Can Big Push Infrastructure Unlock Development?

Evidence from Ethiopia

Niclas Moneke*

LSE

Job Market Paper

November 2019 [click here for latest version]

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%).

JEL classification: F15, J24, L16, O13, O14, O18, Q41, R1

*I am indebted to Oriana Bandiera, Gharad Bryan, Robin Burgess and Daniel Sturm for their continuous support throughout this project. I thank Karun Adusumilli, Jan David Bakker, Tim Besley, Francesco Caselli, Simon Franklin, Doug Gollin, Michael Greenstone, Vernon Henderson, Felix König, David Lagakos, Rocco Machiavello, Ted Miguel, Bart Minten, David Nagy, Shan Aman Rana, Imran Rasul, Claudio Schilter, Fabian Waldinger, Torsten Figueiredo Walter and participants at Bonn, CEPR-ILO Geneva, CURE, InsTED, OxDEV, ONS, SERC, SMYE Brussels and UEA Philadelphia for helpful comments. Financial support from the IGC is gratefully acknowledged. Contact: n.moneke@lse.ac.uk, https://niclasmoneke.com, Department of Economics, LSE, Houghton Street, WC2A 2AE, London, UK.

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 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]

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.

Second, in order to test these predictions empirically, I provide new, geo-identified

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

[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.

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 inter-regional trade (Eaton & Kortum, 2002), to capture a rich geography of heterogeneous locations; general equilibrium implications of road investments 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

[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).

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 exploit locations' proximity to straight transmission lines that connect newly opening hydropower dams with Addis Abeba.[9,10] Intuitively, the electrification instrumental variable exploits 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] The temporal variation in my roads IV arises from realising that, firstly, 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' 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

[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.

[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.

[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.

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 & 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

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

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 of employment in such locations.

For example, the strong spatial heterogeneity in response to infrastructure shocks predicted by the model are directly confirmed in the reduced-form: those districts closest 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 paper is organised as follows: Section 2 develops a simple spatial general equilibrium model. Section 3 introduces the empirical context in Ethiopia and describes the data. I then present my reduced-form empirical strategy (Section 4) and the reduced-form results (Section 5). Section 6 details the structural estimation strategy, provides welfare results and studies policy counterfactuals. Section 7 concludes.

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, we 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 interregional trade.[15] 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.[16] 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 traditional sector final goods, C^T, modern 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:

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

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

[16] 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.

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

$$C_n = \left[\psi^T \left(C_n^T\right)^\rho + \psi^M \left(C_n^M\right)^\rho + \psi^S \left(C_n^S\right)^\rho\right]^{1/\rho} \tag{2}$$

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 (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 $j \in [0, 1]$:

$$C_n^T = \left[\sum_{i \in N} \int_0^1 \left(c_{ni}^T(j)\right)^\nu dj\right]^{\frac{1}{\nu}} \tag{3}$$

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 (4) provides the classic Dixit-Stiglitz price index over traditional sector goods, with the manufacturing sector's Dixit-Stiglitz price index, P_n^M, following an equivalent formulation:

$$P_n^T = \left[\sum_{i \in N} \int_0^1 \left(p_{ni}^T(j)\right)^{1-\sigma} dj\right]^{\frac{1}{1-\sigma}} \tag{4}$$

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.

$$Y_n^T = z^T \left(\frac{L_n^T}{\mu^T}\right)^{\mu^T} \left(\frac{h_n^T}{1-\mu^T}\right)^{1-\mu^T} \tag{5}$$

$$Y_n^M = z^M \left(\frac{L_n^M}{\mu^M}\right)^{\mu^M} \left(\frac{h_n^M}{1-\mu^M}\right)^{1-\mu^M} \tag{6}$$

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

$$F_n^T(z^T) = e^{\left(-A_n^T z^T\right)^{-\theta}} \tag{7}$$

$$F_n^M(z^M) = e^{\left(-A_n^M z^M\right)^{-\theta}} \tag{8}$$

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 across all varieties, since, for example, larger values of A_n^T decrease $F_n^T(z^T)$ and thus increase the probability of higher productivity draws, z^T, for all traditional sector varieties 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 our empirical application focuses on sector-location specific average productivity shocks, we assume the shape parameter, θ, to be the same across sectors and locations.

Trade in both sectors' final goods is costly and we assume trade costs 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. We 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. $d_{nn} = 1$. We also assume that trade costs are the same across sectors ($d_{ni}^T = d_{ni}^M$), are symmetric ($d_{ni} = d_{in}$) and a triangle inequality to hold 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, $p_{ni}^T(j)$, must equal marginal costs, weighted by factor shares, inverse productivity and trade costs:

$$p_{ni}^T(j) = \frac{w_i^{\mu^T} r_i^{1-\mu^T} d_{ni}}{z_i^T(j)} \tag{9}$$

Similarly standard, relative factor demands have to equal factor share-weighted, inverse factor prices[17]:

$$\frac{h_i^T}{L_i^T} = \frac{(1-\mu^T)}{\mu^T}\frac{w_i}{r_i} \tag{10}$$

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

$$p_{ni}^T(j) = min\{p_i^T(j); i \in N\} \tag{11}$$

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 expenditure shares for each destination location n on traditional sector (and equivalently modern sector) final goods produced in origin i:

$$\pi_{ni}^T = \frac{A_i^T\left(w_i^{\mu^T} r_i^{1-\mu^T} d_{ni}\right)^{-\theta}}{\sum_{k\in N} A_k^T\left(w_k^{\mu^T} r_k^{1-\mu^T} d_{nk}\right)^{-\theta}} \tag{12}$$

where 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 homogenous 'services good':

$$Y_n^S = A_n^S\left(\frac{L_n^S}{\mu^S}\right)^{\mu^S}\left(\frac{h_n^S}{1-\mu^S}\right)^{1-\mu^S} \tag{13}$$

Throughout, I assume that agriculture is the most, while services the least land-intensive sector, $\mu^T < \mu^M < \mu^S$. Without trade in services, the non-tradeable services

[17]Since both factors have to be overused in production to account for the iceberg-style loss in produced output quantity during transit from production location i to consumption location n, transport costs d_{ni} cancel out from relative factor demands.

good's price equals marginal cost:

$$P_n^S = \frac{w_n^{\mu^S} r_n^{1-\mu^S}}{A_n^S} \tag{14}$$

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

$$\xi_n^K = \frac{\left(\psi^K\right)^\kappa \left(P_n^K\right)^{1-\kappa}}{\left(\psi^M\right)^\kappa \left(P_n^M\right)^{1-\kappa} + \left(\psi^T\right)^\kappa \left(P_n^T\right)^{1-\kappa} + \left(\psi^S\right)^\kappa \left(P_n^S\right)^{1-\kappa}}, K \in \{T, M, S\} \tag{15}$$

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 (16) presents the simplified T-sector price index. An equivalent formulation holds for the M-sector.

$$P_n^T = \gamma \left[\sum_{k \in N} A_k^T \left(w_k^{\mu^T} r_k^{1-\mu^T} d_{nk}\right)^{-\theta}\right]^{-1/\theta} = \gamma \left(\Phi_n^T\right)^{-1/\theta} \tag{16}$$

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

To arrive at an equilibrium below, I now 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, by M-sector firms and by T-sector firms. We 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 local government that consumes its income from land on goods and land consumption itself.

The land market clearing condition can be stated as follows:

$$
\begin{aligned}
r_n H_n = (1-\alpha)\left[w_n L_n + r_n H_n\right] \\
&+ \sum_{k \in N} \pi_{kn}^{T} \xi_k^{T} \left(1-\mu^{T}\right) \alpha \left[w_k L_k + r_k H_k\right] \\
&+ \sum_{k \in N} \pi_{kn}^{M} \xi_k^{M} \left(1-\mu^{M}\right) \alpha \left[w_k L_k + r_k H_k\right] \\
&+ \pi_{nn}^{S} \xi_n^{S} \left(1-\mu^{S}\right) \alpha \left[w_n L_n + r_n H_n\right]
\end{aligned}
\tag{17}
$$

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:

$$
\begin{aligned}
w_n L_n = & \sum_{k \in N} \pi_{kn}^{T} \xi_k^{T} \mu^{T} \alpha \left[w_k L_k + r_k H_k\right] \\
&+ \sum_{k \in N} \pi_{kn}^{M} \xi_k^{M} \mu^{M} \alpha \left[w_k L_k + r_k H_k\right] \\
&+ \pi_{nn}^{S} \xi_n^{S} \mu^{S} \alpha \left[w_n L_n + r_n H_n\right]
\end{aligned}
\tag{18}
$$

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:

$$
V_n = \bar{V} = \frac{\alpha^{\alpha}(1-\alpha)^{(1-\alpha)} \eta_n w_n}{\left[P_n\right]^{\alpha/(1-\kappa)} r_n^{(1-\alpha)}} \tag{19}
$$

where $P_n = \left(\psi^{M}\right)^{\kappa} \left(P_n^{M}\right)^{1-\kappa} + \left(\psi^{T}\right)^{\kappa} \left(P_n^{T}\right)^{1-\kappa} + \left(\psi^{S}\right)^{\kappa} \left(P_n^{S}\right)^{1-\kappa}$ and, after replacing the sectoral price indices in the denominator with equations (16) and the equivalent M-sector formulation, the labour mobility condition 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. (17)], labour market clearing [eq. (18)] and the labour mobility condition [eq. (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 our 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:

$$w_n L_n^T = \sum_{k \in N} \pi_{kn}^T \xi_k^T \mu^T \alpha \left[w_k L_k + r_k H_k \right] \tag{20}$$

$$w_n L_n^M = \sum_{k \in N} \pi_{kn}^M \xi_k^M \mu^M \alpha \left[w_k L_k + r_k H_k \right] \tag{21}$$

$$w_n L_n^S = \pi_{nn}^S \xi_n^S \mu^S \alpha \left[w_n L_n + r_n H_n \right] \tag{22}$$

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) methodology, we are unable to invert the model to solve for unobserved productivities (and amenities) since rental rates are generally not observable in Ethiopia given the overwhelmingly nationalised status of land ownership during our study period until 2016. Therefore, we cannot simply invert the general equilibrium system to determine productivities, but 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 that features occasionally binding constraints when sectoral employment reaches zero for any one of the two sectors in a location, 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. Firstly, 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). 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 our spatial general equilibrium model: 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 crucial. Since we aim to estimate empirically relevant effects of shocks which manifest themselves in sectoral heterogeneity across locations, we opt for the empirically more realistic assumption of heterogeneous sectoral productivities within locations, that is $A_n^M \neq A_n^T$ everywhere.

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 to increase local manufacturing and services sector productivities in electrified regions.

Since I am interested in structural transformation as a relevant proxy of economic development, the objects of interest are the partial derivatives of any sectoral employment share, 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 our 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 our empirical analysis below: Figure (10) 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 (11) depicts changes in relative manufacturing employment shares as a result of a simulated combined transport cost and electrification shock.

As highlighted in Figures (10) and (11), 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 (10) 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 (11) (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.[18]

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 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.

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

3 Empirical Context and Data

3.1 Why Ethiopia? An ideal study setting

We study the effects of infrastructure complementarities and their effects on structural transformation in the context of Ethiopia over the last two decades. Ethiopia provides an ideal study setting for several reasons: firstly, the country experienced large-scale investments in two separate kinds of infrastructure, namely all-weather roads and electricity. The all-weather road network expanded roughly fourfold between the late 1990s and today, from approximately 16,000km to 70,000km. Figure 2 provides a graphical account of this expansion. Our focus on all-weather roads follows the general understanding in the literature that trade and market access rely on year-round accessibility (ideally by truck) of a given location.
Over the same time period, although with a slight lag, the electricity network doubled in its extent from 95 to 191 major electricity substations. Figure 3 displays this expansion of the electricity network during our sample period. Electric substations are crucial for electrification since they step down the voltage from high-voltage, long-distance overland transmission lines to local, low-voltage distribution-networks that connect individual firms, households and other end-users to the electric grid.

Secondly, the almost complete lack of direct infrastructure substitutes in Ethiopia implies that the all-weather road and electricity network expansions we track do in fact capture genuine extensive margin effects of access to infrastructure. In particular, Ethiopia is a landlocked country without major navigable rivers or canals. During our study period, the single existing railway line (to neighbouring Djibouti and its port) was still out of order.[19] A single new railway construction project was not started before 2015.[20]
With respect to access to energy and substitutes for grid electricity, only a handful of minor, isolated diesel generators originating from the 1960s operated in selected major cities. All of these major cities were electrified before our study period and, thus, do not feature as compliers in the instrumental variables strategy below. Self-generated energy

[19]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.

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

from off-grid solar home systems generated approximately one megawatt of capacity midway through our 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 5-10W each) were purchased by the Ethiopian government for decentralised installation by 2013.[21] Thus, due to both the low penetration and the low voltage and performance of the existing solar home systems in Ethiopia during our 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.

Thirdly, Ethiopia experienced massive structural transformation out of agriculture during our study period. We can confirm this anecdotally reported overall trend using at least two distinct sources of occupational choice data that exist in the case of Ethiopia – both with reasonable spatial coverage, large survey sample sizes and mostly overlapping in time (see Section 3). 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 our study period of interest from the late 1990s to the very recent past.

Based on the above data, Figure 4 presents micro-founded evidence on macroeconomic structural transformation patterns out of agriculture 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 (industry) 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 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), it is not obvious why

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

one should expect positive income and welfare effects of such sectoral shifts, neither at the individual level, nor in the aggregate.
However, as shown in Figure 5, our study period displays an almost exploding time series of GDP per capita and a dramatic reduction in headcount poverty, using either national or international measures. Hence, even from a purely empirical perspective, devoid of theoretical motivation, a study of large-scale infrastructure investments in relation to structural transformation appears *à priori* interesting.[22]

Finally, with a population of more than a hundred million people, an area roughly the size of France and the second largest developing country economy in Sub-Saharan Africa, the study of Ethiopian structural transformation appears of interest in its own right, with potential external validity for other developing countries.

3.2 Data

We provide novel, previously undisclosed information on the electricity grid and the road network expansions in Ethiopia as the foundation of our analysis of infrastructure complementarities since the 1990s.
Resulting from a close collaboration with Ethiopian Electric Power (EEP)[23], the state utility charged with electricity generation and transmission, we obtained confidential information on the exact location, capacity, equipment and commissioning time of each one of the electric grid's substations. 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"), we 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

[22]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 potentially moot: one may 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 natural endowments in terms of agricultural productivity, nor with radial distance to the economic centre.

