Spatial Integration and Agricultural Productivity: Quantifying the Impact of New Roads †

Tasso Adamopoulos
York University

April 2020

Abstract

I study the effects of Ethiopia's 1996-2014 road expansion program on aggregate and local agricultural productivity and development outcomes. I combine a quantitative spatial framework with a novel district-level panel data set on agricultural production and transport costs. I estimate transport costs between district centers and domestic crop markets accounting for the volume and quality of the road network, and the topography of the terrain. The model features multiple rural locations, where delivering crops to market, as well as accessing intermediate inputs is subject to location-good-specific transport costs. The spatial heterogeneity of transport costs affects the distribution of production and mobile inputs across locations, and the allocation of land across crops within locations. I calibrate the model to the 1996 spatial agricultural production structure of Ethiopia, and then change transport costs alone to their 2014 levels. The model implies a substantial increase of 14.2% in the aggregate real yield, which rises by 20% with the direct resource savings from lower transport costs. These gains account for about 10% of the overall yield gain in the data over 1996-2014. The model also delivers a U-shaped pattern of yield gains across districts with respect to transport cost changes, similar to the one observed in the data. This pattern across districts is attributed to the extent of alignment of districts' changes in absolute and comparative advantage implied by the transport cost changes.

JEL classification: O11, O13, O18, O40, O55, Q10, R12.
Keywords: productivity, agriculture, spatial allocation, transportation costs, road infrastructure.

†This paper supersedes a previous version circulated under the title "Spatial Integration, Agricultural Productivity, and Development: A Quantitative Analysis of Ethiopia's Road Expansion Program." The GIS consultant Mohammed Nuru, and Stein Monteiro provided superb research assistance. I thank Simon Alder, Andrew Dickens, Doug Gollin, Berthold Herrendorf, Gustavo Ventura, Pramilla Krishnan, and Theodore Papageorgiou for helpful comments. I have also benefited from comments at York University, and the SED in St. Louis. I gratefully acknowledge financial support from the International Growth Centre, and thank Alemayehu Seyoum Taffesse for facilitating access to Ethiopia's Agricultural Sample Surveys. All potential errors are my own. Contact: Tasso Adamopoulos, aadamo@yorku.ca.

1 Introduction

A key challenge in the study of growth and development is understanding what factors account for the remarkably low real agricultural productivity in developing countries.[1] Agricultural productivity has important implications not only for developing countries' aggregate income and structural transformation process but also for their welfare and poverty reduction efforts. A characteristic of low income countries is that they have poor transport infrastructure and high internal transportation costs,[2] which operate as a major impediment for farmers in accessing both domestic and international markets. Further, the spatial heterogeneity of transportation costs within developing countries can have implications for what crops are produced, where they are produced and with what inputs, all of which can impact agricultural productivity at the aggregate and local level.

I quantify the aggregate and local productivity effects of a major transport infrastructure intervention that altered the spatial distribution of transport costs. In particular, I study Ethiopia's comprehensive road expansion program over 1997-2014, which resulted in a major overhaul of the country's network in terms of both volume and quality. To quantify the gains from improved market access on agricultural productivity I combine a novel district-level panel data set with a spatial model of agricultural productivity. I construct the district-level panel over 1996-2014 by overlaying agricultural production data with geo-coded transport costs between agricultural production sites and crop markets. The spatial model of agricultural productivity features location-good-specific transport costs, which affect the distribution of production across locations, crop choice, and intermediate input use. I find that at the aggregate level the real yield increases by 14.2%, when only transport costs change from their 1996 to their 2014 level. This accounts for about 1/10 of the actual yield gain over 1996-2014. In addition, the change in the distribution of transport costs leads to a restructuring of the agricultural sector, with a shift in the composition of crops produced

[1] There is a large literature that emphasizes the importance of agriculture for development and cross-country income differences, e.g. Schultz (1953), Gollin et al. (2002), Restuccia et al. (2008), Caselli (2005)

[2] See for example Adamopoulos (2011).

towards cash crops, a drop in agricultural employment, and an increase in average farm size. Similar to the data, I find that local yield gains exhibit a U-shaped pattern with respect to changes in transport costs. I attribute this pattern to the (mis-)alignment of changes in absolute and comparative advantages implied by the transport cost changes, and the associated specialization of districts.

