1,039	540

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9.14: Satellite Outcomes, Roads and Elec. (2000-2016)

	Dependent variable:		
	Log Pop.	Built-up	Nightlights
	(1)	(2)	(3)
Log Pop. Initial	1.002^{***}	0.089	-0.010
	(0.006)	(0.076)	(0.094)
Nightlights Initial	0.0001	1.194^{***}	1.028^{***}
	(0.0005)	(0.072)	(0.027)
Road Indicator	0.057^{**}	-1.626^{***}	-1.033^{*}
	(0.023)	(0.527)	(0.531)
Road*Elec Ind.	-0.143^{***}	1.023	2.312^{**}
	(0.039)	(0.643)	(0.980)
Model	2SLS	2SLS	2SLS
Year FE	✓	✓	✓
Controls	✓	✓	✓
Cragg-Donald F.	42.909	30.454	27.263
Windmeijer cond. F. (I)	238.793	105.775	41.396
Windmeijer cond. F. (II)	37.785	24.858	14.465
Observations	2,748	1,374	2,061
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 9.15: Full-panel: Roads and Least-Cost Distances (2000-2016)

	Dependent variable:			
	Log(Sum of Least-cost Distances)			
	(1)	(2)	(3)	(4)
Roads Ind.	-0.203^{***}	-0.114^{***}	-0.045^{***}	-0.040^{***}
	(0.011)	(0.007)	(0.009)	(0.003)
Controls		✓	✓	
Year FE			✓	✓
District FE				✓
Observations	2,752	2,744	2,744	2,752
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 9.16: Parameters for Structural Estimation

Parameter	Value	Source	Description
σ	4	Bernard et al. (2003)	Elasticity of substitution between varieties
$1-\alpha$	0.25	Data (HCES)	Expenditure share on land/housing
κ	0.5	Ngai & Pissarides (2008)	Elasticity of substitution across sectors
μ^M	0.82	Data (LMMIS)	Labour share in M-production
μ^T	0.78	Data (AgSS)	Labour share in T-production
μ^S	0.84	Data (DST)	Labour share in S-production
τ	0.3	Data (RPI)	Elasticity of trade cost with respect to distance
θ	4	Donaldson (2018)	Shape parameter of productivity distribution across varieties & locations

Note: HCES denotes the Central Statistical Agency's (CSA) Household Consumption and Expenditure Surveys; LMMIS denotes the CSA's Large- and Medium-Scale Manufacturing Surveys; AgSS denotes the CSA's Agricultural Sample Surveys; DST denotes the CSA's Distributive and Service Trade Surveys; RPI denotes the CSA's Retail Price Index' raw data.

Appendices

Appendix A

Appendix: Data

For the structural estimation, I require additional data on model inputs that were either not required or not of primary interest in the reduced-form estimation. In particular, I require information on the supply of land and population information for every district in Ethiopia. I also explain how initial sectoral employment shares for the baseline calibration are derived from NLFS and DHS samples. Furthermore, I provide proxy measures for the productivity in either production sector and wages as additional robustness checks on the fit of the model.

With respect to land area, I use arable land derived from satellite imagery, that is land either deemed theoretically inhabitable or suitable for productive use. This choice of proxy is not without potential issues, given that in the model housing for consumers and land used in production are conflated. However, reliable data on the housing stock and its value in Ethiopia is virtually non-existent since the Ethiopian real estate market remains monopolised by government ownership of land, essentially a leftover from the previous, socialist regime in power: all land is owned by the government and firms or residents only obtain non-permanent permission to use their allocated land without technically owning it. Hence, I use arable land as one possible proxy for land supply for which data exists, which appears reasonable given that land enters all sectors' production as input and even consumers' housing demand will reflect demand for land given prevailing housing construction outside of Addis Abeba.

For population data, I employ Census data at the district-level for 2007/2008 in addition to Census-derived, remotely-sensed population estimates for earlier and later years. Although the NLFS and DHS repeated cross-sections do not include useable information on district-level population, I can nonetheless derive information on the share of the working age population and the labour force participation rate from this data. The latter becomes useful in scaling population measures, since large parts of the (on average very young) population are not (yet) active in the labour force. This information is supplemented by the birth histories from the DHS' female questionnaire, which provides useful insights into changes of fertility (and thus population growth) at the district-level across the country and over time.

Regarding initial sectoral employment shares used in the baseline calibration, the manufacturing and service sector shares relate one for one to the manufacturing and services sectors' employment shares in each district as of 1999/2000. In particular, to maximise the sample of available initial data points, I pool both the first National Labour Force Survey (NLFS) round from 1999/2000 and the first Demographic and Health Survey (DHS) round from 2000. Wherever a district contains enumeration areas from both surveys, the manufacturing share of that district represents the average of enumeration areas across surveys. Using both unbalanced samples, I obtain 1999/2000 manufacturing employment share data for 475 out of the total 689 districts used in the analysis. Out of these 475, 181 districts appear only in the NLFS for the initial period, 58 appear only in the DHS and 236 appear in both. For the missing 214 districts, I impute initial employment shares by relying on the fact that both the NLFS and the DHS are representative at the country- and the regional-level. Hence, any interpolation has to preserve the sample mean. We propose three different imputation methods and show sensitivity of my results below: firstly, a naive imputation where every missing district value is replaced with the sample mean. Secondly, a random permutation of this sample mean within one standard deviation, while preserving the overall mean and, thirdly, a more sophisticated regression-based approach that predicts (mean-preserving) employment shares based on observable district characteristics.

Regarding sectoral productivities, I use agricultural yields as a proxy for traditional sector productivity: as shown in recent applications in the remote-sensing literature, remotely-sensed organic carbon content at shallow soil depths (e.g. 5-20cm) performs surprisingly well as a proxy for soil fertility and agricultural productivity when compared against lab-in-the-field measures of either, which are obtained by taking physical soil samples or measuring farmer output.

I can show that district-averages of organic carbon content at five centimetre depths from remotely-sensed data across Ethiopia appear to fit a Weibull distribution of yields well (results available upon request). This empirical finding is particularly interesting given the wide use of extreme value distribution (such as Fréchet, Weibull or Gumbel) properties in spatial general equilibrium models like mine. Since the reciprocal of a two-parameter Weibull-distributed random variable is Fréchet-distributed, one can easily derive the scale and shape parameters of such a distribution. In particular, if $X \sim Frechet(\alpha, s, m = 0)$, then its reciprocal is Weibull-distributed with parameters: $X^{-1} \sim Weibull(k = \alpha, \lambda = s^{-1})$. In my case, fitting a Weibull distribution to the yield data by Maximum Likelihood results in estimates for the scale parameter (A_n^T) of 31.74 ($s.e. = 0.4661$) and for the shape parameter (θ) of 2.75 ($s.e. = 0.0749$).

For the modern sector productivity, I lack any country-wide proxies for it. However, I use crude measures of a TFP residual from the firm-level raw data of repeated cross-sections of the Central Statistical Agency's Large and Medium-Scale Manufacturing and Industry Surveys, as well as the Small-Scale Manufacturing and Industry Surveys. Since the firms sampled in these surveys are predominantly located in towns and cities across Ethiopia, I derive the correlation structure between district-level agricultural yields and survey TFP proxies, whenever available, to then extrapolate from this for all Ethiopian locations.

Finally, I can derive wages or proxies thereof from two distinct sources: firstly, the Retail Price Index raw data contain monthly information on the day rate for unskilled labour in 119 markets from 1998 until 2017 (cf. Subsection 6.1 for more information). This represents a very direct measure of wages, albeit it is only available for the subset of market towns, which may vary quite dramatically from rural or more remote labour market conditions. Secondly, I can exploit household expenditure information per capita provided in the nationally representative HCES surveys (cf. Subsection 3.2 for more information) for sampled enumeration areas throughout Ethiopia, from which at least a subset of district wage proxies can be derived. The remaining gaps in coverage can be filled by small area estimation such as in Balboni (2019).