[23]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 (EEP) and a distribution utility, Ethiopian Electric Utility (EEU).

equipment (e.g. reactors and capacitators) associated with each line. For welfare calculations, we 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, we also collected locations, capacity, operational status and commissioning time information on all power plant to track generation. The Ethiopian electricity supply is mostly provided by hydropower from eight major dams, as well as at least three wind farms, one geothermal power plant and by-generation from at least three sugar refineries. Recent dam openings after 2016 are currently ignored in our analysis due to the lack of outcome variables spanning this very recent past (see below).

Although the opening of a substation certainly implies a large decrease in the potential cost of energy in a given location, it does, however, not perfectly capture distribution-level connections at the town-neighbourhood or village-level. Therefore, we also obtained 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 town-level electrification status. Although originally lacking exact information on the timing of distribution network expansion, we obtained complementary records on town- and village-level electrification status combined with the year of electrification. This comprehensive distribution network coverage and expansion dataset is not yet exploited in our district-level analysis (see Section 4), although it is current work in progress to refine our analysis.

We obtained information on the expansion of all-weather roads mostly from the Ethiopian Roads Authority (ERA). In particular, we employ several historical and present maps and GIS data from various, partially undisclosed records. For the years 2006, 2012 and 2016, we 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.[24] Earlier maps were supposedly based on partial road surveys and/or records of road construction projects. However, we cannot verify this claim given the lack of centrally recorded road construction documentation at the project level.

[24]The 2016 ERA road survey also contains estimates of the original year of each road's construction, which we use for cross-validation of earlier surveys and maps.

In addition, we use various other sources for cross-validation and to obtain better visibility on the pre-sample road network: we use GIS data from OpenStreetMap for the year 2014 to cross-verify the earlier and later ERA records. We 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 our first cross-section in the sample and the 2009 map for cross-validation of ERA records. Furthermore, we 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 we employ in our main analysis.

With respect to the outcome variables of interest, we are first and foremostly interested in structural transformation, which we interpret in line with the literature (Herrendorf et al., 2014) as sectoral employment. Thus, we require information on relative employment, which we 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, we use respondents' answer to questions about their "current occupation", which we then group into sectors three sectors, agriculture, manufacturing/industry 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 our outcome variables, we 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 our 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.
With respect to the geo-identification of enumeration areas (and, thus, households), two qualifications are due: firstly, the enumeration area locations of NLFS EAs are provided in codified form, which may not always be 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. Secondly, 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.[25] Therefore, although we have exact geo-identified information on infrastructure placement, we are constrained by the available outcome variable data with respect to the highest spatial resolution our analysis can support.

4 Empirical Strategy

A natural starting point to estimate the effects of a treatment, such as access to an all-weather road, on an outcome of interest, such as sectoral employment, would be to compare treated locations to untreated ones over time, i.e. a differences-in-differences or fixed effects identification strategy. However, in the particular context of infrastructure investments, which are usually extremely costly and long-term in nature, both underlying identifying assumptions of parallel trends between treatment and control locations in the absence of the treatment, and the stable unit treatment value assumption (SUTVA) are likely to be violated.

The endogeneity concerns in a district-level regression of infrastructure access on economic outcomes, therefore, loom large: first of all, districts should be expected to be targeted or selected to receive access or not, and connection timing of to-be-connected locations should follow some form of prioritisation. For example, one could think of districts with the largest potential for economic growth to come first, or those lagging furthest behind to obtain priority. Omitted variable bias due to other factors affecting both the infrastructure expansion treatment and the outcome of structural transformation (such as natural resource windfalls, global economic cycles, capital flows, etc.) are entirely thinkable in the Ethiopian context. Likewise, reverse causality in the form of sectoral shifts causing greater demand for infrastructure investments can also not be ruled out ex ante. Finally, measurement error in our right-hand side variable (e.g. due to potentially inaccurate timing information of electric grid expansion) and our left-hand

[25] In principle, these displacements do, 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, several displacement errors were corrected manually (see Data Appendix).

side variables (e.g. due to missing information on primary and secondary occupations for the same individual, or the intensive margin of labour supply) are not unlikely.

In the particular case of Ethiopia, for example, we know that the government formulated an explicit policy to connect all of the more than 650 district capitals with an all-weather road by 2020 – an objective that was successfully achieved by 2016 already. Hence, in our analysis of road network investments, a key endogeneity concern is the timing of a district's connection (in contrast to the usual issue of endogenous district selection into treatment and control, since all districts obtain road access treatment eventually). Appendix Figure A9 confirms that more densely populated districts were in fact connected to an all-weather road earlier than more rural, sparsely populated districts, pointing towards the hypothesised endogenous connection timing.
For electric grid access, however, the extremely high cost involved and given existing engineering guidelines that the main cost driver to be minimised was primarily the length of transmission lines, spatial targeting of electricity infrastructure is an obvious feature of electrification. Engineers involved in the Ethiopian grid expansion have also stated privately that to minimise cost, only politically demanded locations would obtain a transmission line connection (and substation), unless they accidentally lie on a straight line between supply (e.g. a hydropower plant) and demand (e.g. the major load centre(s)).

Therefore, 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[26], we resort to employing an instrumental variables identification strategy.
For electricity access, our instrumental variable (IV) is founded on the general idea 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. We thus apply 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. Thus, from the year of dam opening onwards, all districts lying along the straight line connecting

[26] Unlike, for example, Asher and Novosad's (2016) dichotomous targeting of rural roads based on Indian villages' population size above some idiosyncratic threshold.

the dam to Addis will be considered hypothetically electrified.

With respect to such an IV's identifying assumptions, the validity assumption would read: the hypothetical electrification status of districts along 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. We draw straight line connection corridors of 25km radius for at least eight dams and two large-scale wind farms.

The random assignment assumption of our 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 usually would span several hundred kilometres, are not systematically different than nearby locations off the straight line corridor. 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 does appear to have some merit in the Ethiopian context.

Finally, the exclusion restriction requires that the straight line from a dam to Addis does not affect structural transformation in the years and locations exposed to the (now online) hypothetical transmission line, other than through actual electrification.

In sum, our instrumental variable for electricity is similar in nature to Michaels (2008) and Kassem (2018), who use similar exposure to artificial lines to instrument for infrastructure expansions. Figure 6 provides a graphical representation of our instrument and how differential proximity to straight line corridors (and their opening years) generate spatial and temporal variation in districts' hypothetical electrification status.

Regarding our instrumental variable for the timing of a district's road connection, we construct a hypothetical least-cost network constructed in the following way: starting from the historic Italian colonial road network, which provides a plausibly exogenous baseline all-weather road network cross-section for Ethiopia. We 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 our 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 algorithms with respect to their order), but do not yet provide temporal variation in districts' connection timing. Therefore, we apply 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. We 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. Figure 2 provides a graphical representation of the road IV's variation that we exploit in our two-stage least squares estimation below.

In addition to the Kruskal's algorithm-derived road IV, we also provide an alternative instrumental variable for a district's road connection derived solely from 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 occupiers 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 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 port at Assab 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.

Actual Italian road construction started in 1936 followed to a surprising extent 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 road. Appendix Figure A6 provides a historic picture of the construction efforts during the late 1930s.

On the territory of today's Ethiopia, approximately 3,378km (2,023m) of paved 'highways' were constructed, of which at least 1,970km (1,180m) 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.[27]

[27] 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

For the purposes of our study, we exploit the fact that Ethiopian road construction in the 1990s started reconstruction of its road network from the formerly Italian trunk network and subsequently, during the period of our 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 A7 and A8 provides two exemplary cases of how Ethiopian road construction initiated from (previously reconstructed) Italian colonial roads, connecting nearby settlements almost orthogonally.

We 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), we calculate orthogonal, shortest distances to every district capital, as the crow flies.[28] One should also note that this distance is calculated from the plausibly exogenous straight lines designed by the Italians, not from the actual roads that were constructed (and re-constructed) under these designs. Since road construction in the Federal Democratic Republic of Ethiopia is politically a regional matter, we run the following algorithm separately and simultaneously for each of the eleven regions of Ethiopia.
Given the total length of (straight line) road connections to be built in every region to connect every district in that region, we calculate the annual mileage per region of road construction to achieve this goal of universal district road access by the end of our sample period, i.e. over a seventeen year period (2000-2016).[29] Given this annual mileage goal, we allow each region to build the shortest stretches of (straight line, orthogonal) district connections first until the goal for a given year has been reached. For every subsequent year, we 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.[30] 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,

Afdem range and the Semien Mountains.

[28] We exclude districts which contain arteries, which are considered already treated by the IV.

[29] As confirmed by the maps of the Ethiopian road network in 1999, the reconstruction of the original Italian road network was finished by then. Therefore, we assume that new construction started from the year 2000 onwards.

[30] 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 build-out observed on the ground.

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 closest to the artery district 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, we do not update distance and always have the algorithm build (relatively more unrealistic) connections to the Italian artery irrespective of any districts already connected in between.
In sum, our second roads IV provides temporal and spatial variation in district road access derived from a purely exogenous source, namely straight line distance to Italian straight line arteries. The instrument takes a value of one from the district-year in which a given district got connected onwards.

Our research design builds on two separate instrumental variables for each kind of infrastructure investment we observe in the data: one for the electricity network expansion and one for the road network expansion. Of key interest to us is also a third endogenous variable, that is the interaction of both roads and electricity access.
As shown in Table 2, however, our sample does not feature three genuine treatments, namely a roads treatment, an electricity treatment and a roads combined with electricity treatment. Instead, we do not observe isolated electricity investments in districts without all-weather road access (beyond a handful of isolated cases).[31] Therefore, we face a situation of effectively two endogenous variables for which we require two instrumental variables: a roads instrument, and an instrument for the interaction between roads and electricity investment, which we construct in a standard fashion by interaction our roads IV with the electricity IV. To account for this feature, we always drop the level effect for electricity in all of our results below.

First stage results are presented in Tables 3 and 4 and show a strong and statistically significant relationship between instrumental variables and endogenous regressors. Cragg-Donald, Sanderson-Windmeijer and classic F-test statistics all reject the joint null.
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 rug-

[31] All of our results below are robust to excluding these nine cases.

gedness. The Online Appendix shows results on the full sample of district-years available without the restrictions imposed on sampling by the repeated cross-section nature of the DHS and NLFS surveys used as our dependent variable of interest. These complementary first stages show strong instruments even when including both year and district fixed effects, for which the 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:

$$\begin{aligned} Agriculture_{d,t} = \alpha + \beta_1 Roadind_{d,t} + \beta_2 Roadind_{d,t} * Stationind_{d,t} \\ + \delta_d + \lambda_t + \epsilon_{d,t} \end{aligned} \tag{23}$$

However, following above comments on the endogeneity concerns associated with the likely biased OLS estimator, we instead run two-stage least squares (2SLS) on the following specifications, with year-fixed effects and district-level initial values as controls:

$$\begin{aligned} Roadind_{d,t} = \kappa + \eta_1 RoadIV_{d,t} + \eta_2 RoadIV_{d,t} * StationIV_{d,t} \\ + X_d' \tau + \rho_t + \nu_{d,t} \end{aligned} \tag{24}$$

$$\begin{aligned} Roadind_{d,t} * Stationind_{d,t} = \kappa + \eta_3 RoadIV_{d,t} + \eta_4 RoadIV_{d,t} * StationIV_{d,t} \\ + X_d' \tau + \rho_t + \nu_{d,t} \end{aligned} \tag{25}$$

$$\begin{aligned} Agriculture_{d,t} = \alpha + \beta_1^{2SLS} \widehat{Roadind}_{d,t} + \beta_2^{2SLS} Roadind_{d,t} \widehat{* Stationind}_{d,t} \\ + X_d' \gamma + \lambda_t + \epsilon_{d,t} \end{aligned} \tag{26}$$

In our specification of interest, equation (26), the outcome variables $Agriculture_{d,t}$, $Services_{d,t}$ or $Manufacturing_{d,t}$ represent the share of people reporting an agricultural, services or manufacturing sector occupation, respectively, in district d, aggregated from all EA's (villages) i in that district, in the year of DHS (NLFS) survey round t.
$Roadind_d, t$ represents a dummy if district d contains an all-weather road in year t, while $Roadind_d, t * Stationind_d, t$ 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 (ruggedness, distance to market town, distance to administrative zone capital, soil quality) or would be bad controls if included as time-varying controls (nightlights, or even temperature anomalies).

Our coefficient of interest on the effect of access to a road is captured by β_1^{2SLS}, while β_2^{2SLS} captures what we deem the potentially complementary effect of roads access when combined with electricity.

5 Reduced-form Results

We begin by estimating local average treatment effects of the two treatments outlined above: the effects of roads only and roads in combination with electricity access on sectoral employment at the district-year level. Table 5 provides the results from regression equation (26) run on the National Labour Force Surveys (NLFS) sectoral employment, in a specification with year fixed effects and aforementioned controls of district-level initial values. Standard errors are clustered throughout at the district-level, which is the level of the treatment.[32] Table A10 provides the same specification on a NLFS sample without 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.
We find that compared to districts without road access, an all-weather road causes a 12 per cent decrease in relative manufacturing, while employment associated with the services sector increases by a statistically significant 12 per cent.
However, the effects once road access is combined with electricity access are markedly different: compared to a district with only a road connection (row 1 of Table 5), the interaction of roads with electricity appears to decrease agricultural employment, although insignificantly so. The large standard errors point to substantial underlying heterogeneity (see findings on spatial heterogeneity below), whereas manufacturing employment increases by approximately 10 per cent, with small, insignificant effects on services.[33]

The above results show sectoral employment splits based on first-digit International Standard Classification of Occupations (ISCO) occupational groups, which is our preferred definition. For robustness, Table 6 confirms the above qualitative results based on sectoral classifications derived from first-digit International Standard Industrial Classification of All Economic Activities (ISIC) industry groups.

We provide further supporting evidence on our preferred interpretation of comple-

[32] In order to also allow for spatial correlation structures beyond arbitrary district borders, we also test for robustness using Conley standard errors. Results are unchanged.

[33] The three sector-coefficients in each row of Table 5 do not sum to zero, however, which implies that the share of people responding to 'not work' decreased (results not shown)

mentary infrastructure investments that cause structural transformation by breaking down previous two-stage least squares results into occupational subgroups, separately for the DHS (into Major Occupational Groups) and the NLFS (into both first-digit ISCO and first-digit ISIC subgroups). Figure 9 presents the DHS breakdown by occupational major group graphically in two panels: the upper panel corresponds to results from the same specification that yielded coefficient estimates in row 1 of Table A7, although with more than three outcome variables (i.e. previously, the relative sectoral employment shares). Instead, outcome variables do now correspond to relative employment shares for three different agricultural occupations, five different services occupations and either skilled or unskilled manufacturing occupations. The lower panel of Figure 9 then corresponds to results obtained from the same specification as those in row 2 of Table A7, again using relative employment shares for ten occupational subgroups as outcome variables, instead of three sectors.