In the late 1990s Ethiopia was a low income country, with its economy heavily skewed towards agriculture, an employment share in agriculture of over 85%, and its agricultural productivity at 55% of its 1960s level, in real terms.[3] At the same time, Ethiopia had very low road network density, and high domestic transport costs.[4] These characteristics were shared by many other developing countries, particularly in Sub-Saharan Africa. Since Ethiopia embarked on its road expansion program in 1997, the entire road network increased 3-fold, while the volume of the rural road network increased 4.7-fold. The proportion of asphalt roads in good condition increased from 17% in 1997 to 73% in 2010. Since then real agricultural productivity has not only rebounded, but surpassed its 1960s levels. This paper measures the contribution of road improvements to the surge in Ethiopia's productivity over 1996-2014, using micro-level data and a structural model.

The micro-level data allow me to construct a district-level panel over 1996-2014, consisting of an agricultural production component and a geographic component. The agricultural production component of the panel draws from repeated waves of household-level data from the Ethiopian Agricultural Sample Surveys, on the type and quantity of crops produced, land allocations by crop, as well as input use. For the geographic component, I estimate travel times between district centers and crop markets, using detailed GIS information on the road infrastructure network at each point in time and high resolution data on the topography of the terrain that has to be travelled to reach the relevant market. I find that in 1996 transport costs are on average very high and exhibit substantial spatial heterogeneity within Ethiopia. By 2014 there has been a considerable drop in the level - 36% on average - and the dispersion of transport costs.

[3] Based on data from the Groningen Growth and Development (GGDC) 10-sector database.

[4] Based on data from Adamopoulos (2011), and World Development Indicators, World Bank.

I then develop a simple spatial equilibrium model featuring an urban center and multiple rural agricultural production locations. Each rural location can produce a food crop for domestic consumption, or a cash crop for the export market. Consumers in the urban location have non-homothetic preferences over the consumption of food, and non-agricultural goods produced in the urban center. Shipments of crops to the urban center for consumption or export are subject to domestic crop-location-specific transportation costs. Transport costs also raise the cost of disbursing imported intermediate inputs from the urban center to the multiple rural locations. The food farming technology also requires labor, as does non-agricultural production in the urban center. Rural locations are heterogeneous along three dimensions: the total amount of agricultural land; the productivities with which they produce crops; the crop-specific transportation costs they face in delivering crops to markets and accessing intermediate inputs. Changes in the distribution of good- and location-specific transportation costs reallocate food production across locations, alter the allocation of land across crops within locations, as well as the distribution of intermediate input use and labor across locations. The model has implications for both aggregate and local outcomes, associated with changes in transport costs.

To isolate the effects of transport cost changes over 1996-2014 on productivity, my empirical approach involves three steps. First, I calibrate the spatial production structure of the model to aggregate and district-level data for the Ethiopian economy for 1996, before the road program began. Second, keeping all else equal, I feed into the model exogenously only the observed changes in transportation costs, implied by my data on the actual changes in the volume and quality of the road network in Ethiopia. Third, I compare the equilibrium changes implied by the model with only transport cost changes to the actual changes in the data over the period 1996-2014, in terms of both aggregate statistics and spatial distributional patterns across districts. The quantitative experiment described essentially answers the following question: What would real productivity have been in 1996, if the road infrastructure was that of 2014 rather than 1996?

The model implies a substantial increase of 14.2% in the aggregate economy-wide real yield. This number is 20% higher if the direct resource savings from lower transport costs are taken into account. To appreciate the magnitude of these gains, I note that they account for about 10% of the overall yield gain experienced by Ethiopia over the period 1996-2014. In terms of the mechanism, as transport costs fall overall, food production is increasingly undertaken by relatively more productive rural districts. Given that the demand for food is inelastic, this allows for an overall shift to cash crops in the economy. The labor required for food production falls, generating a structural shift towards non-agriculture, with an associated increase in average farm size. These changes encapsulate the structural transformation of the economy induced by the transport cost changes, and given the size of the agricultural sector imply substantial gains in aggregate income, with real GDP per capita increasing 20.9%.