Appendix B

Additional Figures

Figure A1: Sectoral Value-Added in Ethiopia (1980-2016)

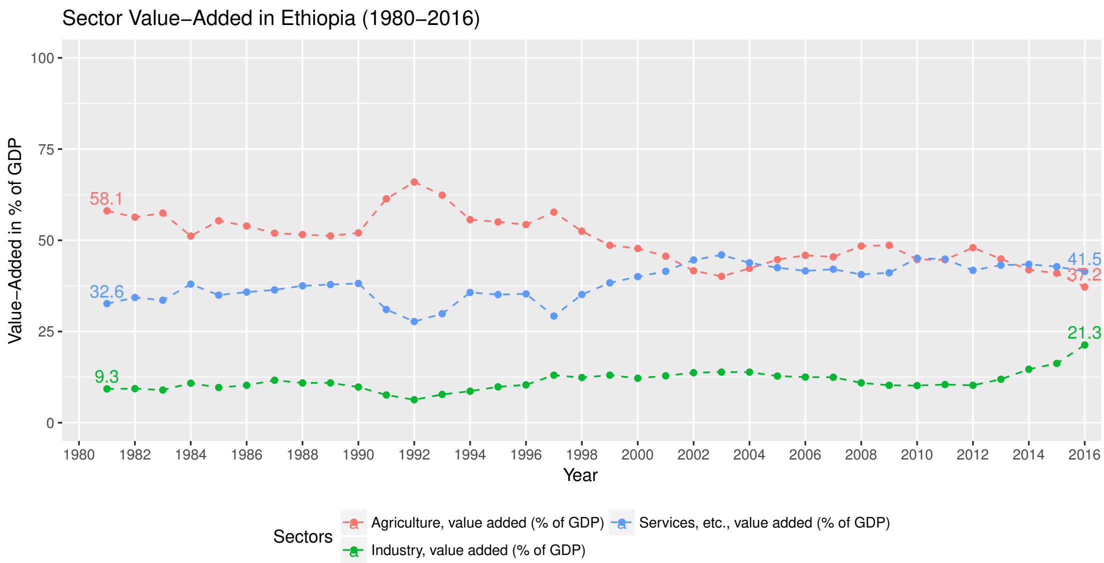


Figure A2: DHS Enumeration Area Locations by Survey Round (2000-2016)

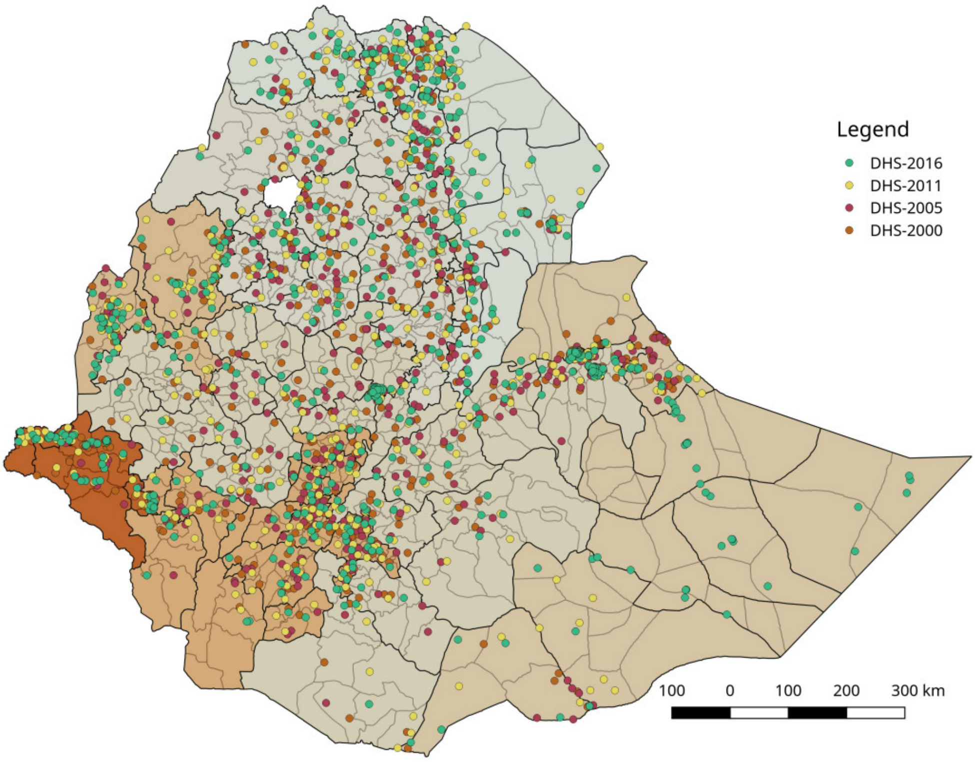


Figure A3: Spatial Variation in Population Density across Ethiopia (2015)