We find that road access especially decreases relative employment for agricultural labourers, but increases employment for both skilled agricultural market workers and self-employed farmers – in line with a speculative interpretation of road access leading to greater market access for agricultural producers (and thus demand for market skills), and potentially greater returns to farming that may lead to higher self-employment. Overall, however, as shown above, relative agricultural employment decreases, mostly at the expense of increased services.

This strong positive effect on services employment is confined to employment in retail and sales, which is mostly informal in nature. It is exactly this anecdotally and macroeconomically widespread shift (see Figure 4 from above), caused by all-weather road access, from agriculture and manufacturing into small, informal petty retail trade which we frame as low-level structural transformation, with unknown aggregate productivity implications.

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 global competition, especially regarding Chinese manufactured goods.

In contrast, when combined with electricity, both infrastructure investments cause strong positive effects on agricultural labourers and a negative effect on self-employed farming employment, which may be related to the adoption of capital-intensive irrigation technology by some.

Regarding services, the combined effect on services (in comparison with a location with only roads access) masks a reduction in retail and other services occupations, but a precisely estimated increase in clerical services employment, usually associated with more white-collar activities and office work.

Returning to a 'big push' argument, we confirm large positive effects on both skilled and unskilled (i.e. mostly factory) employment caused by the addition of electricity access to road access, although the latter is not statistically significant at conventional levels. To provide more detailed results on this latter aspect, Figure A10 provides a breakdown of results from the NLFS by ISCO (upper panel) and ISIC (lower panel) first-digit occupational/industry subgroups. The effect of roads alone is concentrated on increases in relative employment in retail services and sales, at the expense of traditional artisinal manufacturing and crafts. With additional electrification, skilled agricultural employment falls, crafts and elementary occupations increase, albeit insignificantly, whereas plant operations and professional occupations increase siginificantly.

The lower panel of Figure A10 provides additional insights into which industry subgroups are mostly affected: roads cause employment decreases in manufacturing, mining, construction and administrative industries, while it increases in wholesale-retail and education industries. In contrast, electrification leads to significantly more employment in manufacturing, construction, accommodation and food, as well as mining.

One key result that motivates our spatial general equilibrium model below is shown in Table 8: the substantial spatial heterogeneity in structural transformation outcomes across space. For example, 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, in live with the low-level structural transformation results discussed above. Quite in contrast, the employment effects on manufacturing in column (3) of Table 8 are relatively more muted and statistically insignificant. The overall services employment increase, however, appears predominantly driven by far-from zone capital districts (+20.8%). Interestingly the agriculture effects, although insignificant, have opposing signs across close-to-town vs far-from-town districts. In sum, these results appear to point firstly to structural transformation patterns along districts' likely comparative advantage, and secondly to potential agglomeration economies of manufacturing sector employment. Both inform our spatial general equilibrium model in Section 2 below.

An additional, often overlooked aspect of structural transformation processes is its heterogeneity across gender: as we highlight in Figure A12 and Table A11, results across sexes are notably different. Whereas the relative decrease in manufacturing employment at the expense of services due to road access is mostly driven by females, the infrastructure complementarity effect out of agriculture into manufacturing with electricity access is mostly driven by males. To the best of our knowledge, this is the first empirical evidence of strongly diverging structural transformation processes in a modern developing country by gender.

Regarding the mechanisms behind our core results presented above, we study several supplementary outcomes of interest: migration, demographics, education and labour force participation.

No differential migration responses can be detected by roads access, whereas additional electricity does lead to economically meaningful, but only marginally significant immigration responses (see Table 9).

The demographic makeup of electrified districts also changes markedly: workers are on average 2.2 years older than in road-connected districts (Table A12). As Figure 12 highlights, this effect is mostly driven by a considerable narrowing of the age distribution in road- and electricity-connected districts around 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 estimation, as shown in Figure A11, where electrified districts add especially workers between the second and sixth decile of the age distribution, i.e. between 18 and 33 years.

Column 4 of Table A12 also shows how the share of divorced workers increases in road and electricity districts, which may be interpreted as a proxy for the arrival of greater economic opportunities that decrease the economic value of marriage.

With respect to education, the overall education results (cf. Table A13 are ambiguous since only road-connected districts show increases in literacy (column 1), whereas years of educational attainment only increases insignificantly. A placebo test if these education effects are indeed driven by infrastucture is provided in Table A14: 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.[34] Finally, overall labour force participation shows an insignificant positive effect from roads, and a strong negative effect from electrification, in line with the age pyramid

[34]Taken at face value, the negative coefficient in Table A14, column 5, row 2 would thus indicate negative selection of migrants into road- and electricity-connected districts. This result may or may not be counterintuitive, depending on the skill requirements of newly arising plant operations and construction subsector jobs.

narrowing highlighted above: many teenagers opt out of the labour force, either due to lower fertility ('missing youths') or young people staying in education.

Finally, we attempt to shed light on the likely growth and welfare implications of our core reduced-form results. As a first proxy, Table 13 reports two-stage least-squares estimates of satellite-derived outcomes on treatments: we do find roads to lead to geater 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 electrification appears to result in more noticeable economic development. One caveat, 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.

Results for households' real expenditure, a potentially more concrete proxy of economic development following Young's (2012) approach, are presented in Tables 11 and 12. Confirming our interpretation of our structural transformation results, we do not find statistically significant improvements in real expenditure on either durables (Table 11) or housing (Table 12) from road access alone, whereas real consumption of eight out of twelve categories does markedly increase with additional electrification. If higher relative sectoral employment in manufacturing is in fact representative of higher aggregate productivity, one would expect exactly this pattern of real consumption results.

Current work in progress analyses household survey responses from several rounds of the Ethiopian Household, Income and Expenditure surveys, which have a similar spatial and temporal coverage as the NLFS and should confirm the tentative results on material effects of infrastructure on consumption presented here.

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 structure to characterise a spatial general equilibrium, we turn to structurally estimating the general equilibrium relationship between infrastructure investments and structural transformation.

6.1 Data for structural estimation

For the structural estimation below, we require additional data on model inputs that were not of primary interest in the reduced-form estimation above. In particular, we require a matrix of district to district trade costs, information on the supply of land, proxy measures for the productivity in either production sector, as well as population information for every district in Ethiopia.

Regarding trade cost, we compute an exhaustive matrix of district centroid to distance centroid least-cost distances by means of an Dijkstra algorithm employed on a tailored cost surface. The algorithm then determines the least-costly route to connect each district centroid to each other, separately. We generate the underlying cost surface from a terrain raster image overlayed with the year-specific rasterised all-weather road vector layers. We 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 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 geography of Ethiopia is represented by a graph of approximately 12,000 quadratic pixels, each representing approximately an area of 180 times 180 meters (when measured at the equator). Hence, for pixels without an all-weather road in it, we measure the cost to cross the pixel as the distance (in kilometers) 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, we 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, our approach is equivalent to understanding a given all-weather road in a pixel to virtually level the terrain in trade cost terms.

With respect to land area, we use fertile land derived from satellite imagery, that is land either deemed theoretically inhabitable or suitable for productive use. This choice of proxy is problematic, however, if one thinks of land in the model as representing mostly housing, since both are obviously distinct and not necessarily even correlated. Data on housing stock and its value in Ethiopia is virtually nonexistent since the Ethiopian real estate market remains monopolised by government ownership of land, essentially a leftover from former socialist regimes in power: all land is owned by the government and firms or residents only obtain non-permanent permission to use any land without owning it. Hence, we use fertile land as one possible proxy for land supply for which data exists, which appears reasonable especially for the model extension in which land enters traditional sector production as a second input.

Regarding sectoral productivities, we 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.
Appendix Figure **??** shows 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. 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 ours. 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, which are also reported in Table 15.[35]

For the modern sector productivity, we lack any country-wide proxies for it. However, we 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, we 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, for population data, we 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 in-

[35]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 our case, fitting a Weibull distribution to our 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 ($s.e.$ $= 0.0749$).

formation on district-level population, we 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.

6.2 Calibration

For the calibration of the model, we use data on initial district population, initial sectoral employment shares and district land area as inputs to pin down sectoral productivities and amenity shocks.[36] These initial values refer to the years 1999/2000, which maps directly to the earliest survey rounds employed in our reduced-form analysis above.

In addition to the data vectors L_n, λ_n and H_n, we also use the Dijkstra algorithm-derived effective distance matrix between the 689 Ethiopian district capitals (or centroids). As detailed in Table 15, we estimate the elasticity of trade cost with respect to distance from the rich, raw panel data underlying the Central Statistical Agency's Retail Price Index (RPI). We follow Atkin and Donaldson's (2015) procedure to estimate this elasticity solely from goods prices, albeit employing price data with greater temporal coverage and exploiting several more confirmed product origin locations to construct destination-origin price gaps.

The sectoral employment share for the modern sector relates one for one to the manufacturing sector's employment share in each district as of 1999/2000. In particular, to maximise the sample of available initial data points, we 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, we obtain 1999/2000 manufacturing employment share data for 475 out of the total 689 districts used in our analysis. Out of these 475, 181 districts appear only in the NLFS for the initial period,

[36] This setup resembles an earlier version of Michaels, Rauch and Redding's (2012) calibration, although our calibration exercise features one noticeably simpler aspect (i.e. Cobb-Douglas upper-tier consumer expenditure shares) and one more complex aspect (i.e. endogenous within-location wage equalisation).

58 appear only in the DHS and 236 appear in both. For the missing 214 districts, we 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 our 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.

[parameter settings and calibration procedure here]

6.3 Shocks to the system

As argued above, we model road infrastructure as directly affecting transport costs via its empirically realistic effect on enabling bulk transport of goods to locations that were previously unreachable or only intermittently reachable by lorry. Regarding electrification, in an empirical setting where no direct substitutes for a constant supply of energy exist apart from decentralised, diesel generators, the arrival of the grid reduces the cost of energy and may even represent an extensive margin change with respect to the application of power-driven means of production. We thus opt to understand electrification in our context as a shock directly affecting improving the productivity of an electrified location.

Our modelling choices link directly to empirical reality: Table 14 highlights that there is a strong positive association between the district-level all-weather road indicator, our treatment indicator used in the reduced-form estimation in Section 5 above, and the transport cost matrix between district centroids or capitals, which we obtain from a Dijkstra least-cost algorithm employed on a terrain-weighted distance graph. Since we are not constrained by gaps in the coverage of outcome variables for either districts or years in this descriptive exercise, we 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 we have distinct information on the extent of the all-weather road network as described in Section 3 above.[37] As columns (3) and (4) in Table 14 show, a district getting connected to the all-weather road network (such that the roads indicator switches from zero to one) is associated with a 4% reduction in the sum of

[37] This implies, necessarily, that we also run the Dijkstra least-cost algorithm four times on distinct cost surfaces.

that district's least-cost distances to all other districts. In other words, the Dijkstra least-cost algorithm is actually proposing least-cost connections between district capitals and/or centroids that 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 some transformation, trade costs).

Regarding the transport cost shock amplitudes, Appendix Figure A13 shows that the long difference in relative changes in Dijkstra algorithm least-cost distances from the earliest (1999) to the latest point in our 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, we 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[38] are affected most: at least seven distinct zones affected by large transport cost shocks emerge from Appendix Figure A13, 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, circa 350-500 km south-southwest of Addis), as well as central Oromia (Arsi, circa 150-250 km south of Addis) and eastern Oromia (Harerge, circa 300 km west of Addis).

We are therefore confident that the stylised shocks to the general equilibrium system do in fact capture empirically realistic changes on the ground and provide a useful proxy for the reduced-form measures of infrastructure investments we employed above.

6.4 Estimation

A standard approach to structural estimation would be to calibrate the numerically solved spatial general equilibrium model to initial outcome variable levels. In our case, this would involve a grid search over parameter values to match sectoral employment and population levels across Ethiopian districts in our initial sample year of 1999. The resulting, calibrated model could then be used to solve for sectoral employment and

[38] Cf. Appendix Figure A3 for the distribution of population towards the end of our sample.

population changes in response to shocks to the system, such as changes to the 1999 trade cost matrix, for example due to newly constructed all-weather roads that change trade cost of all districts to that newly connected district. Likewise, a naive mapping of electrification to productivity would allow the empirically observed roll-out of the grid over time to be exploited as shock to the system.

One crucial underlying assumption of the standard calibration approach, however, would be that shocks to the system are orthogonal to conditions on the ground that would feed as inputs into the structural estimation, such as initial population levels. Such an orthogonality assumption appears unlikely to hold in our context: as discussed in greater detail in Subsection 4, infrastructure investments in Ethiopia were targeted on observable and unobservable characteristics of to-be-connected locations following various policy objectives.

Therefore, we propose an alternative estimation strategy that exploits the plausibly causal, quasi-experimental reduced-form estimates from above to calibrate not the model's initial levels, but instead to match its estimated treatment effects. This approach, similar in spirit to recent advances in the spatial economics such as Faber and Gaubert (2016) and Adão, Arkolakis and Esposito (2019), builds on the realisation that two-stage least-squares delivers a local average treatment effect that, conditional on the instrumental variable assumptions, captures a causal effect net of general equilibrium repercussions of the shock affecting economic fundamentals in locations not directly shocked, but indirectly affected by reallocations of labour either across locations or across sectors.

More specifically, given our estimates of β^{RF}_{2SLS} reported in Section 5 above, we solve our model for some initial parameter settings and initial input values, such as population, land area or productivity. We then let the model predict the relevant outcome variable, i.e. sectoral employment shares across all locations, in response to the empirically observed variation in treatment settings for each pseudo-panel year separately (i.e. 1999, 2005 and 2013 for the NLFS sample; 2000, 2005, 2011 and 2016 for the DHS sample). We then proceed by running our preferred two-stage least-squares specification on the resulting simulated sectoral employment shares across treatment settings over time and report 2SLS estimates from this regression. The final step in this procedure is then to alter the initial parameter settings of the model until both the reduced-form and the structurally-simulated 2SLS esimates converge.

[describe structural estimation results here]

6.5 Counterfactuals

[describe counterfactuals results here]

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 effects of combining roads and electrification investments 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 occupations. 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 potential interaction effects taken into consideration in the planning of infrastructure investments to maximise their impact.

References

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

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).

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.

Bryan, G. & Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from Indonesia. *Journal of Political Economy*, *127*(5), 2229–2268.

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).

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

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

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.*

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*.

Michaels, G., Rauch, F. & Redding, S. J. (2012). Urbanization and structural transformation. *The Quarterly Journal of Economics*, *127*(2).