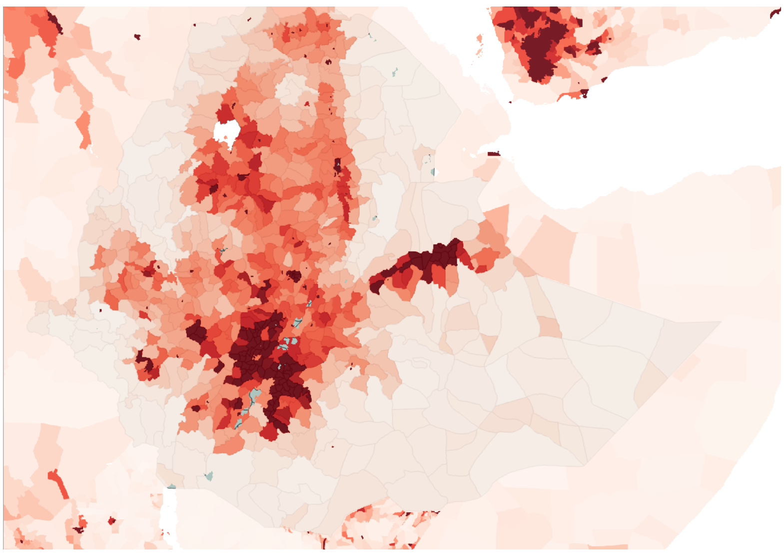


Figure A4: Spatial Variation in Elevation across Ethiopia

Figure A5: Spatial Variation in Terrain Ruggedness across Ethiopia

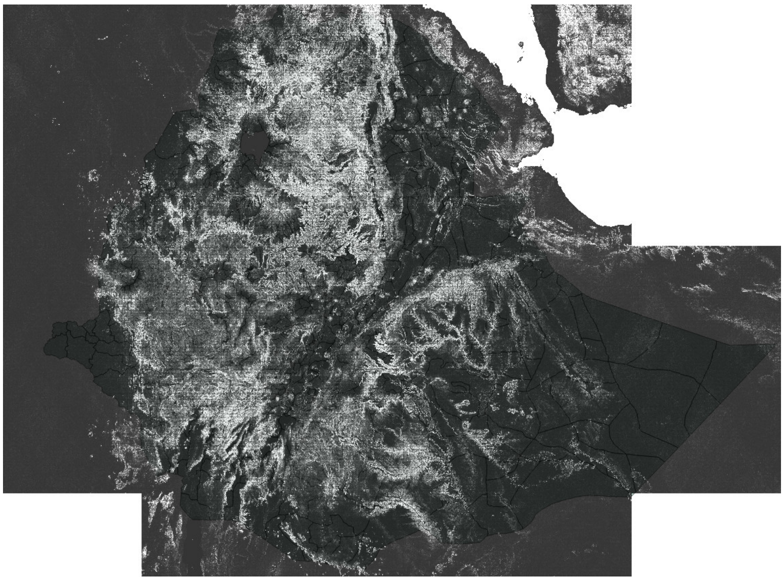


Figure A6: Endogeneity of Infrastructure Allocation across Ethiopia

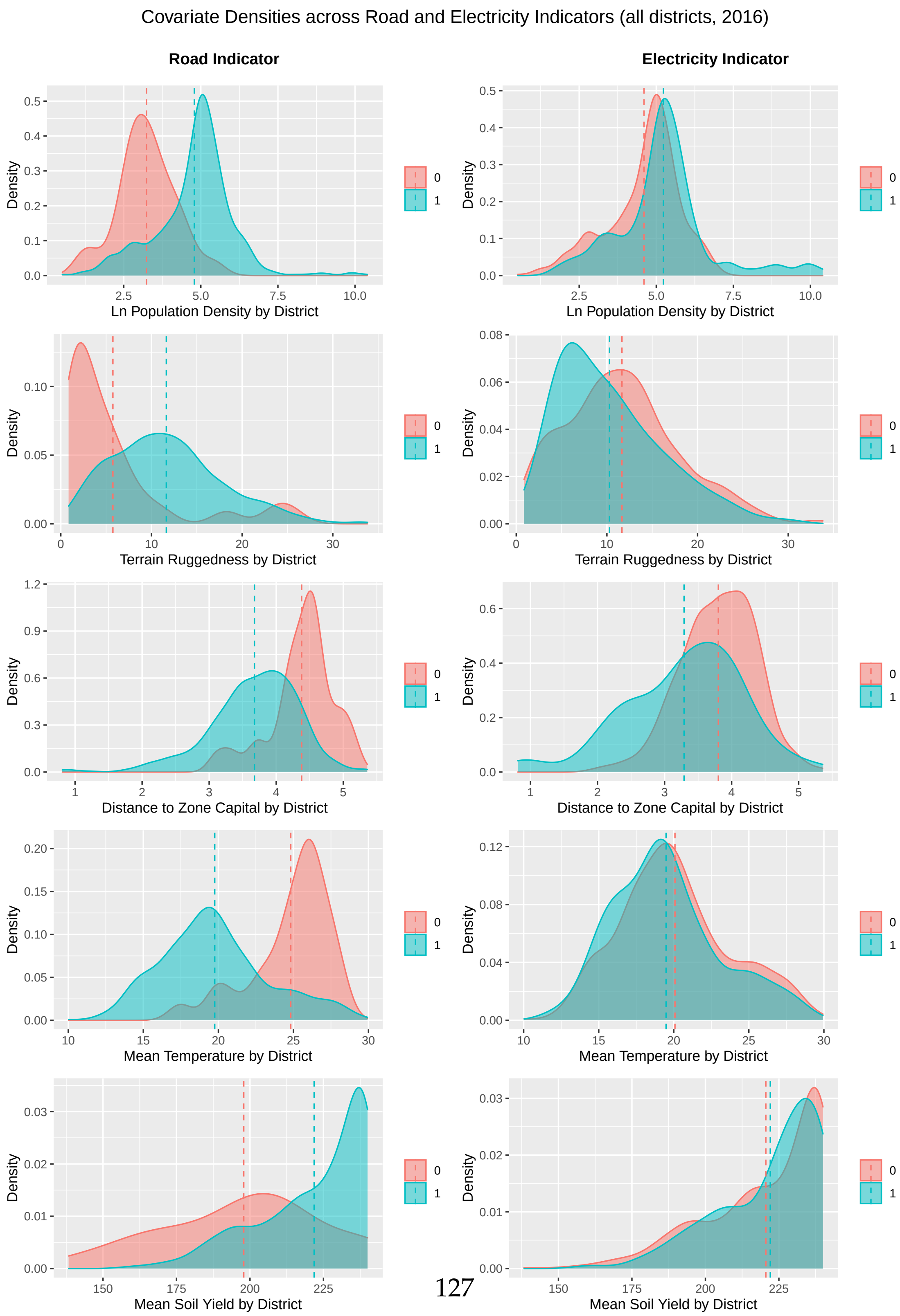


Figure A7: Historic Italian Road Construction in Ethiopia and Eritrea

Figure A8: Reconstructed Italian Colonial Roads and Orthogonal Feeder Roads to Nearby Districts around Debre Berhan (along Dessie–Addis Abeba corridor)

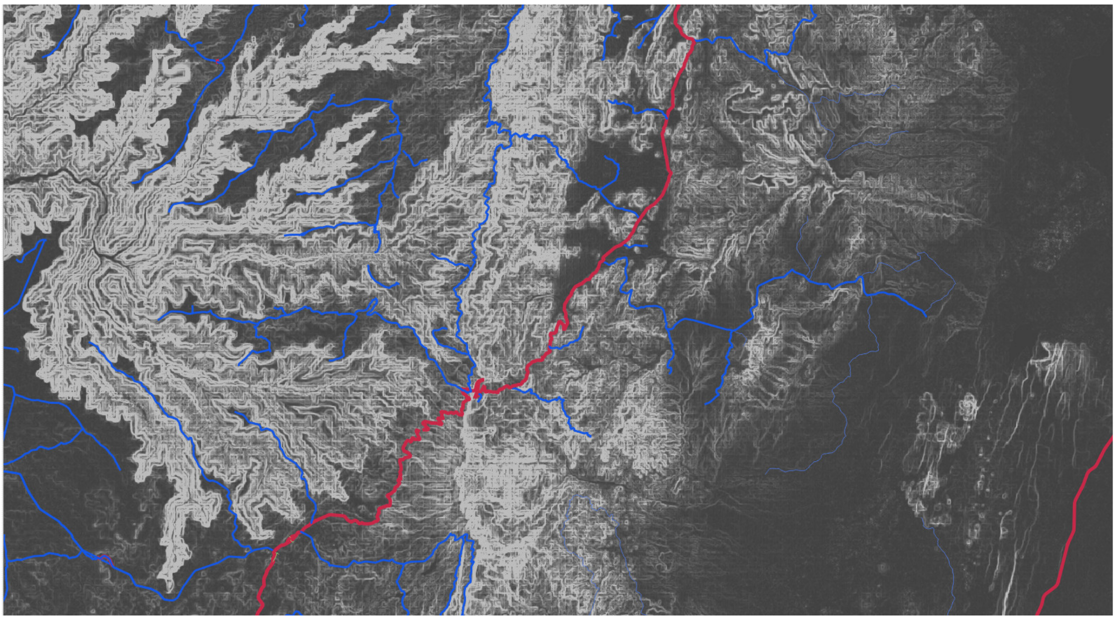


Figure A9: Reconstructed Italian Colonial Roads and Orthogonal Feeder Roads to Nearby Districts around Kulubi (along Harar–Addis Abeba corridor)

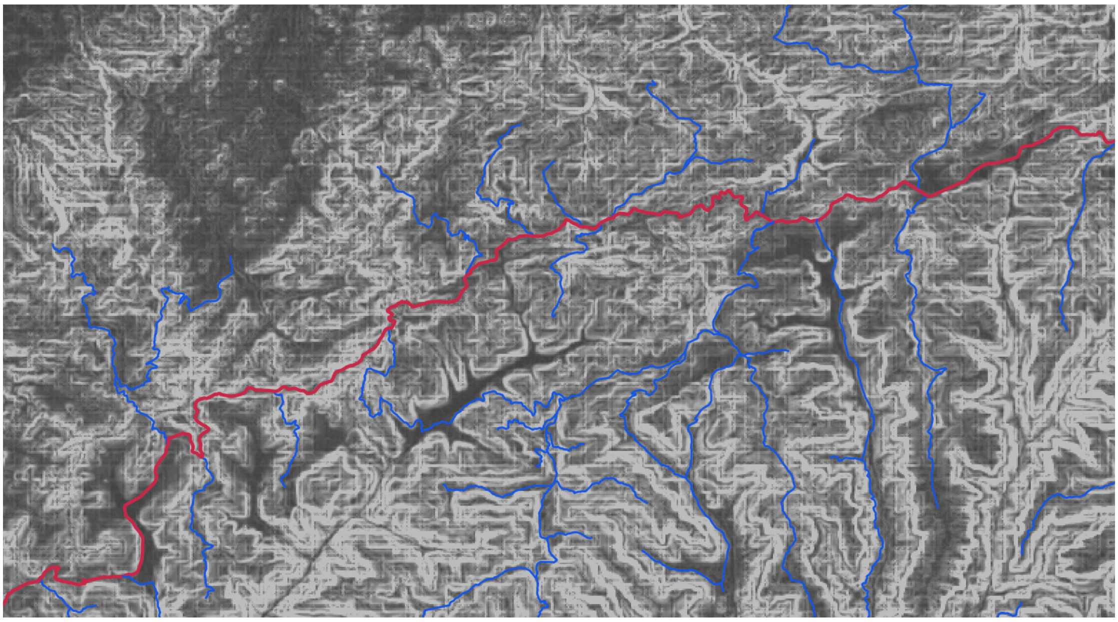


Figure A10: Random Assignment of Electricity Instrument: Covariate Values of Original Districts under Instrument Line Buffer vs Neighbouring Districts