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*.

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).

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.

Young, A. (2013). Inequality, the urban-rural gap and migration. *The Quarterly Journal of Economics*, *128*(4), 1727–1785.

8 Figures

Figure 1: Kuznets' Growth Fact: Structural Transformation out of Agriculture

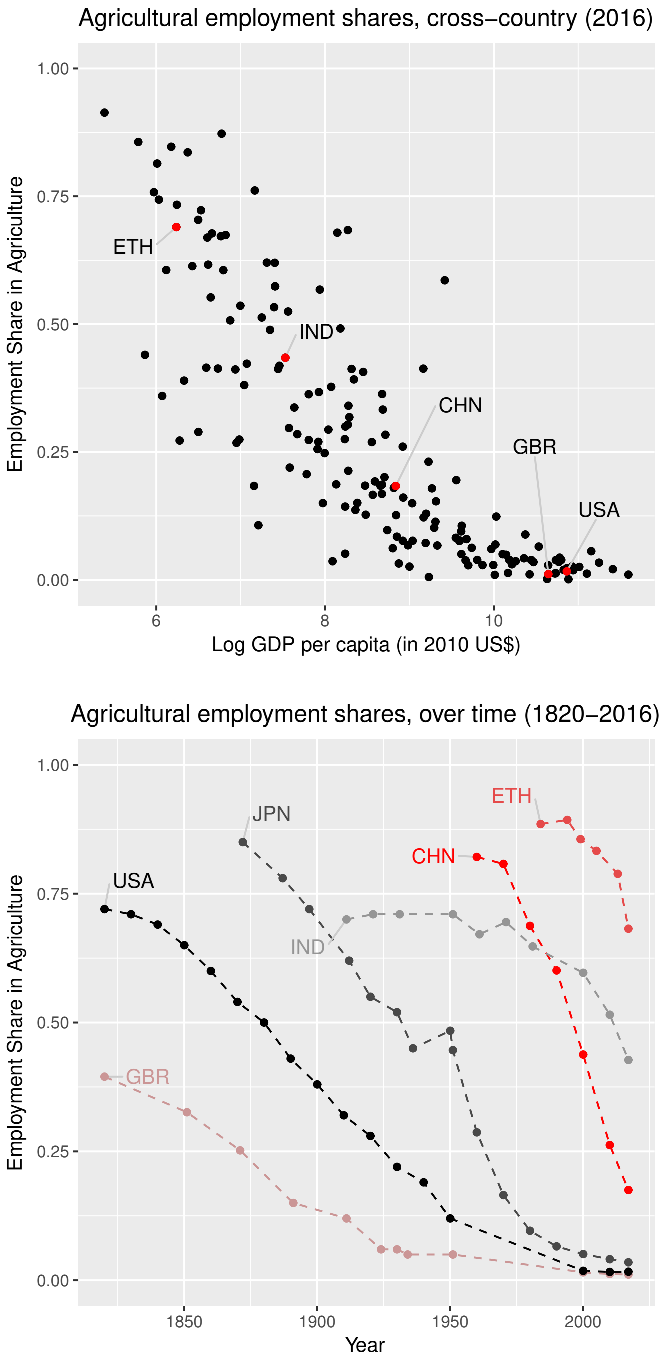
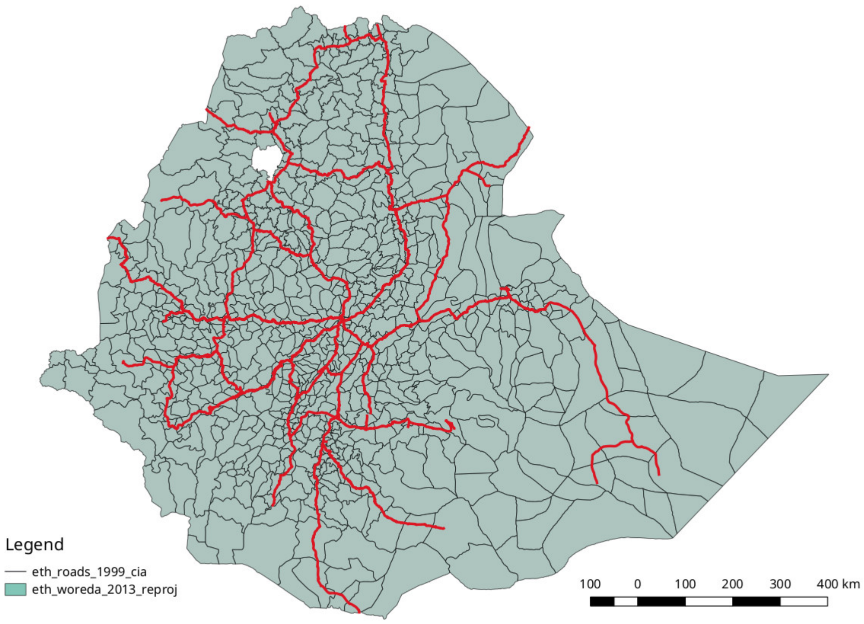


Figure 2: Large-scale Road Network Expansion (2000-2016)



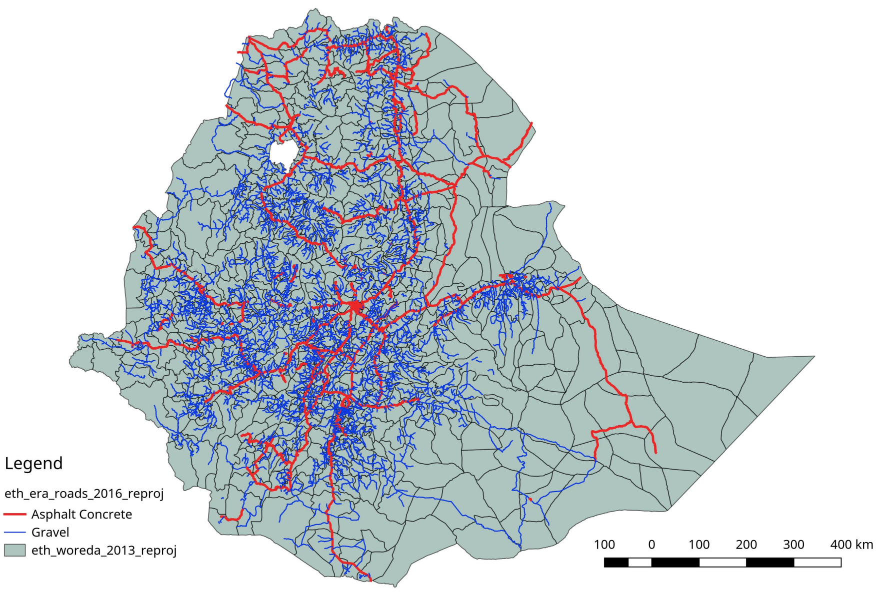


Figure 3: Large-scale Electricity Network Expansion (1991-2013)

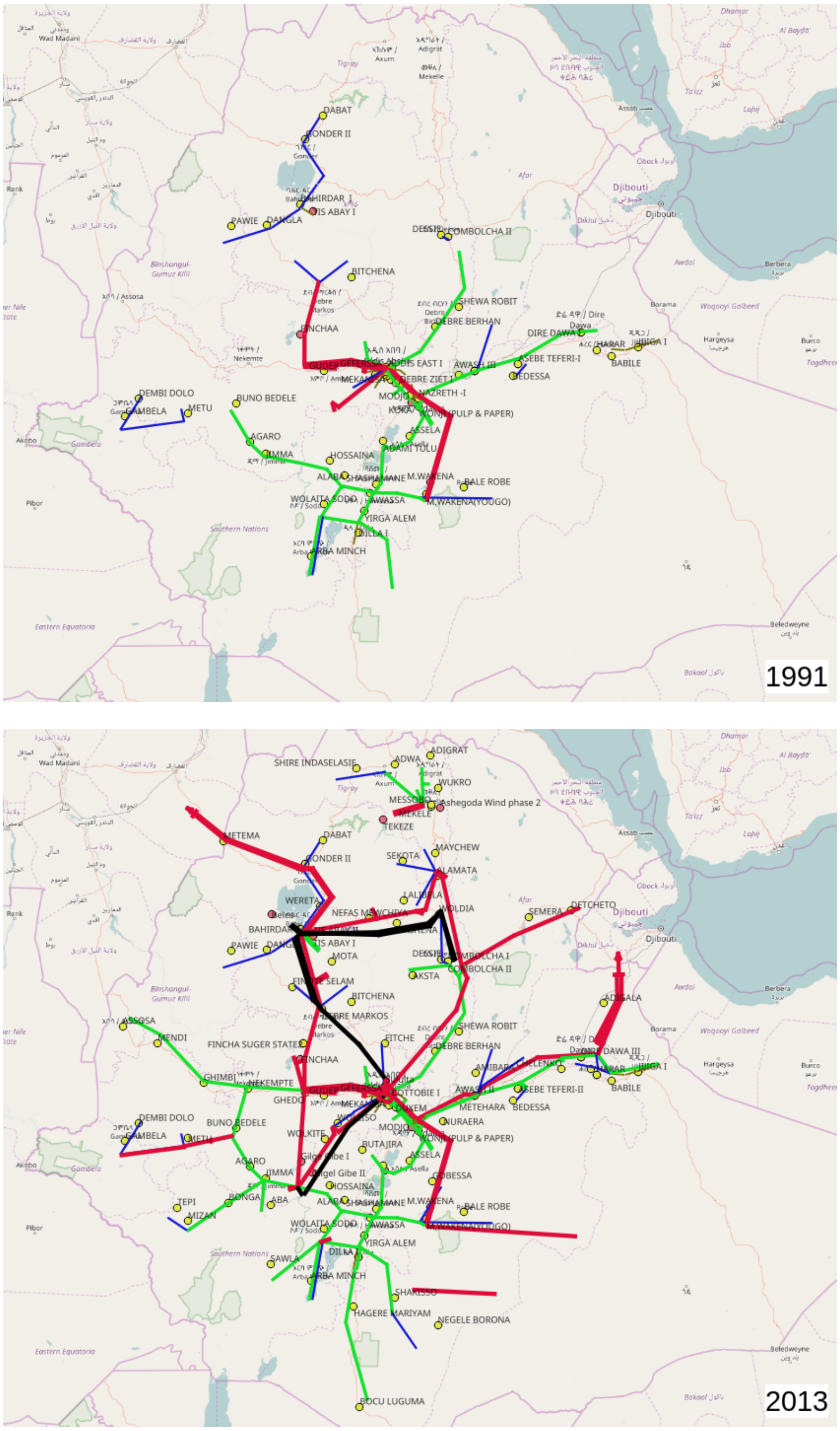


Figure 4: Sectoral Employment in Ethiopia (1994-2016)

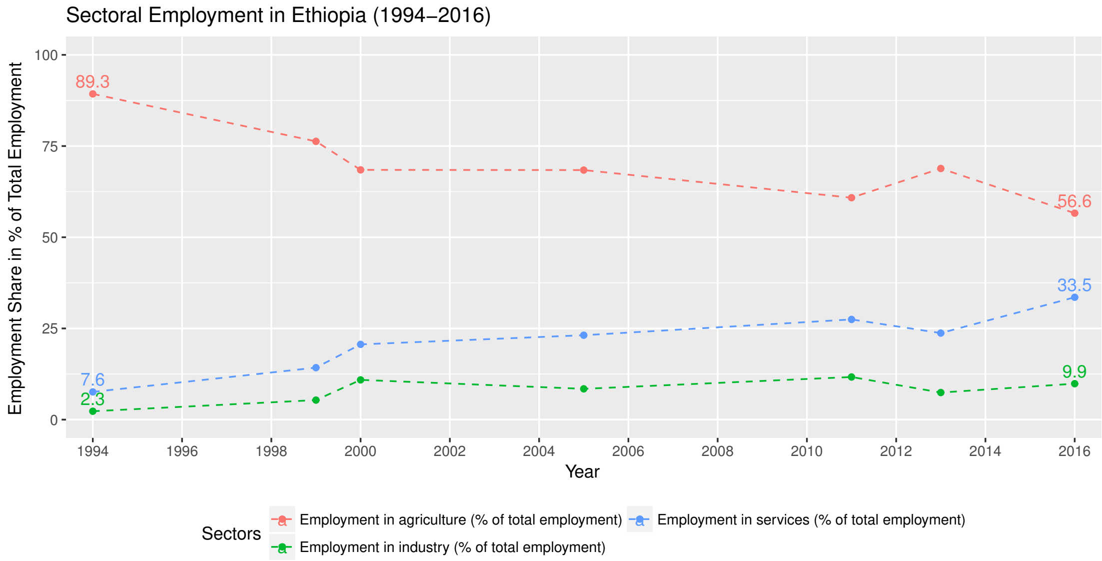


Figure 5: Poverty Headcounts and GDP per Capita in Ethiopia (1994-2016)

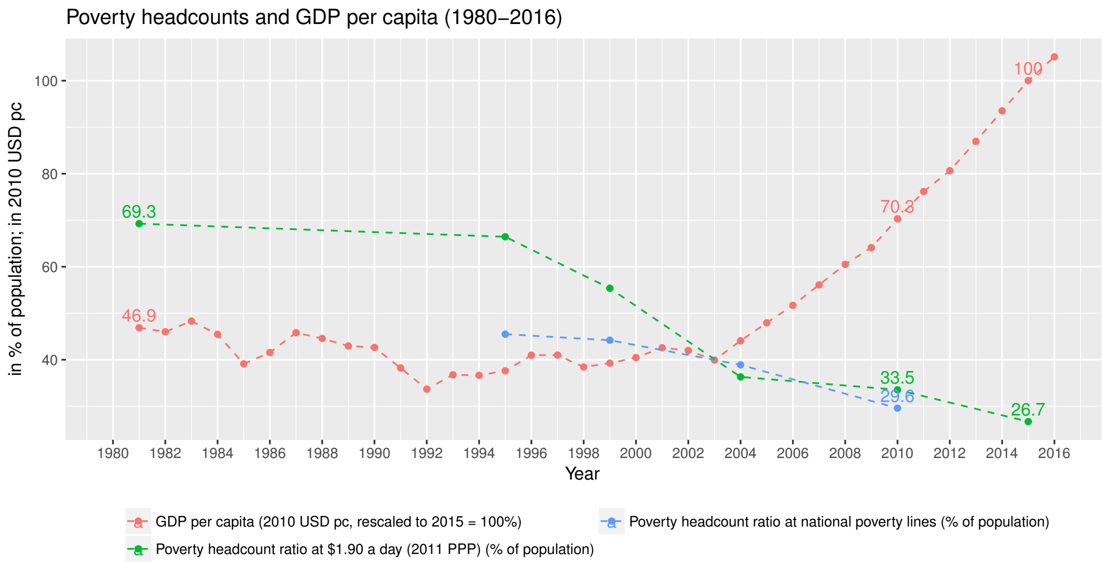


Figure 6: Electrification IV Corridors and Times, Connecting Dams with Addis Abeba