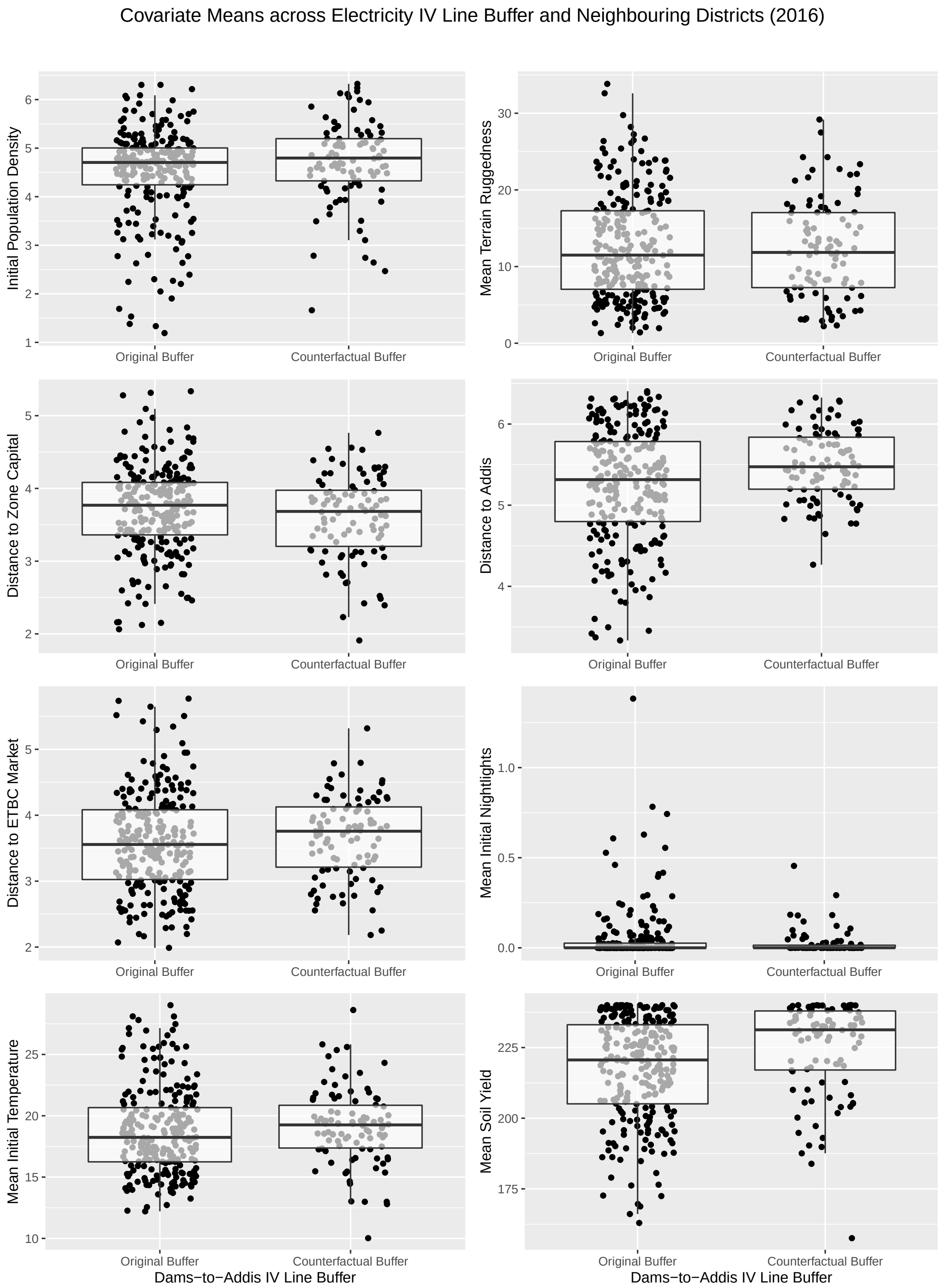
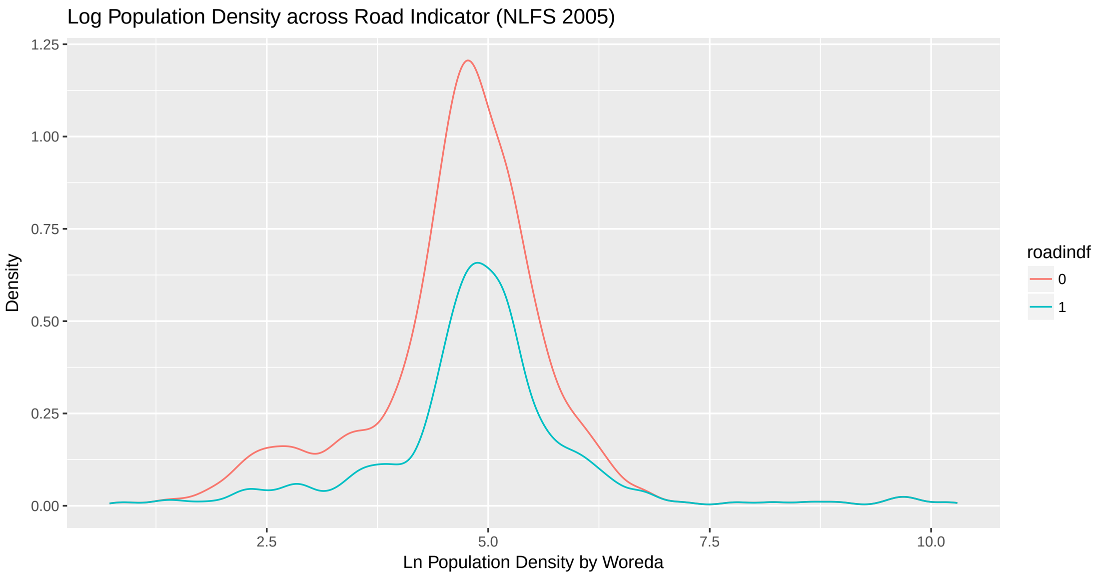


Figure A11: Districts' Road Access Status as Function of Population Density (2005-2013)



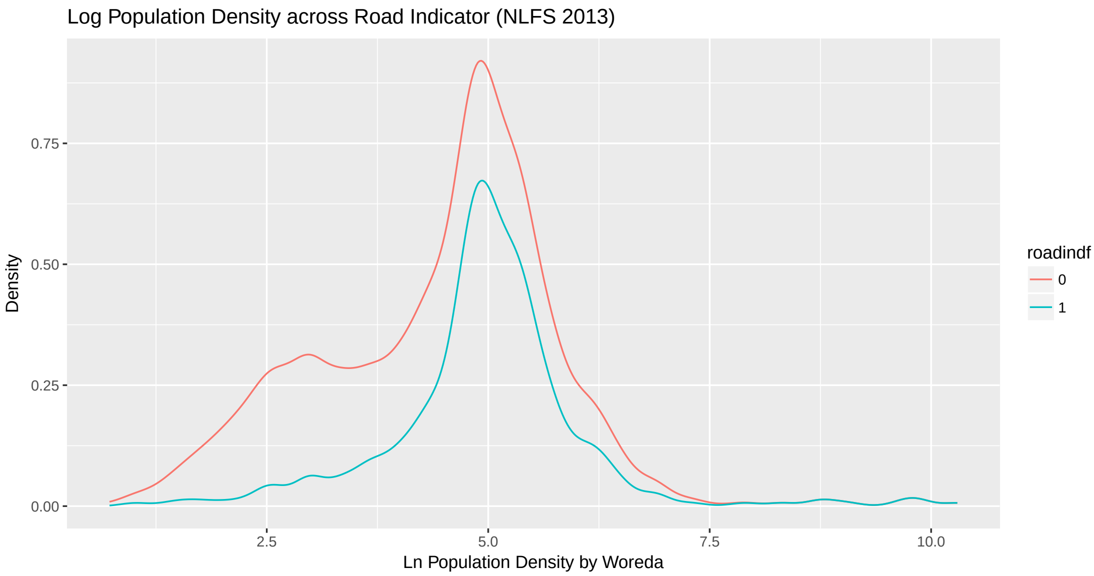


Figure A12: Random Assignment of Roads Instrument: Covariate Values of Pre-Sample Year Districts vs Post-Sample Year Districts