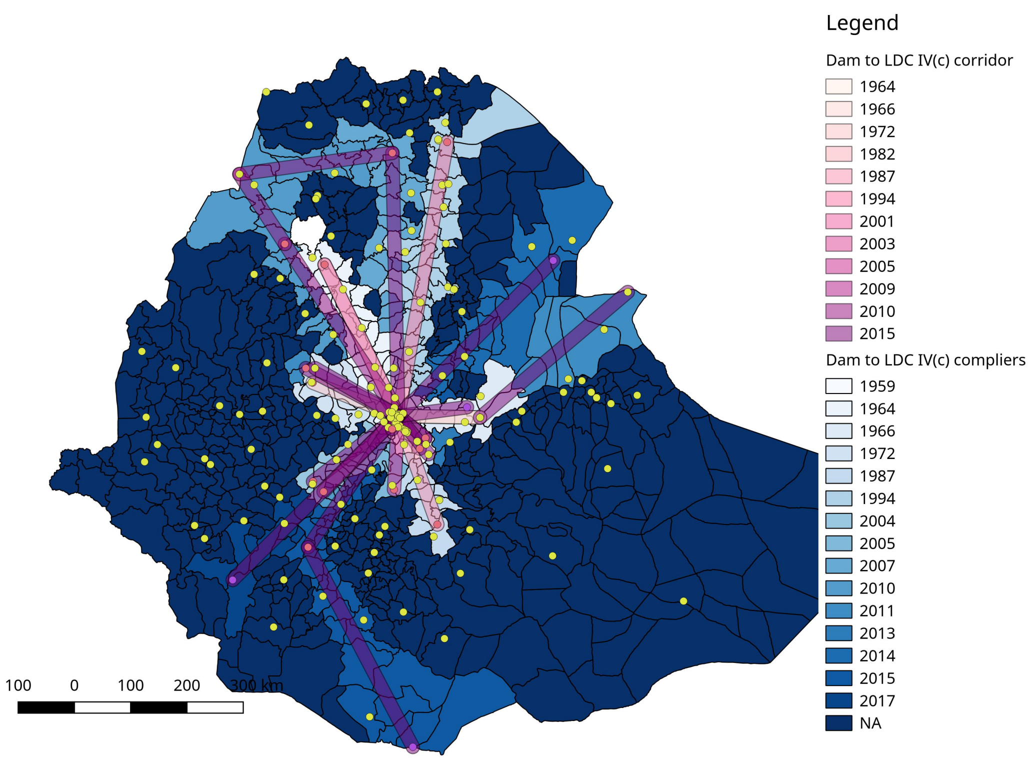


Figure 7: Road IV (Kruskal) District Connection Year to All-weather Road

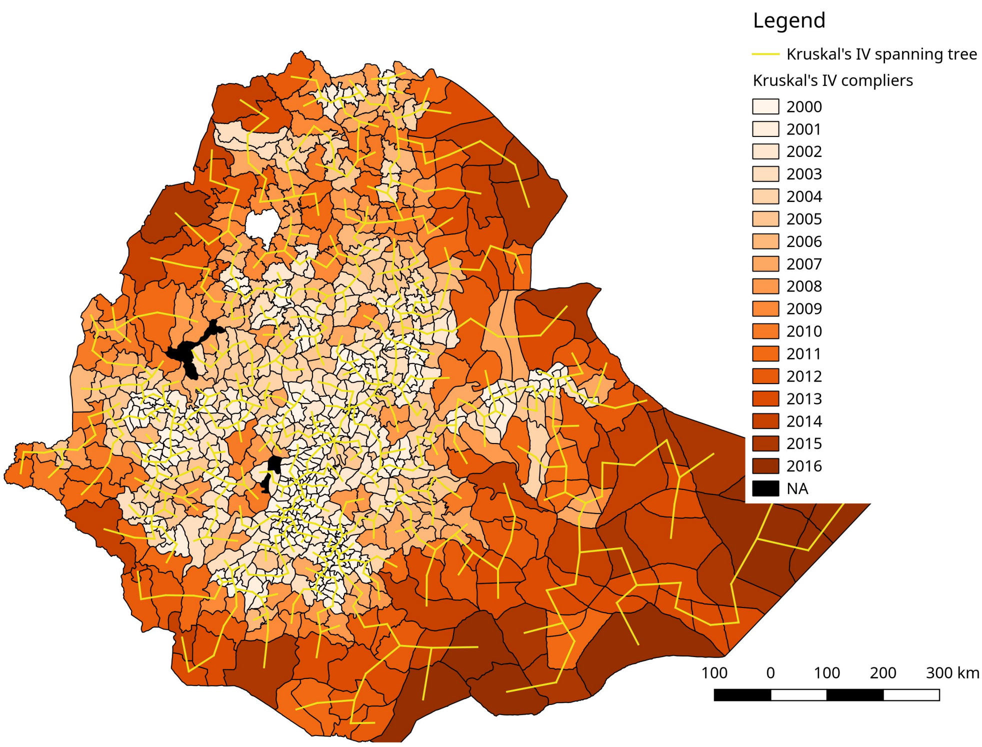


Figure 8: Road IV (Italian) District Connection Year to All-weather Road

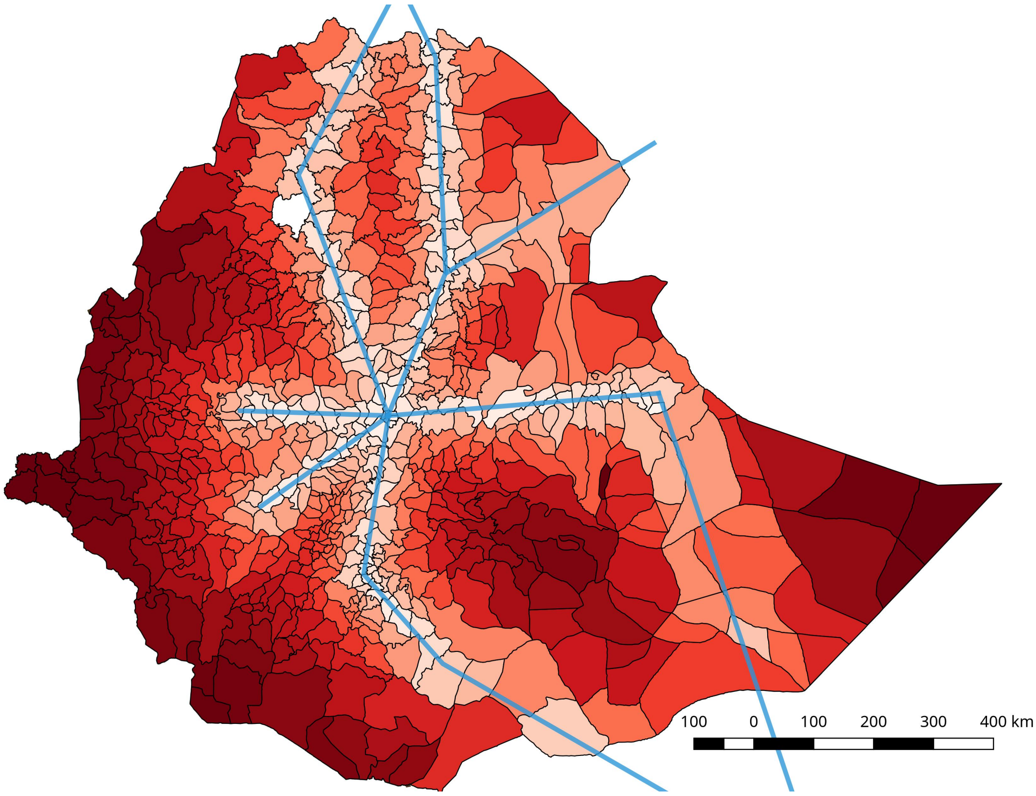
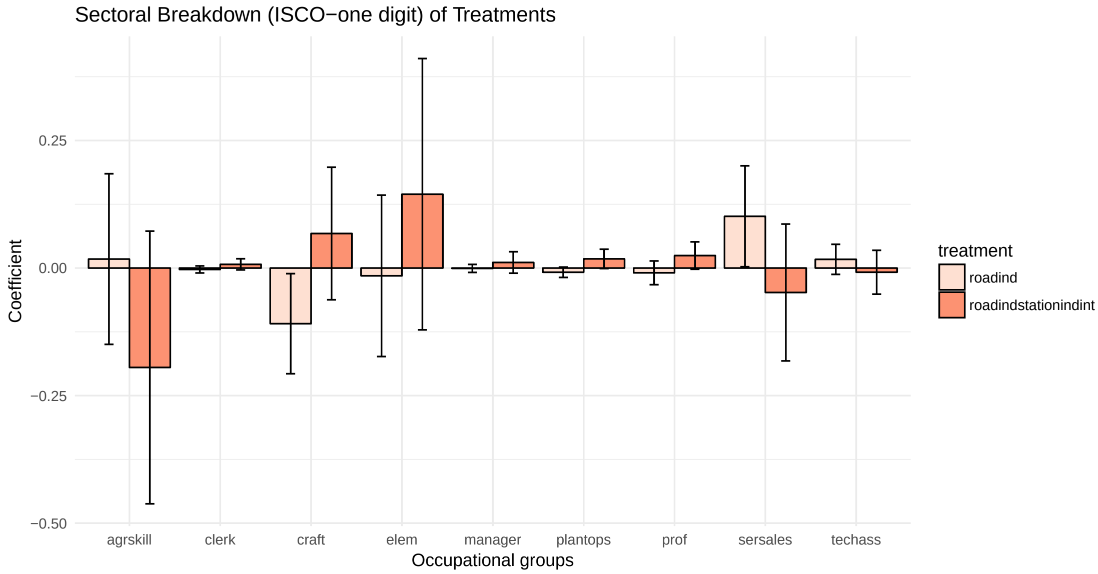


Figure 9: Roads and Roads & Electricity Interaction Coefficients by Occupational Groups (in NLFS or DHS-R dataset)



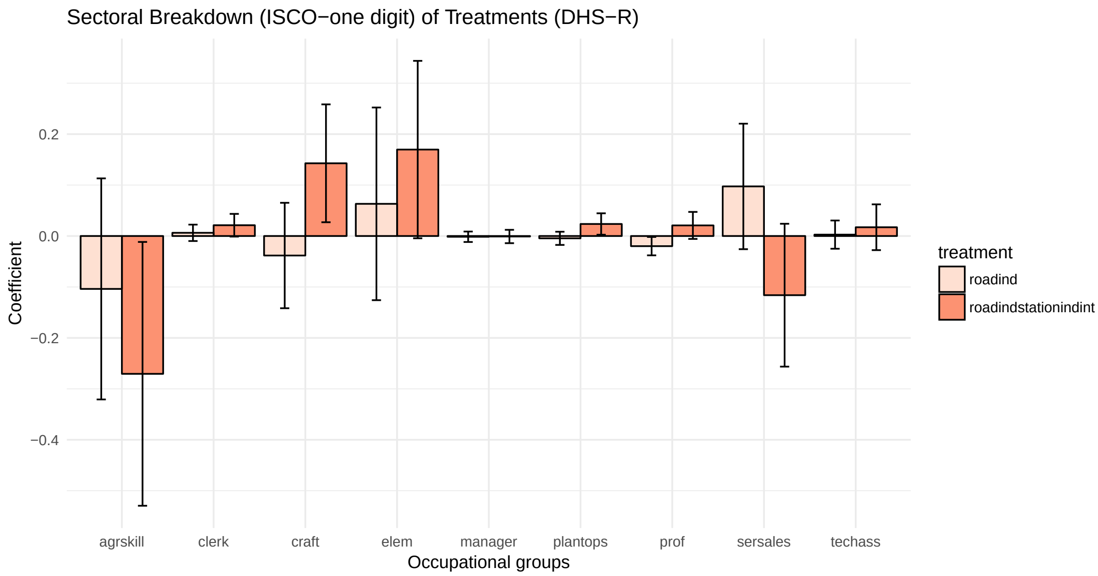


Figure 10: Simulated Change in Manufacturing Shares from Trade Cost Shock

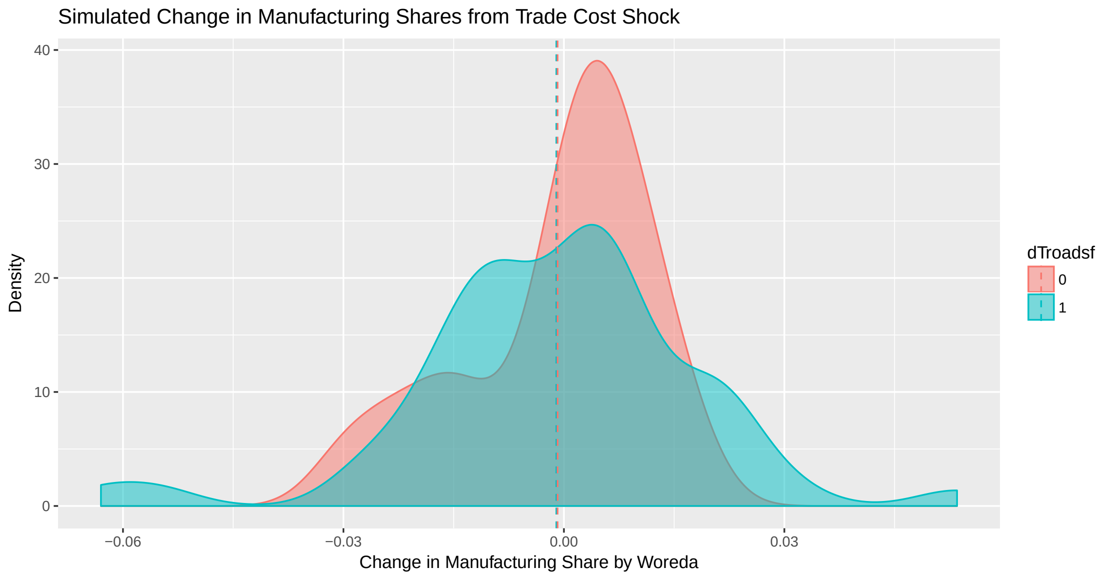


Figure 11: Simulated Change in Manufacturing Shares from Combined Trade Cost and Electrification Shock