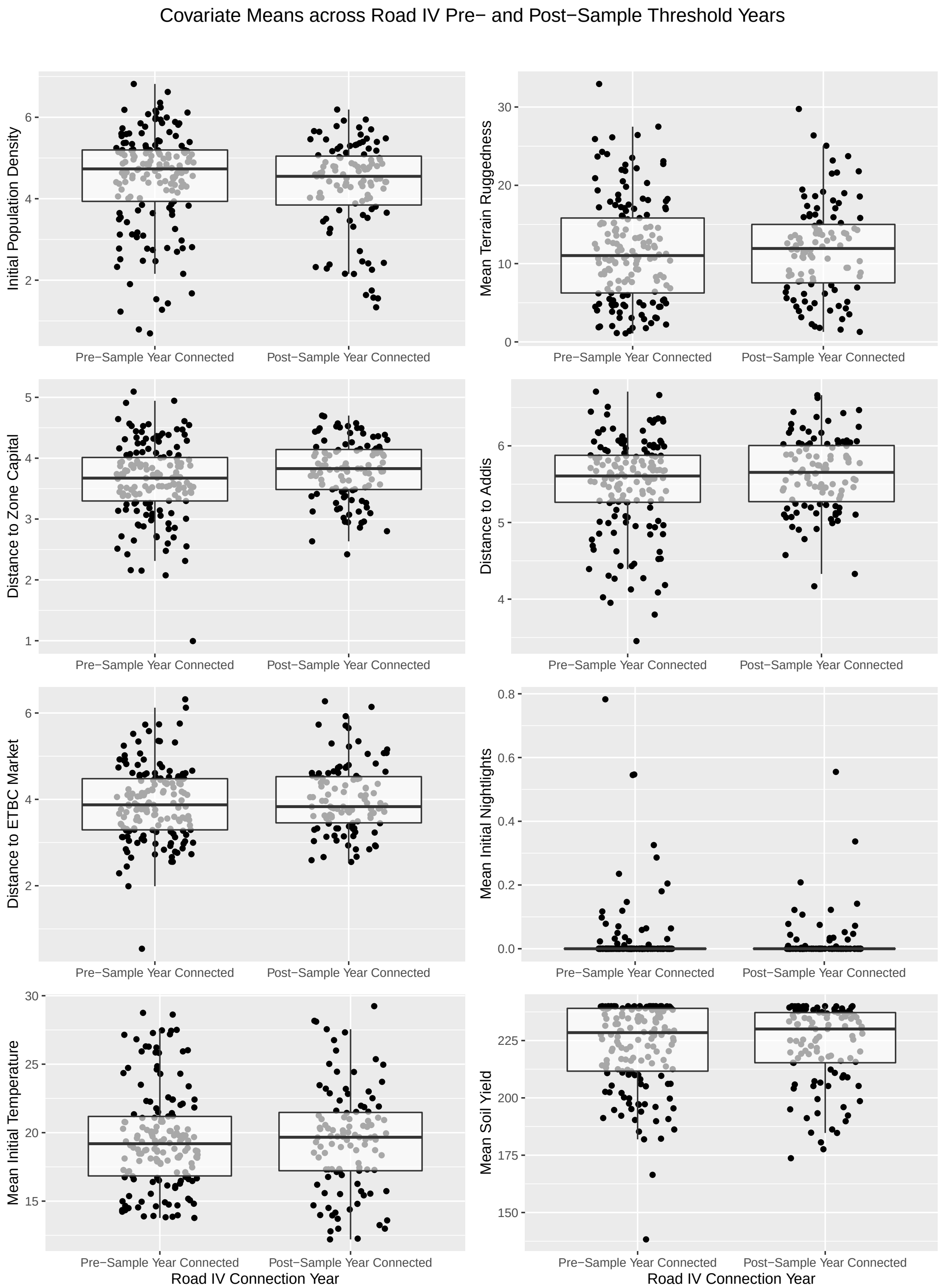


Figure A13: Road IV (Kruskal) District Connection Year to All-Weather Road

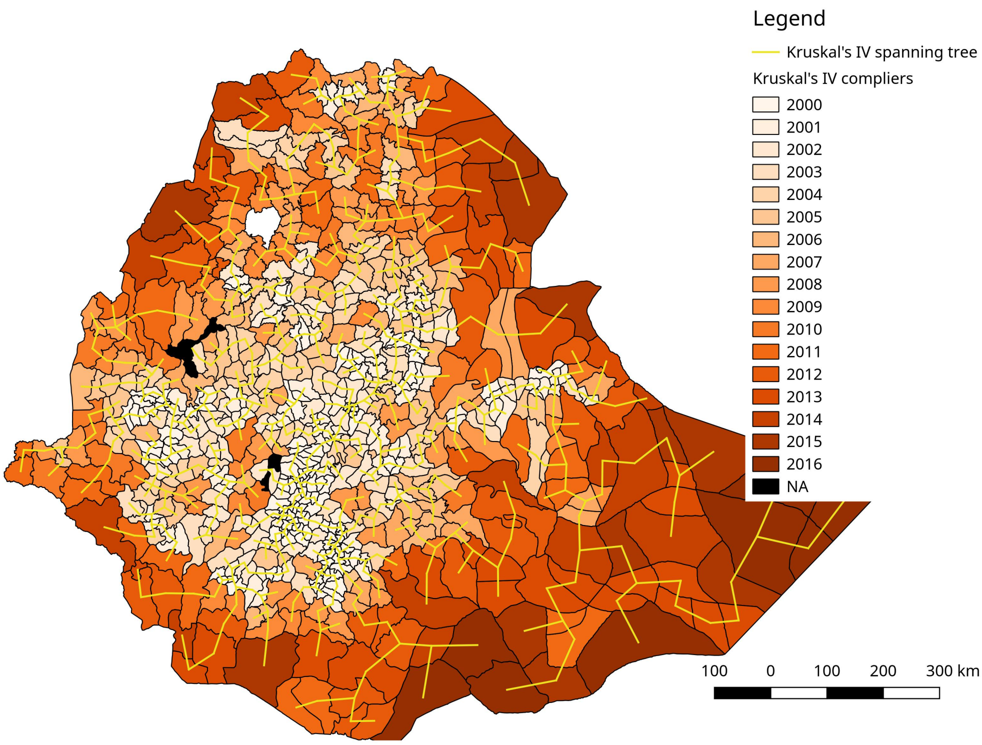
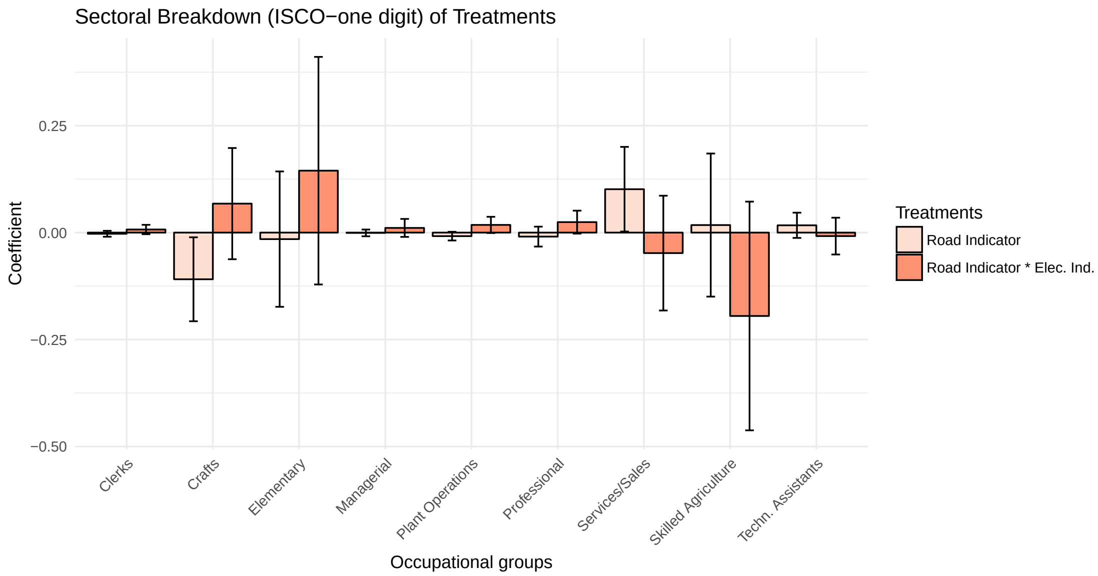