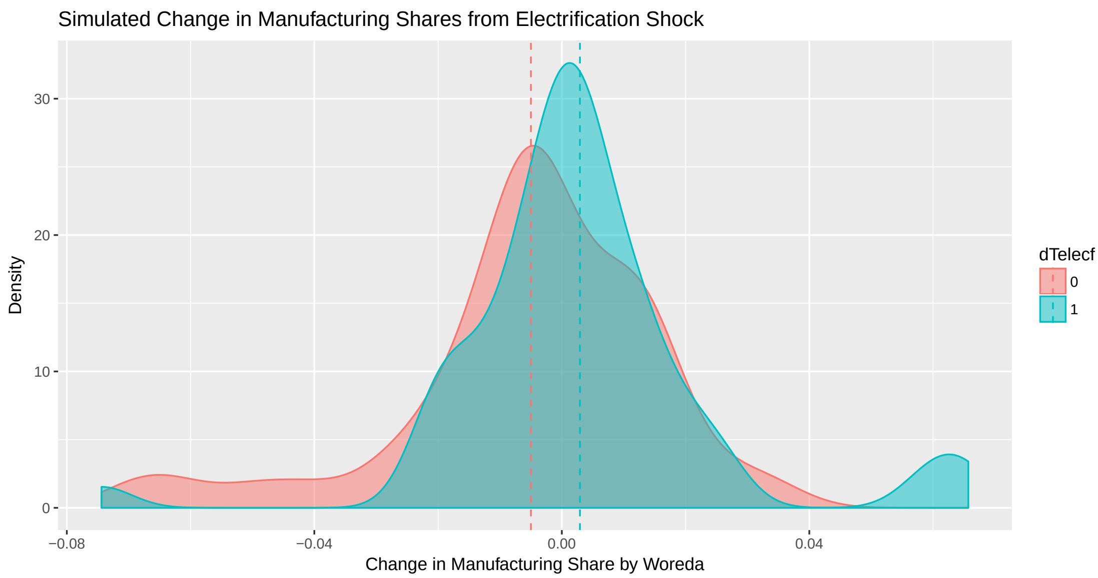


Figure 12: Age Distributions by Treatment Complier Status

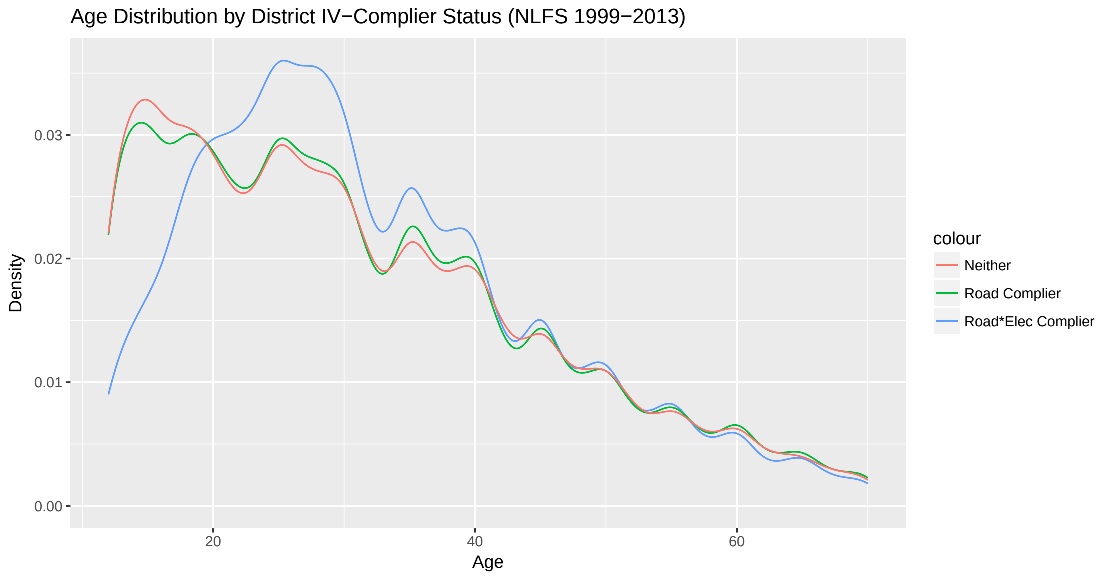
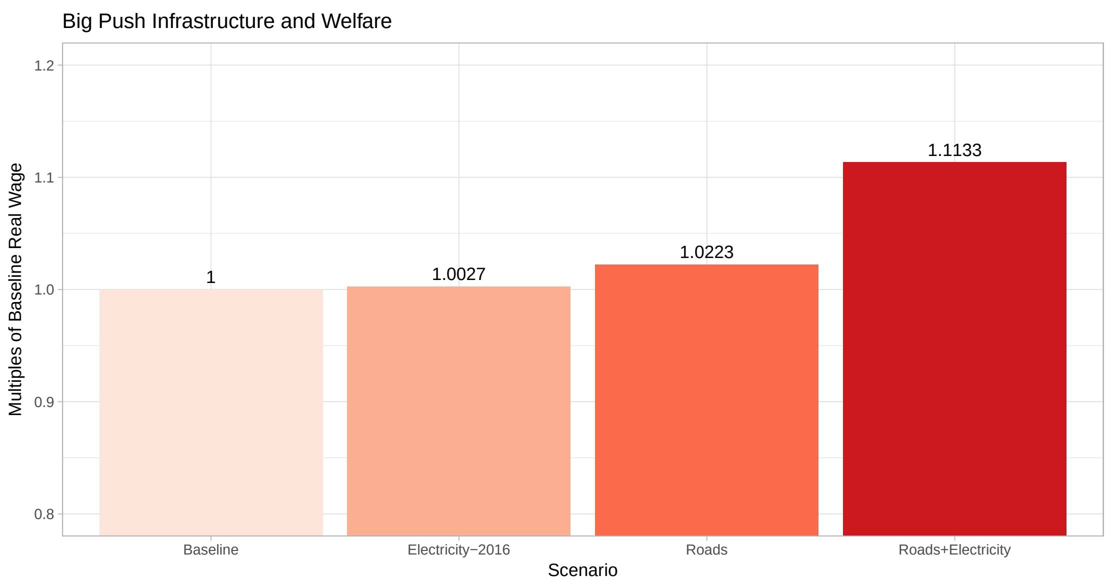


Figure 13: Welfare Estimates of Big Push Infrastructure



9 Tables

Table 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 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 3: First Stage: Roads-IV (Kruskal) and Elec.-IVc 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 4: First Stage: Roads-IV (Kruskal) and Elec.-IVc 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 5: Occup. Change (NLFS), Roads (Kruskal) and Elec. (IVc) (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 6: Occup. Change (NLFS-ISIC, excl. Somali), Roads (Kruskal) and Elec. (IVc) (1999-2013)

	Dependent variable:		
	Agr. [isic]	Ser. [isic]	Man. [isic]
	(1)	(2)	(3)
Road Indicator	-0.074	0.183^{*}	-0.104^{*}
	(0.113)	(0.094)	(0.058)
Road*Elec Ind.	-0.229^{*}	0.029	0.199^{***}
	(0.118)	(0.096)	(0.065)
Model	2SLS	2SLS	2SLS
Year FE	✓	✓	✓
Controls	✓	✓	✓
Cragg-Donald F.		10.327	
Windmeijer cond. F.	17.545	13.208	
p-val $\beta_1 + \beta_2 = 0$	0.0051	0.014	0.1022
Observations	1,188	1,188	1,188
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 7: Occup. Change (DHS-R), Roads (Kruskal) and Elec. (IVc) (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 8: Occup. Change (NLFS, zone cap. dist.), Roads (Kruskal) and Elec. (IVc) (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: Migration (NLFS), Roads (Kruskal) and Elec. (IVc) (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 10: Consumption (HICES), Roads (Kruskal) and Elec. (IVc) (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 11: Durables Exp. (DHS-HR), Roads (Kruskal) and Elec. (IVc) (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 12: Housing Exp. (DHS-HR), Roads (Kruskal) and Elec. (IVc) (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 13: Satellite Outcomes, Roads (Kruskal) and Elec. (IVc) (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 14: 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 15: Parameters for Baseline Structural Model

Parameter	Value	Source	Description
σ	4	Bernard et al. (2003)	Elasticity of substitution between varieties
$1-\alpha$	0.25	Data (HICES)	Expenditure share on land/housing
κ	0.5	Ngai & Pissarides (2008)	Elasticity of substitution across sectors
μ^M	0.85	Data (LMMIS)	Labour share in M-production
μ^T	0.78	Data (AAgSS)	Labour share in T-production
τ	0.3	Data (RPI)	Elasticity of trade cost with respect to distance
θ	4	Donaldson (2010)	Shape parameter of productivity distribution across varieties & locations

Note: HICES denotes the Central Statistical Agency's Household, Income, Consumption and Expenditure Surveys; NLFS denotes the National Labour Force Surveys; RPI denotes the Retail Price Index' raw data; ORCDRC denotes Soilgrids' remotely-sensed Organic Carbon Content data.

Table 16: Structural Estimation: Roads and GE-Model Outputs (2000-2016)

	Dependent variable:					
	λ	L	w	Income	M-Expend.	T-Expend.
	(1)	(2)	(3)	(4)	(5)	(6)
Roads Ind.	-0.004^{***}	1.680^{***}	-0.001	5.596	2.768	19.669^{**}
	(0.001)	(0.630)	(0.003)	(14.884)	(7.467)	(8.600)
Year FE	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
Observations	2,752	2,752	2,752	2,752	2,752	2,752
Note:				*p<0.1; **p<0.05; ***p<0.01		

Appendices

I Appendix: Additional Figures

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

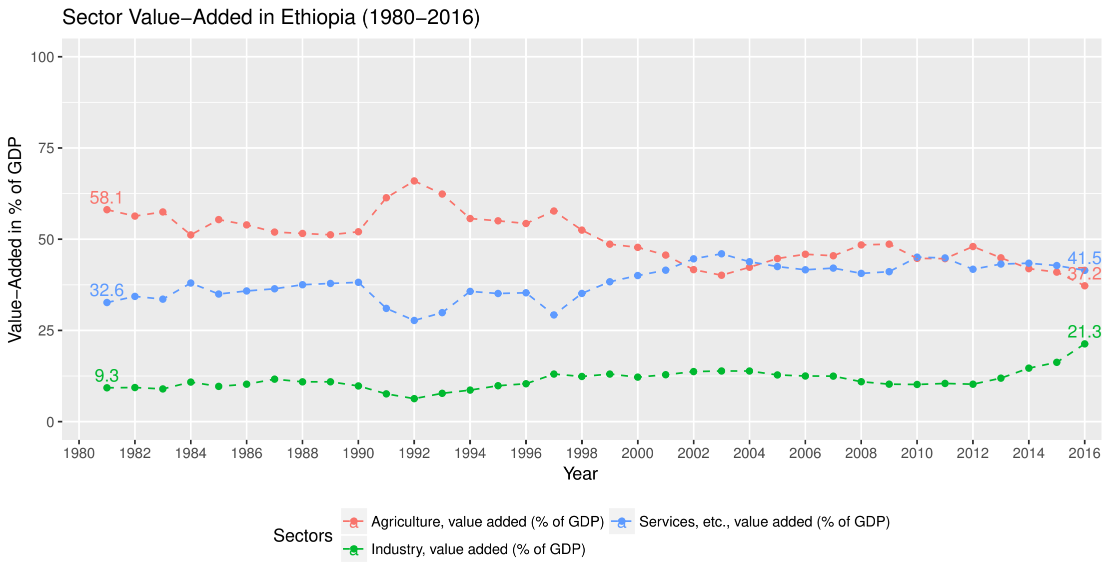


Figure A2: DHS Enumeration Area Locations by Survey Round (2000-2016)

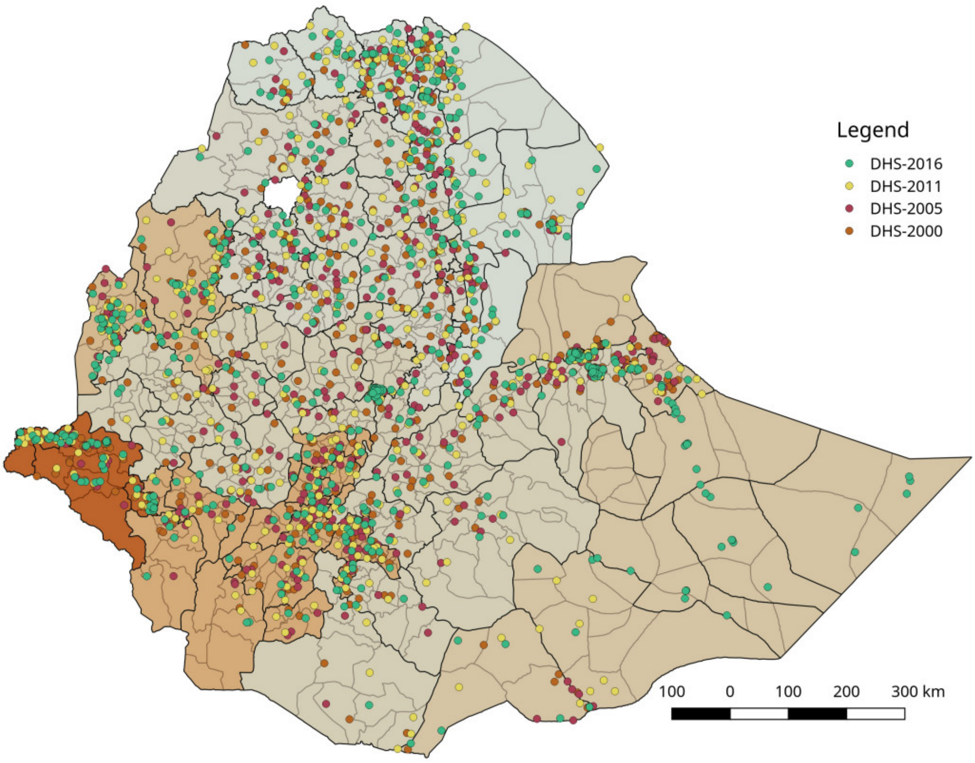


Figure A3: Spatial Variation in Population Density across Ethiopia (2015)