Figure A14: Sectoral Breakdowns (ISCO and ISIC–one digit) of Treatments



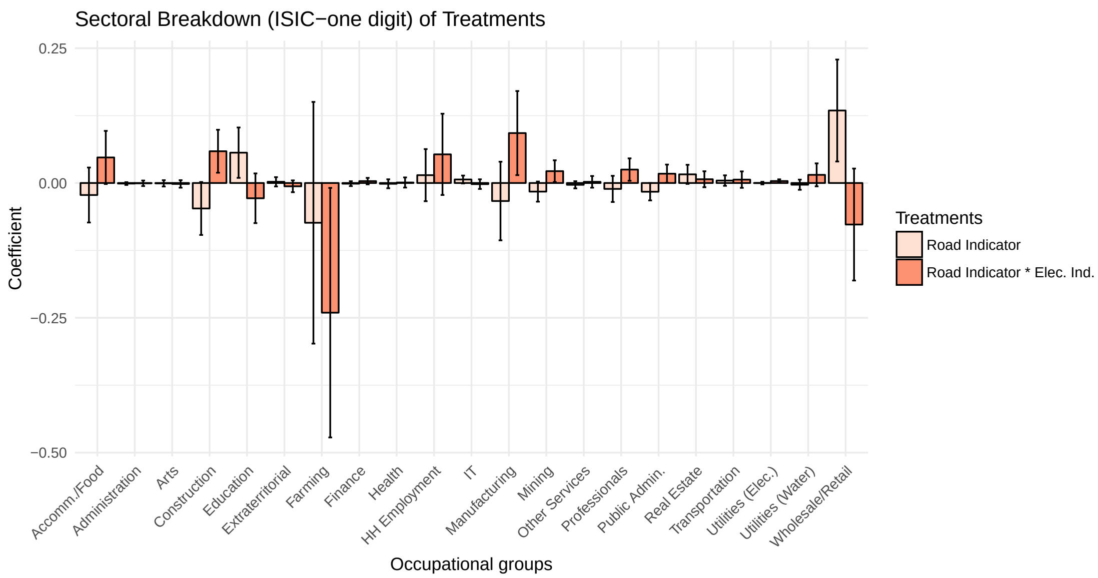


Figure A15: Quintile Treatment Effects by Age

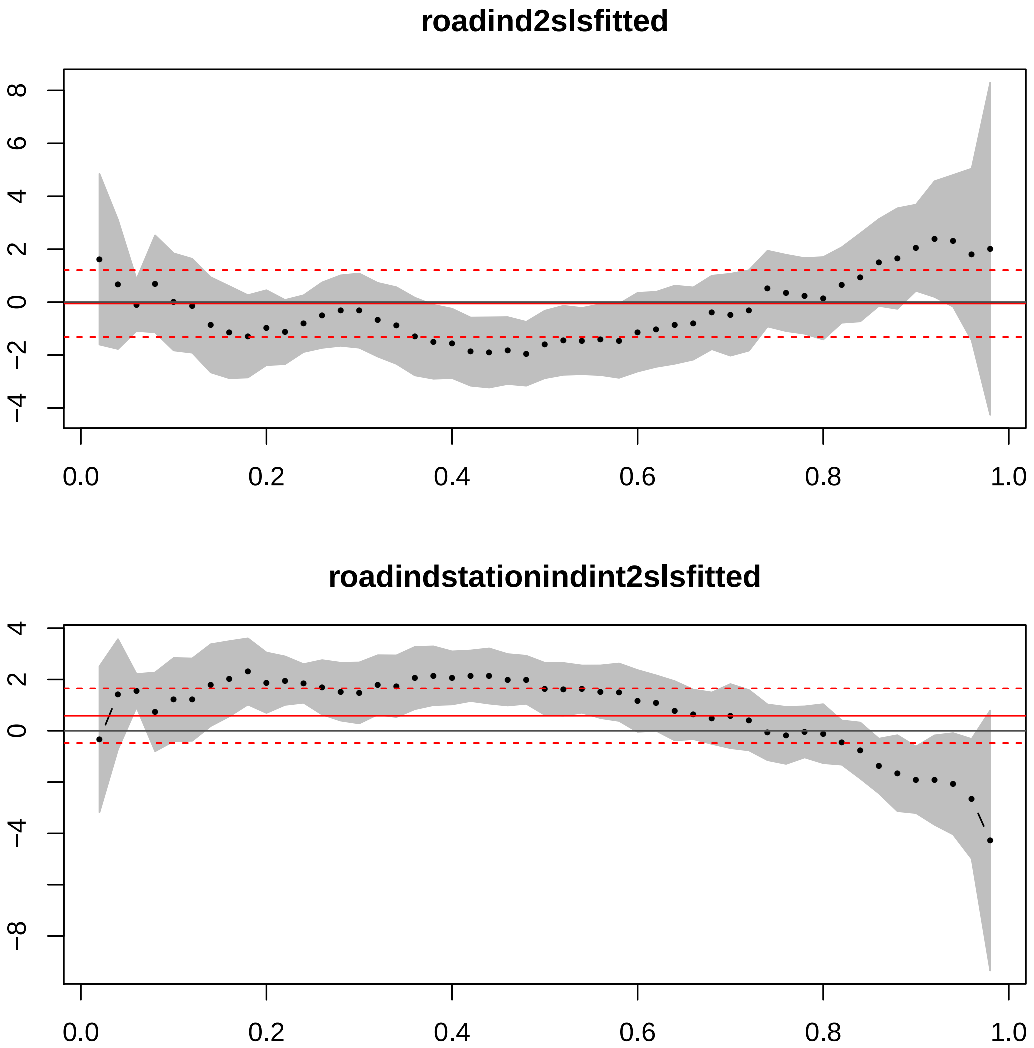


Figure A16: Sectoral Breakdown of Treatments by Gender

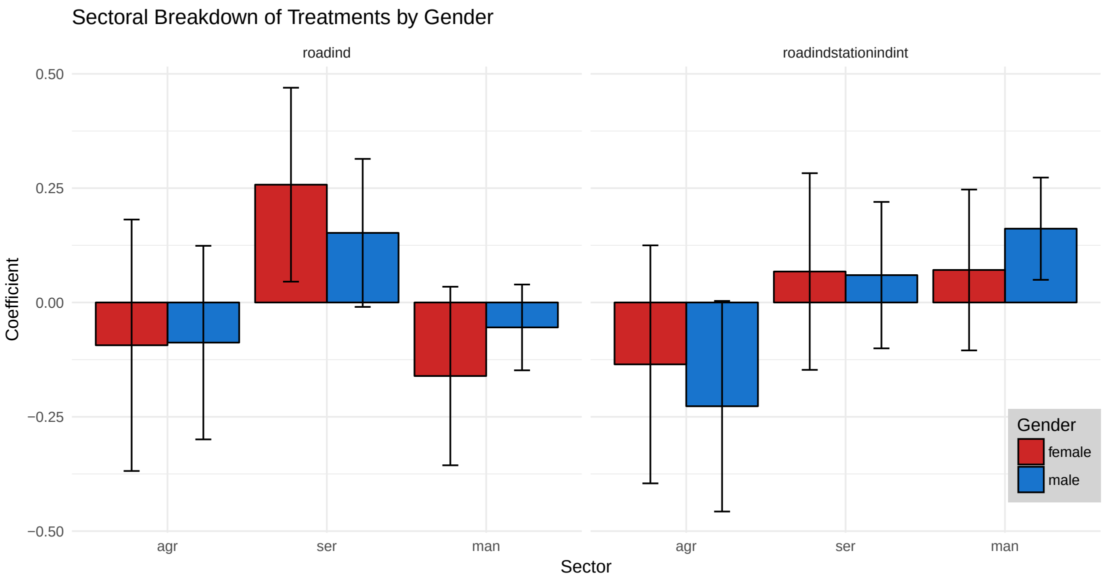
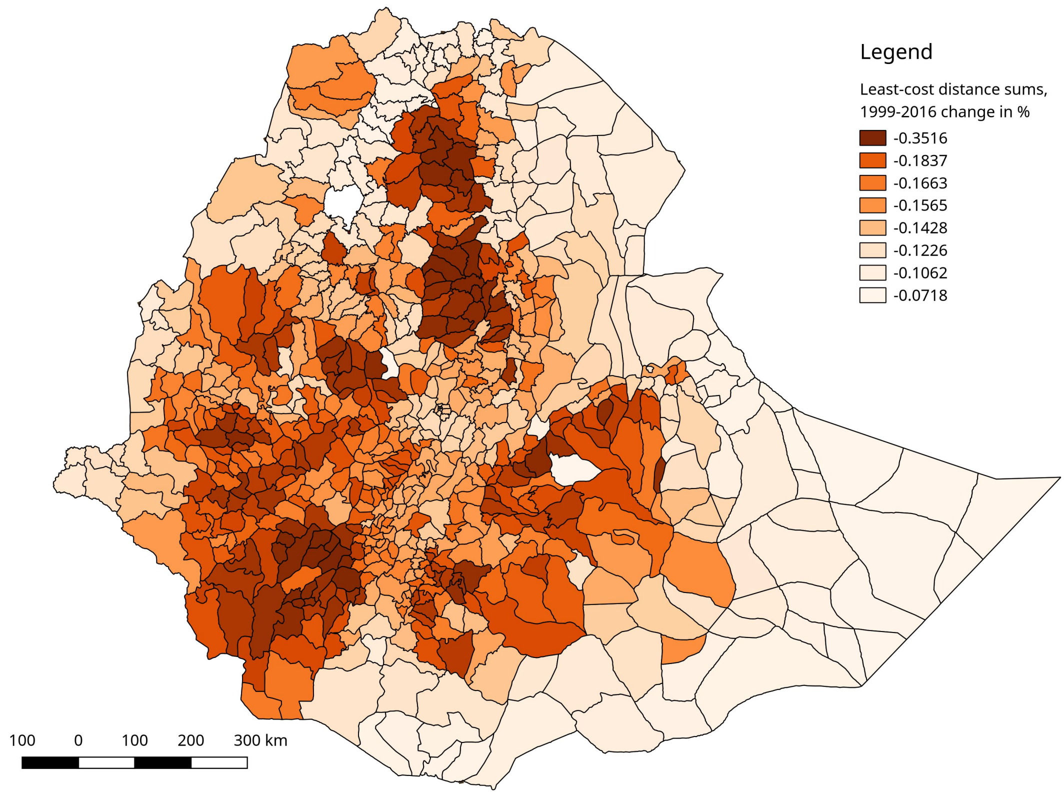