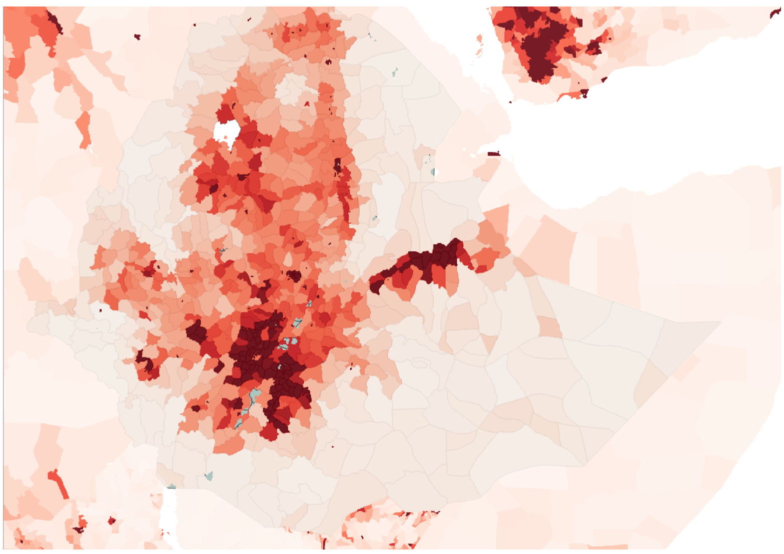


Figure A4: Spatial Variation in Elevation across Ethiopia

Figure A5: Spatial Variation in Terrain Ruggedness across Ethiopia

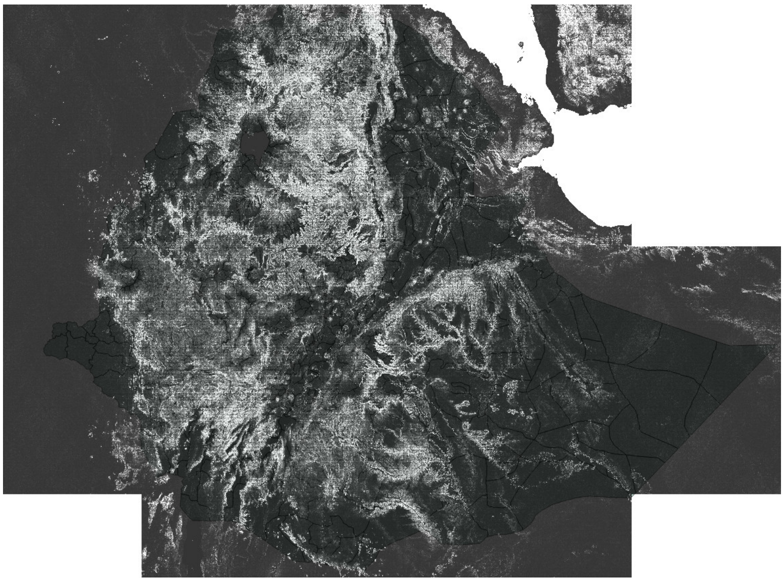


Figure A6: Historic Italian Road Construction in Ethiopia and Eritrea

Figure A7: Reconstructed Italian Colonial Roads and Orthogonal Feeder Roads to Nearby Districts around Debre Berhan (along Dessie–Addis Abeba corridor)

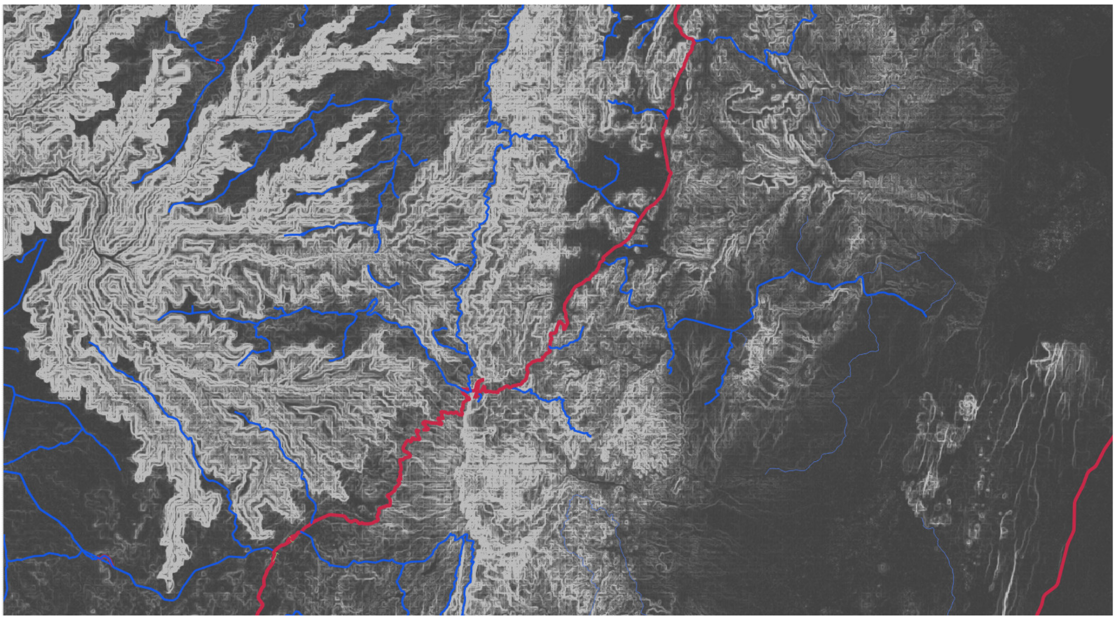


Figure A8: Reconstructed Italian Colonial Roads and Orthogonal Feeder Roads to Nearby Districts around Kulubi (along Harar–Addis Abeba corridor)

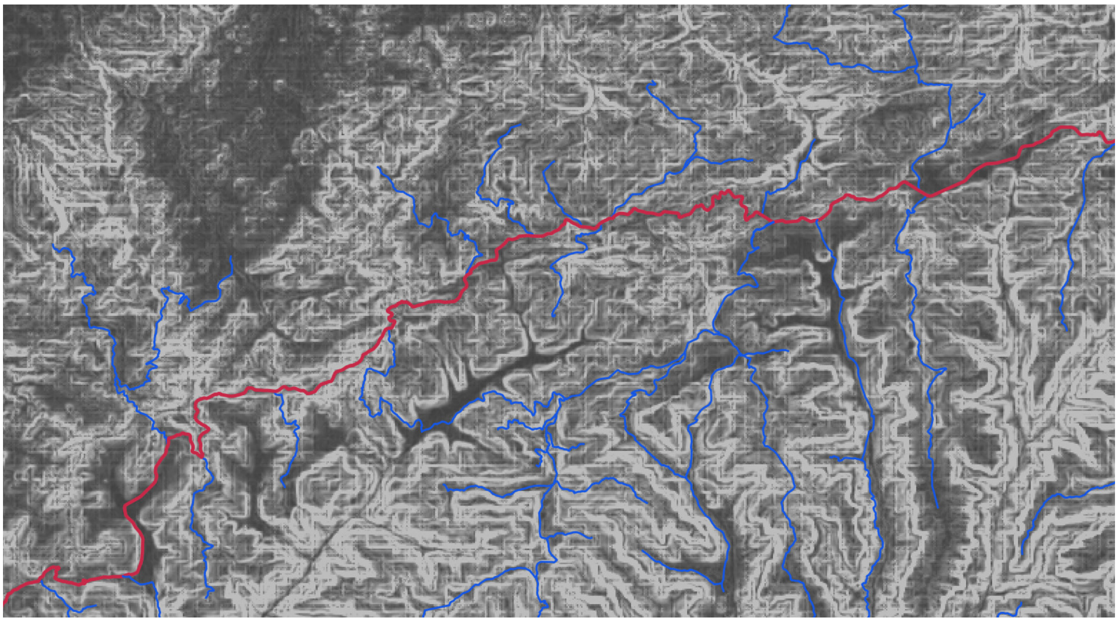


Figure A9: Districts' Road Access Status as Function of Population Density (2005-2013)

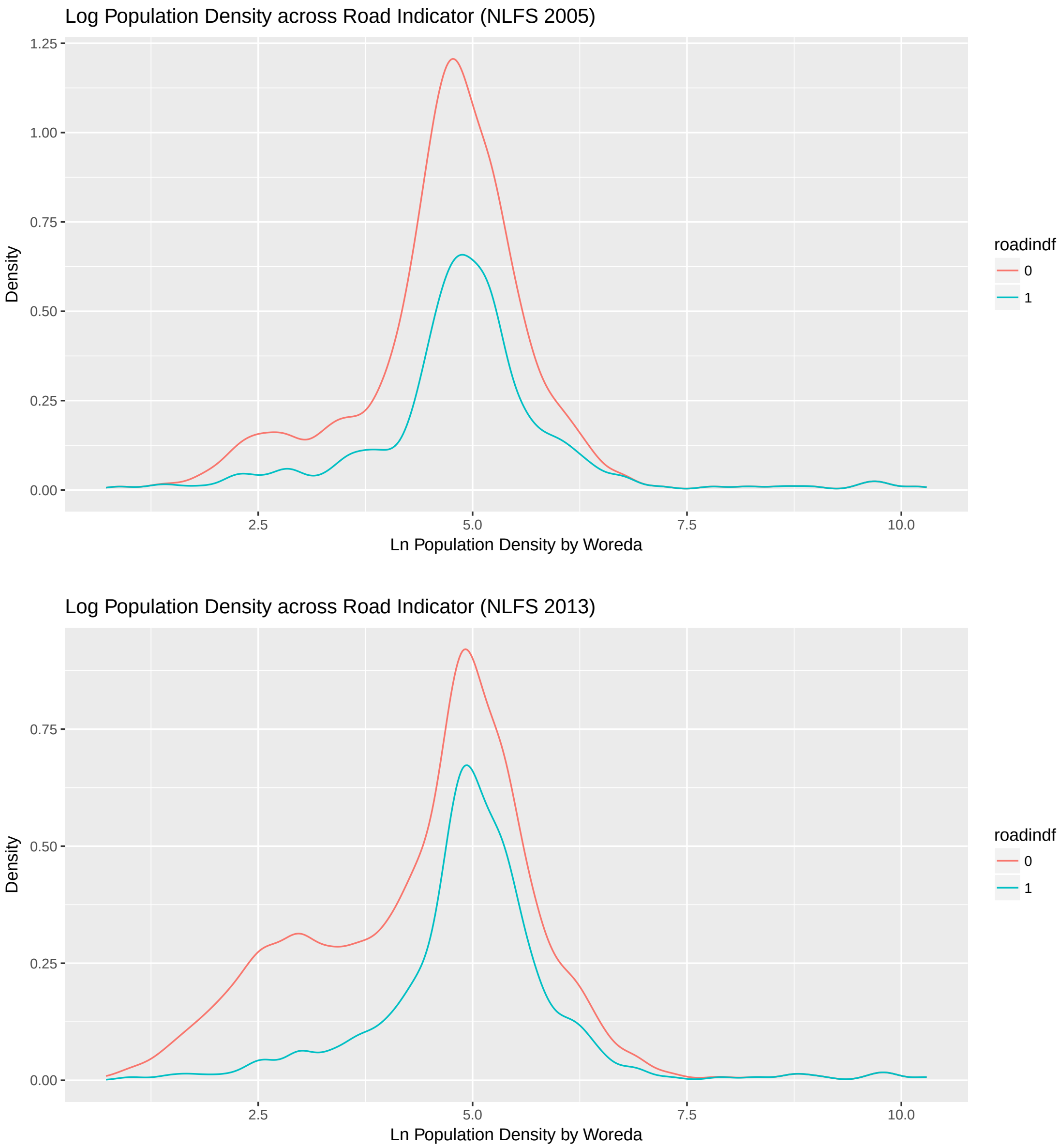


Figure A10: Sectoral Breakdowns (ISCO and ISIC–one digit) of Treatments

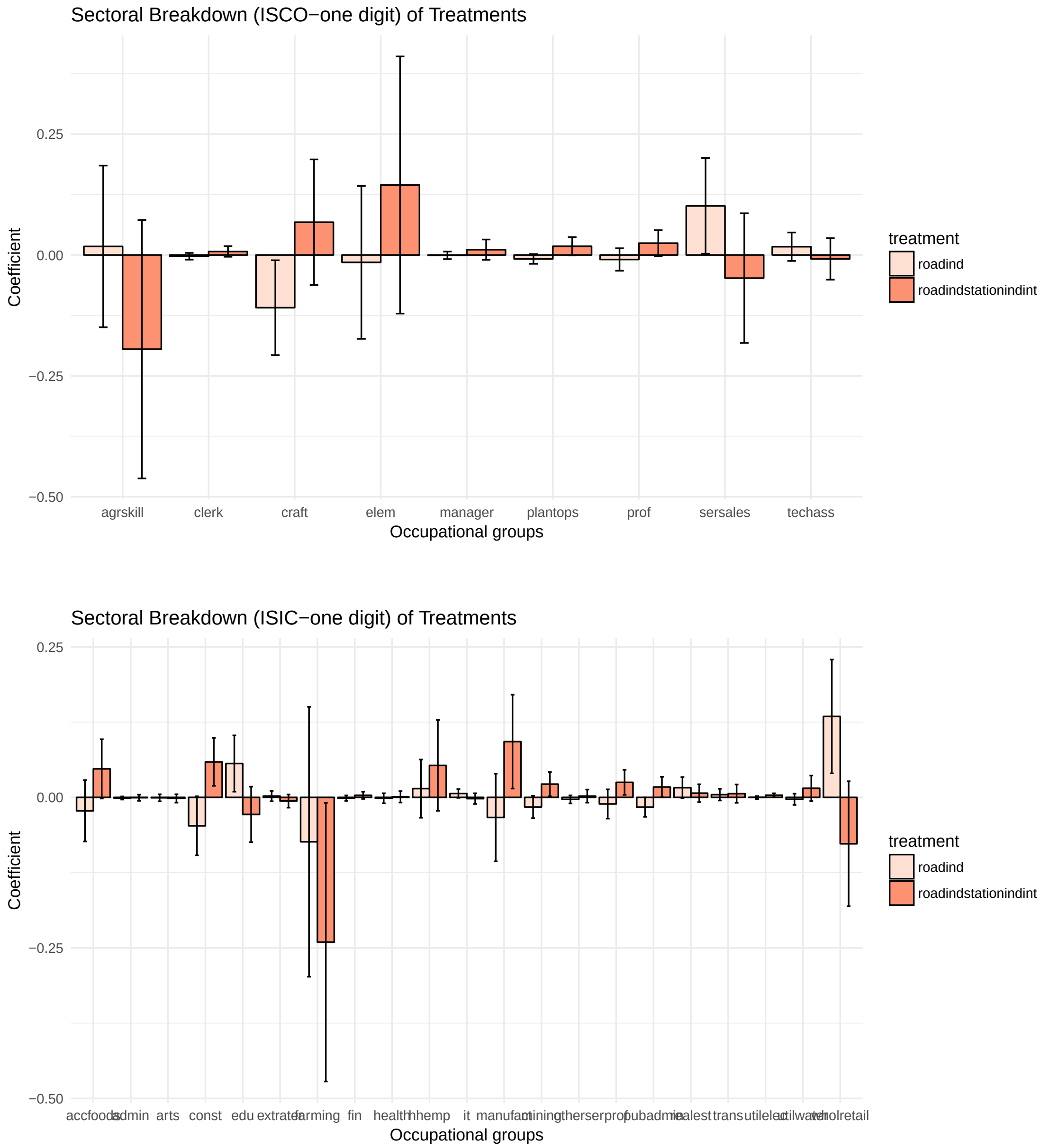
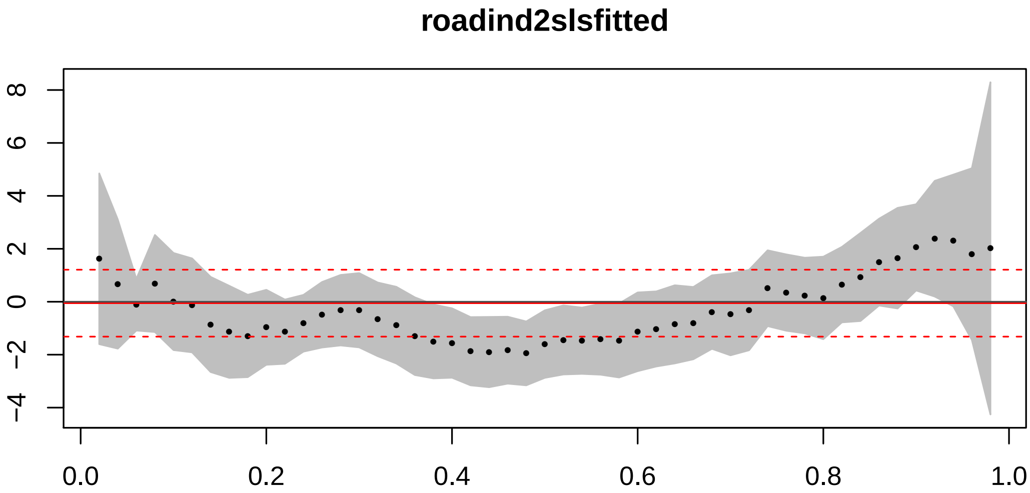