Figure A17: Relative Dijkstra Algorithm Least-cost Distance Changes across Districts, Single Long Difference (1999-2016)



Appendix C

Additional Tables

Table A1: New Sampling Conditional on Population: Correlation with Treatments by Year (DHS-R)

	Dependent variable:			
	District Newly Sampled Indicator			
	(1)	(2)	(3)	(4)
Road Ind.	0.391***	−0.042	−0.108	−0.018
	(0.131)	(0.064)	(0.067)	(0.073)
Elec. Ind.	0.659***	−0.028	−0.003	−0.014
	(0.212)	(0.102)	(0.051)	(0.022)
Log Pop.	0.010	0.154***	−0.010	−0.003
	(0.085)	(0.041)	(0.022)	(0.013)
Year	2000	2005	2011	2016
Controls	✓	✓	✓	✓
New Sampled	294	114	80	47
Resampled	0	175	228	273
Unsampled	393	398	379	367
Observations	687	687	687	687
R^2	0.137	0.038	0.025	0.029
Adjusted R^2	0.125	0.025	0.012	0.016
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table A2: OLS: Occupational Change (NLFS), Roads and Electricity (1999-2013)

	Dependent variable:											
	Agriculture				Services				Manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Road Indicator	−0.020	−0.001	−0.053***	−0.033**	0.040***	0.028***	0.046***	0.032***	−0.019**	−0.027***	0.006	0.0003
	(0.014)	(0.014)	(0.015)	(0.015)	(0.009)	(0.010)	(0.010)	(0.011)	(0.008)	(0.008)	(0.008)	(0.008)
Road*Elec Ind.	−0.218***	−0.143***	−0.217***	−0.146***	0.150***	0.099***	0.150***	0.101***	0.068***	0.044***	0.067***	0.046***
	(0.025)	(0.020)	(0.025)	(0.019)	(0.018)	(0.014)	(0.018)	(0.014)	(0.009)	(0.008)	(0.009)	(0.008)
Model	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Year FE			✓	✓			✓	✓			✓	✓
Controls		✓		✓		✓		✓		✓		✓
Observations	1,208	1,208	1,208	1,208	1,208	1,208	1,208	1,208	1,208	1,208	1,208	1,208
R^2	0.148	0.273	0.189	0.309	0.166	0.298	0.178	0.309	0.060	0.123	0.164	0.221
Adjusted R^2	0.147	0.267	0.186	0.302	0.165	0.293	0.175	0.303	0.058	0.117	0.161	0.214

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A3: OLS: Occupational Change (DHS-R), Roads and Electricity (2000-2016)

	Dependent variable:											
	Agriculture				Services				Manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Road Indicator	−0.005	0.008	0.002	0.026	0.033**	0.019	0.021	−0.002	−0.025**	−0.026**	−0.019	−0.021*
	(0.020)	(0.019)	(0.022)	(0.021)	(0.014)	(0.014)	(0.017)	(0.016)	(0.011)	(0.011)	(0.012)	(0.011)
Road*Elec Ind.	−0.202***	−0.092***	−0.204***	−0.091***	0.142***	0.056***	0.144***	0.055***	0.059***	0.035***	0.059***	0.036***
	(0.035)	(0.025)	(0.035)	(0.025)	(0.025)	(0.018)	(0.025)	(0.018)	(0.013)	(0.012)	(0.012)	(0.012)
Model	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Year FE			✓	✓			✓	✓			✓	✓
Controls		✓		✓		✓		✓		✓		✓
Observations	1,039	1,039	1,039	1,039	1,039	1,039	1,039	1,039	1,039	1,039	1,039	1,039
R^2	0.097	0.285	0.106	0.298	0.098	0.271	0.105	0.283	0.035	0.107	0.056	0.128
Adjusted R^2	0.095	0.279	0.102	0.289	0.097	0.265	0.101	0.275	0.033	0.100	0.052	0.118

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A4: OLS: Occupational Change (NLFS, tercile zone capital distance), Roads and Electricity (1999-2013)

	1st terc. zone capital dist.			2nd terc. zone capital dist.			3rd terc. zone capital dist.		
	Agr.	Ser.	Man.	Agr.	Ser.	Man.	Agr.	Ser.	Man.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Road Indicator	-0.031	0.021	0.007	-0.013	0.022	-0.010	-0.057^{**}	0.043^{**}	0.015
	(0.036)	(0.026)	(0.016)	(0.023)	(0.017)	(0.014)	(0.024)	(0.019)	(0.011)
Road*Elec Ind.	-0.131^{***}	0.087^{***}	0.043^{***}	-0.076^{***}	0.063^{***}	0.014	-0.028	0.007	0.022
	(0.034)	(0.023)	(0.014)	(0.029)	(0.024)	(0.016)	(0.030)	(0.021)	(0.015)
Model	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	371	371	371	393	393	393	399	399	399
R^2	0.249	0.230	0.250	0.218	0.217	0.197	0.141	0.137	0.198
Adjusted R^2	0.226	0.206	0.227	0.196	0.194	0.174	0.117	0.112	0.175

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

1st to 2nd tercile at log 3.428 (≈30.8km); 2nd to 3rd tercile at log 3.903 (≈49.55km)

Table A5: OLS: Occupational Change (DHS-R, tercile zone capital distance), Roads and Electricity (1999-2013)

	1st terc. zone capital dist.			2nd terc. zone capital dist.			3rd terc. zone capital dist.		
	Agr.	Ser.	Man.	Agr.	Ser.	Man.	Agr.	Ser.	Man.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Road Indicator	−0.002	−0.006	0.010	0.093***	−0.052*	−0.036**	−0.033	0.042	−0.006
	(0.051)	(0.037)	(0.028)	(0.036)	(0.028)	(0.018)	(0.033)	(0.026)	(0.017)
Road*Elec Ind.	−0.141***	0.089***	0.054***	0.020	−0.008	−0.013	0.027	−0.062*	0.029
	(0.042)	(0.031)	(0.019)	(0.030)	(0.022)	(0.015)	(0.037)	(0.034)	(0.018)
Model	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	342	342	342	343	343	343	354	354	354
R^2	0.380	0.372	0.175	0.209	0.173	0.120	0.115	0.119	0.127
Adjusted R^2	0.357	0.349	0.145	0.181	0.143	0.088	0.084	0.089	0.096

Note: *p<0.1; **p<0.05; ***p<0.01

1st to 2nd tercile at log 3.428 (≈30.8km); 2nd to 3rd tercile at log 3.903 (≈49.55km)

Table A6: Occup. Change (DHS-R), Roads and Elec. (2000-2016)

	Dependent variable:		
	Agriculture	Services	Manufacturing
	(1)	(2)	(3)
Road Indicator	-0.192	0.228^{**}	-0.035
	(0.128)	(0.105)	(0.065)
Road*Elec Ind.	-0.223	0.014	0.216^{***}
	(0.143)	(0.105)	(0.078)
Model	2SLS	2SLS	2SLS
Year FE	✓	✓	✓
Controls	✓	✓	✓
Cragg-Donald F.		22.328	
Windmeijer cond. F.	19.613	16.27	
p-val $\beta_1 + \beta_2 = 0$	0.0051	0.0291	0.0201
Observations	1,039	1,039	1,039
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A7: Occup. Change (NLFS, excl. Somali), Roads and Elec. (1999-2013)

	Dependent variable:		
	Agriculture	Services	Manufacturing
	(1)	(2)	(3)
Road Indicator	-0.071	0.196^{**}	-0.119^{**}
	(0.110)	(0.088)	(0.060)
Road*Elec Ind.	-0.202^{*}	0.060	0.141^{**}
	(0.114)	(0.089)	(0.061)
Model	2SLS	2SLS	2SLS
Year FE	✓	✓	✓
Controls	✓	✓	✓
Cragg-Donald F.		10.327	
Windmeijer cond. F.	17.545	13.208	
p-val $\beta_1 + \beta_2 = 0$	0.007	9e-04	0.6813
Observations	1,188	1,188	1,188
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A8: Construction Robustness: Occup. Change (NLFS-ISIC, excl. Somali), Roads and Elec. (1999-2013)

	Dependent variable:				
	Constr.	Ser. [isic]	Ser. incl. Constr.	Man. [isic]	Man. excl. Constr.
	(1)	(2)	(3)	(4)	(5)
Road Indicator	-0.023	0.216^{**}	0.193^{*}	-0.125^{**}	-0.102^{**}
	(0.018)	(0.099)	(0.100)	(0.062)	(0.051)
Road*Elec Ind.	0.055^{*}	-0.041	0.014	0.232^{***}	0.177^{**}
	(0.028)	(0.126)	(0.128)	(0.089)	(0.072)
Model	2SLS	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
Cragg-Donald F.		6.052			
Windmeijer cond. F.	15.261	10.377			
p-val $\beta_1 + \beta_2 = 0$	0.1164	0.0483	0.0206	0.1181	0.1664
Observations	1,188	1,188	1,188	1,188	1,188
Note:					*p<0.1; **p<0.05; ***p<0.01

Table A9: Instrument Validity: Initial MA proxy on IVs/Ts (1999-2013)

	Dependent variable:			
	Initial MA proxy		MA proxy	
	(1)	(2)	(3)	(4)
Road IV	−0.068		0.008***	
	(0.063)		(0.002)	
Road IV*Elec IV	0.015		−0.013***	
	(0.076)		(0.003)	
Road		−0.009		−0.002
		(0.055)		(0.002)
Road*Elec		0.021		0.003
		(0.080)		(0.003)
Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Observations	1,208	1,208	1,208	1,208
R^2	0.016	0.015	0.998	0.998
Adjusted R^2	0.007	0.006	0.998	0.998
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A10: Occup. Change (NLFS, gender split), Roads and Elec. (1999-2013)

	Female			Male		
	Agr.	Ser.	Man.	Agr.	Ser.	Man.
	(1)	(2)	(3)	(4)	(5)	(6)
Road Indicator	-0.094	0.258^{**}	-0.161	-0.088	0.152^{*}	-0.055
	(0.140)	(0.108)	(0.100)	(0.108)	(0.083)	(0.048)
Road*Elec Ind.	-0.135	0.068	0.071	-0.227^{*}	0.060	0.161^{***}
	(0.133)	(0.110)	(0.090)	(0.117)	(0.082)	(0.057)
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
Cragg-Donald F.		9.993				
Windmeijer cond. F.	16.747	13.143				
p-val $\beta_1 + \beta_2 = 0$	0.0426	5e-04	0.2234	0.0025	0.0037	0.0278
Observations	1,208	1,208	1,208	1,208	1,208	1,208

Note: *p<0.1; **p<0.05; ***p<0.01

Table A11: Employment Relations Break-Down (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:							
	Empl.-Gov.	Empl.-NGO	Empl.-Priv.	Empl.-Para.	Other Empl.	Self-Empl.	Unpaid	Employer
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Road Indicator	0.029	0.016	-0.074^{*}	0.069^{***}	-0.012	0.041	-0.061	-0.003
	(0.026)	(0.022)	(0.043)	(0.024)	(0.009)	(0.053)	(0.069)	(0.007)
Road*Elec Ind.	0.049	-0.026	0.181^{***}	0.010	0.026^{**}	-0.090	-0.161^{*}	0.011
	(0.037)	(0.024)	(0.063)	(0.025)	(0.013)	(0.077)	(0.097)	(0.008)
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓
Cragg-Donald F.		6.272						
Windmeijer cond. F.	16.334	11.003						
p-val $\beta_1 + \beta_2 = 0$	4e-04	5e-04						
Observations	1,208	1,208	1,208	1,208	1,208	1,208	1,208	1,208

Note: *p<0.1; **p<0.05; ***p<0.01

Table A12: Demographics (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:			
	Age	Never Married	Married	Divorced
	(1)	(2)	(3)	(4)
Road Indicator	0.223	0.056	-0.027	-0.033
	(0.984)	(0.043)	(0.043)	(0.032)
Road*Elec Ind.	2.162*	-0.053	-0.052	0.087***
	(1.272)	(0.048)	(0.050)	(0.033)
Model	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Cragg-Donald F.		9.993		
Windmeijer cond. F.	16.747	13.143		
p-val $\beta_1 + \beta_2 = 0$	0.0317	0.9411	0.0473	0.0927
Observations	1,208	1,208	1,208	1,208
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table A13: LFP (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:					
	L-Sampled	L-Force	L-Act. Force	LFP rate	LFP-S rate	Notwork
	(1)	(2)	(3)	(4)	(5)	(6)
Road Indicator	20.138	3.890	4.447	-0.002	0.050	-0.006
	(15.343)	(9.256)	(8.689)	(0.029)	(0.050)	(0.006)
Road*Elec Ind.	-37.381^{**}	-11.040	-14.737	0.033	-0.080	0.001
	(18.630)	(9.723)	(9.299)	(0.041)	(0.059)	(0.006)
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
Cragg-Donald F.		9.993				
Windmeijer cond. F.	16.747	13.143				
p-val $\beta_1 + \beta_2 = 0$	0.3146	0.4075	0.2215	0.3441	0.5304	0.3218
Observations	1,208	1,208	1,208	1,208	1,208	1,208

Note: *p<0.1; **p<0.05; ***p<0.01

Table A14: Education (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:	
	Read/Write	Edu. (Years)
	(1)	(2)
Road Indicator	0.198**	1.083
	(0.094)	(0.712)
Road*Elec Ind.	−0.119	−0.182
	(0.114)	(0.956)
Model	2SLS	2SLS
Year FE	✓	✓
Controls	✓	✓
Cragg-Donald F.		9.993
Windmeijer cond. F.	16.747	13.143
p-val $\beta_1 + \beta_2 = 0$	0.4228	0.2739
Observations	1,208	1,208
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table A15: Education R/W (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:					
	R/W	R/W Teens	R/W Young Ad.	R/W Old Ad.	R/W Mig.6	R/W Nonmig.6
	(1)	(2)	(3)	(4)	(5)	(6)
Road Indicator	0.198**	0.265**	0.253**	0.071	0.342*	0.188**
	(0.094)	(0.123)	(0.118)	(0.073)	(0.184)	(0.091)
Road*Elec Ind.	−0.119	−0.136	−0.205	0.045	−0.393*	−0.114
	(0.114)	(0.137)	(0.145)	(0.093)	(0.211)	(0.112)
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
Cragg-Donald F.		9.993				
Windmeijer cond. F.	16.747	13.143				
p-val $\beta_1 + \beta_2 = 0$	0.4228	0.7077	0.1409	0.7555	0.4486	
Observations	1,208	1,208	1,208	1,208	1,112	1,208

Note: *p<0.1; **p<0.05; ***p<0.01

Table A16: Education Years (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:					
	Edu. Years	Yrs. Teens	Yrs. Young Ad.	Yrs. Old Ad.	Yrs. Mig.6	Yrs. Nonmig.6
	(1)	(2)	(3)	(4)	(5)	(6)
Road Indicator	1.083	1.378**	1.572*	0.442	3.523*	1.023
	(0.712)	(0.692)	(0.937)	(0.618)	(1.885)	(0.665)
Road*Elec Ind.	−0.182	−0.950	−0.540	0.618	−3.979*	−0.176
	(0.956)	(0.859)	(1.229)	(0.872)	(2.219)	(0.931)
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
Cragg-Donald F.		9.993				
Windmeijer cond. F.	16.747	13.143				
p-val $\beta_1 + \beta_2 = 0$	0.2739	0.3422	0.1412	0.784	0.297	
Observations	1,208	1,208	1,208	1,208	1,112	1,208

Note: *p<0.1; **p<0.05; ***p<0.01

Table A17: Edu. R/W-Mig. (NLFS), Roads and Elec. (1999-2013)

	Dependent variable:						
	R/W	R/W T. NM	R/W Yg. Ad. NM	R/W Old Ad. NM	R/W T M	R/W Yg. Ad. M	R/W Old Ad. M
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Road Indicator	0.198**	0.304**	0.335**	0.133*	0.247	0.168	0.272**
	(0.094)	(0.135)	(0.137)	(0.080)	(0.185)	(0.136)	(0.119)
Road*Elec Ind.	−0.119	−0.126	−0.307*	−0.069	−0.272	0.032	0.018
	(0.114)	(0.146)	(0.169)	(0.107)	(0.197)	(0.163)	(0.147)
Model	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Year FE	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓
Cragg-Donald F.		9.993					
Windmeijer cond. F.	16.747	13.143					
p-val $\beta_1 + \beta_2 = 0$	0.4228	0.1795	0.8499	0.4739	0.878	0.1427	0.019
Observations	1,208	1,205	1,203	1,196	971	1,137	1,149

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$