Figure A11: Quintile Treatment Effects by Age



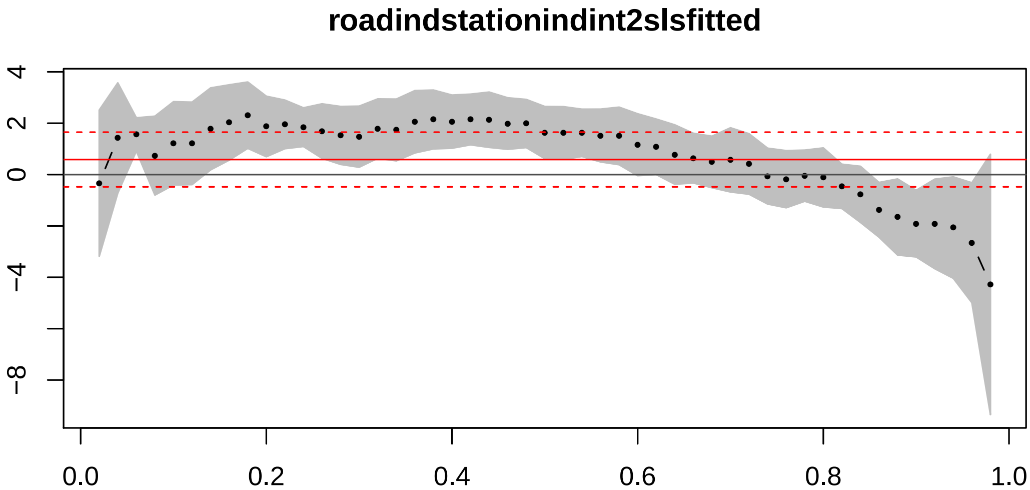


Figure A12: Sectoral Breakdown of Treatments by Gender

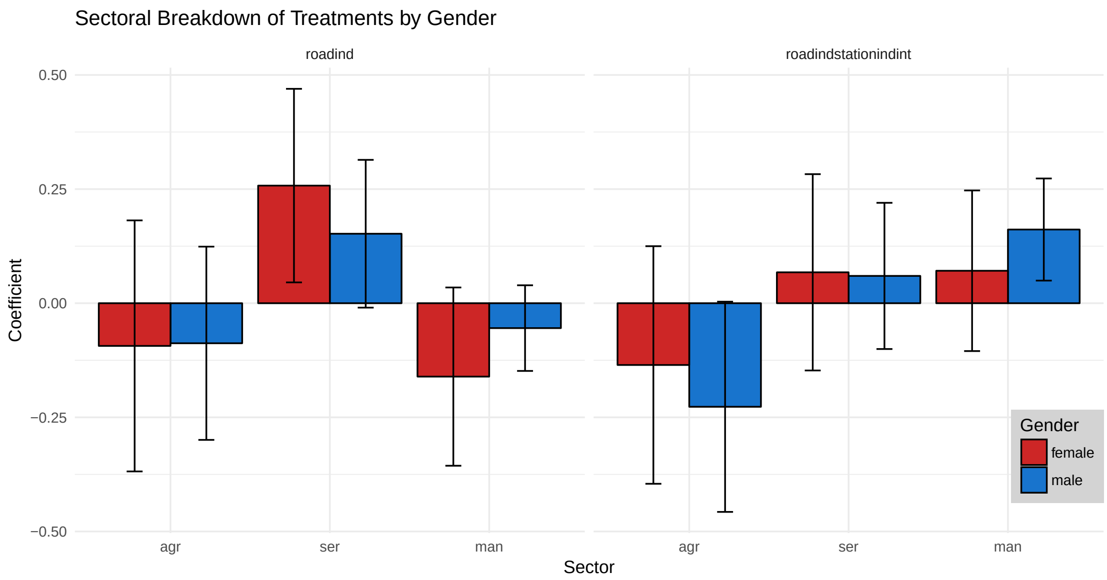
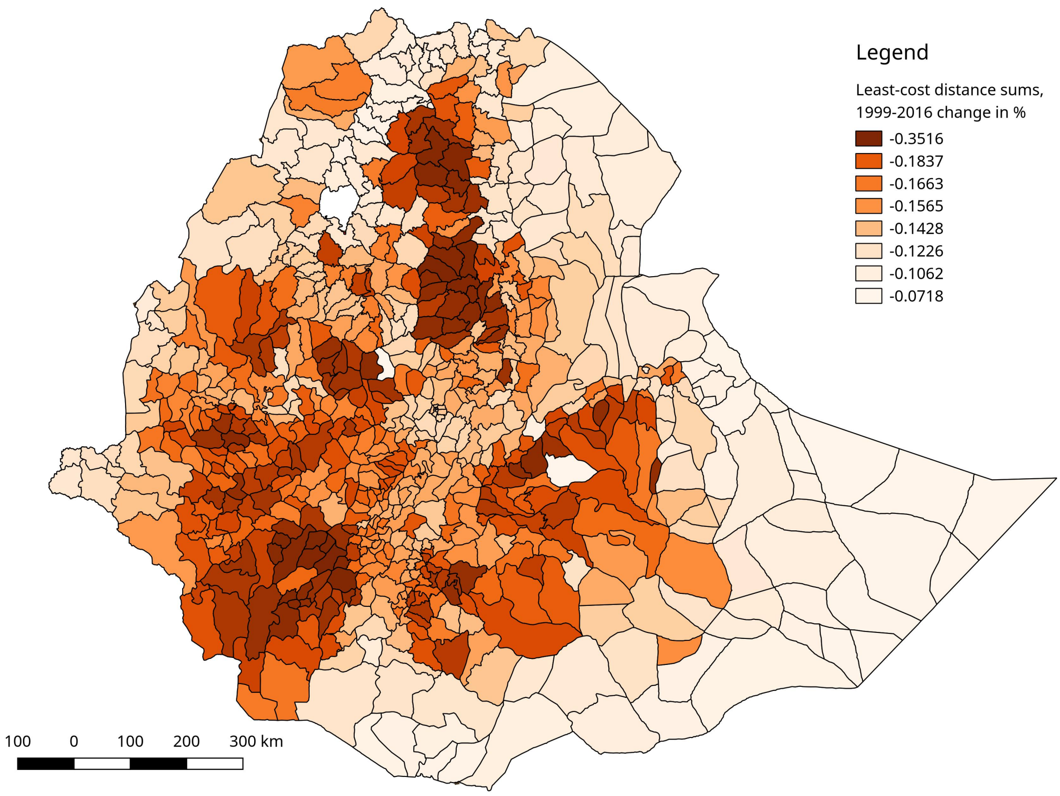


Figure A13: Relative Dijkstra Algorithm Least-cost Distance Changes across Districts, Single Long Difference (1999-2016)



II Appendix: Additional Tables

Table A1: OLS: Occupupational 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 A2: OLS: Occupupational Change (NLFS-excluding Addis/Somali), Roads and Electricity (1999-2013)

	Dependent variable:											
	Agriculture				Services				Manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Road Indicator	-0.033^{**}	-0.032^{**}	-0.037^{**}	-0.036^{**}	0.032^{***}	0.032^{***}	0.035^{***}	0.034^{***}	0.0003	-0.0003	0.002	0.002
	(0.015)	(0.015)	(0.015)	(0.015)	(0.011)	(0.011)	(0.011)	(0.011)	(0.008)	(0.008)	(0.008)	(0.008)
Road*Elec Ind.	-0.146^{***}	-0.145^{***}	-0.098^{***}	-0.096^{***}	0.101^{***}	0.100^{***}	0.067^{***}	0.065^{***}	0.046^{***}	0.046^{***}	0.031^{***}	0.030^{***}
	(0.019)	(0.020)	(0.020)	(0.020)	(0.014)	(0.014)	(0.014)	(0.014)	(0.008)	(0.008)	(0.009)	(0.009)
Model	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
excl. Somali		✓		✓		✓		✓		✓		✓
excl. Addis			✓	✓			✓	✓			✓	✓
Observations	1,208	1,188	1,183	1,163	1,208	1,188	1,183	1,163	1,208	1,188	1,183	1,163
R^2	0.309	0.307	0.226	0.223	0.309	0.306	0.220	0.214	0.221	0.222	0.196	0.196
Adjusted R^2	0.302	0.300	0.219	0.215	0.303	0.299	0.212	0.207	0.214	0.214	0.188	0.188

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

Table A3: OLS: Occupupational 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: Occupupational Change (DHS-R-excluding Addis/Somali), Roads and Electricity (2000-2016)

	Dependent variable:											
	Agriculture				Services				Manufacturing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Road Indicator	0.026	0.020	0.024	0.018	−0.002	0.002	−0.002	0.001	−0.021*	−0.019	−0.019*	−0.016
	(0.021)	(0.021)	(0.021)	(0.021)	(0.016)	(0.016)	(0.017)	(0.016)	(0.011)	(0.011)	(0.011)	(0.011)
Road*Elec Ind.	−0.091***	−0.093***	−0.068***	−0.070***	0.055***	0.056***	0.042**	0.043**	0.036***	0.036***	0.026**	0.026**
	(0.025)	(0.025)	(0.025)	(0.025)	(0.018)	(0.019)	(0.019)	(0.019)	(0.012)	(0.012)	(0.011)	(0.011)
Model	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Year FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
excl. Somali		✓		✓		✓		✓		✓		✓
excl. Addis			✓	✓			✓	✓			✓	✓
Observations	1,039	1,005	999	965	1,039	1,005	999	965	1,039	1,005	999	965
R^2	0.298	0.304	0.164	0.165	0.283	0.292	0.161	0.164	0.128	0.130	0.085	0.085
Adjusted R^2	0.289	0.295	0.153	0.154	0.275	0.284	0.150	0.153	0.118	0.120	0.074	0.073

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

Table A5: OLS: Occupupational 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 A6: OLS: Occupupational 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 A7: Occup. Change (DHS-R), Roads (Kruskal) and Elec. (IVc) (2000-2016)

	Dependent variable:		
	Agriculture	Services	Manufacturing
	(1)	(2)	(3)
Road Indicator	-0.172^{*}	0.252^{***}	-0.102^{**}
	(0.089)	(0.071)	(0.048)
Road*Elec Ind.	-0.238	0.050	0.210^{***}
	(0.152)	(0.123)	(0.062)
Constant	0.844^{***}	0.025	0.134^{***}
	(0.054)	(0.042)	(0.030)
Cragg-Donald F.	6.207	6.207	
Windmeijer cond. F.	23.127	4.433	23.127
District and Year FE	X	X	
Observations	1,039	1,039	1,039
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A8: 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 A9: Full-panel First Stage: Roads-IVs and Elec. IVc., FEs (2000-2016)

	Dependent variable:					
	Roads Ind.	Roads*Elec Ind.	Roads Ind.	Roads*Elec Ind.	Roads Ind.	Roads*Elec Ind.
	(1)	(2)	(3)	(4)	(5)	(6)
Road IV (Kru-It)	0.089^{*}		0.240		0.002	
	(0.046)		(0.213)		(0.023)	
Road IV (Kru-It)*Elec IV	-0.418^{**}		-0.093		-0.392^{***}	
	(0.173)		(0.369)		(0.150)	
Road IV (Bor-It)		0.009				
		(0.014)				
Road IV (Bor-It)*Elec IV		0.058^{**}				
		(0.023)				
Road IV (wRank)				-0.033^{**}		
				(0.015)		
Road IV (wRank)*Elec IV				0.063^{***}		
				(0.023)		
roadivwrankind						-0.007
						(0.013)
roadivwrankindstationivcindint						0.074^{***}
						(0.025)
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	2,756	2,756	2,756	2,756	2,756	2,756
Note:						*p<0.1; **p<0.05; ***p<0.01

Table A10: Occup. Change (NLFS, excl. Somali), Roads (Kruskal) and Elec. (IVc) (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 A11: Occup. Change (NLFS, gender split), Roads (Kruskal) and Elec. (IVc) (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 A12: Demographics (NLFS), Roads (Kruskal) and Elec. (IVc) (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: Education (NLFS), Roads (Kruskal) and Elec. (IVc) (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 A14: Education R/W (NLFS), Roads (Kruskal) and Elec. (IVc) (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 A15: LFP (NLFS), Roads (Kruskal) and Elec. (IVc) (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 A16: Edu. R/W-Mig. (NLFS), Roads (Kruskal) and Elec. (IVc) (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

Table A17: Education Years (NLFS), Roads (Kruskal) and Elec. (IVc) (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