In terms of local outcomes, I find that the distribution of the gains is uneven across districts. The model, under the 1996-2014 changes in transportation costs delivers a U-shaped pattern of district-level yield gains with respect to food transport costs across districts, a relationship that is also present in the data. In the model, among districts that are completely specialized in food crops, the biggest gains are experienced by those that observe the largest drops in their transport costs. For these districts changes in the level of their food transport costs and their relative food-to-cash transport costs are strongly aligned. Among districts that produce both crops (incompletely specialized) the largest gains are experienced by those with the smallest change in the level of their food transport costs. For these districts, while the level of their food transport costs falls, their relative food-to-cash transport costs tend to increase.

While the model is stylized it is rich enough to capture key aspects of the spatial structure of agricultural production in developing economies. In particular, it allows for "within" location choices, such as crop and intermediate input choices, as well as "across" location production reallocation. In addition, it allows me to treat each district in the data as the unit of observation. As a result, I do not have to rely on parametric assumptions about the distributions from which transportation costs and productivities would be drawn from, for each crop-district pair.

The importance of agriculture for development has been emphasized in the earlier development literature, e.g. Schultz (1953), and in a more recent quantitative macroeconomics literature, which shows that agriculture plays a key role in understanding the large productivity disparities across countries, Gollin et al. (2002), Restuccia et al. (2008), Caselli (2005). Developing countries are much more unproductive in agriculture than in non-agriculture when compared to developed countries, and in addition employ most of their labor in agriculture. An important challenge for policy and academic research alike is to understand why agricultural productivity is so low in developing countries. There are several recent contributions in the macro-development literature that study this question, among others Lagakos and Waugh (2013), Adamopoulos and Restuccia (2014), Gollin et al. (2014), Tombe (2015), Donovan (2018), Adamopoulos and Restuccia (2018). This paper contributes to this literature by studying a distinct factor, the importance of farm connectivity to markets.

A recent literature in macroeconomics shows that internal transport costs matter at the aggregate level for development and the sectoral composition of the economy: Adamopoulos (2011), Herrendorf et al. (2012), Gollin and Rogerson (2014). This paper contributes to this literature by overlaying micro-data on farm production and detailed geo-coded market access data, to evaluate the impact of a particular road expansion program.

This paper relates to a large literature studying the economic impacts of transport infrastructure investments, in the form of roads, highways or railroads. One strand of the literature uses general equilibrium trade or economic geography models to measure the effects of transport infrastructure projects, e.g., Donaldson (2018), Donaldson and Hornbeck (2016), Allen and Arkolakis (2014), Alder (2019), Asturias et al. (2019) among others. A more recent literature studies the welfare impact of changes in the transportation network in a general equilibrium setting, Allen and Arkolakis (2019), Fajgelbaum and Schaal (2019), Felbermayr and Tarasov (2015). None of these papers however focus on agriculture per se. A related literature estimates local effects of transport infrastructure expansion, e.g., Banerjee et al. (2012), Faber (2014) Baum-Snow et al. (2017), Storeygard (2016), and most closely related to this paper in context Asher and Novosad (2020). A key characteristic of this literature is the use of appropriate identification strategies to address the potential endogeneity of the placement of the relevant transport infrastructure, and estimate its causal effects. Brooks and Donovan (2020) use a hybrid approach to show the effect of improved market access in rural Nicaragua.

This paper is most closely related to two notable papers, Costinot and Donaldson (2016) and Sotelo (2019), who also employ multi-region spatial frameworks that link domestic trade frictions with agricultural productivity, and welfare, when factors are allocated on the basis of comparative advantage. In addition, Sotelo (2019) examines the effects of counterfactual changes in the infrastructure policy in Peru. Besides the country contexts, models and calibration approaches being different, in addition I evaluate the effects of the actual changes in the road network of Ethiopia over time, and compare outcomes to the data. My framework also permits a quantification of the implications for average farm size and the restructuring of the agricultural sector.

The paper proceeds as follows. Section 2 describes the road network data, and how they are used to estimate the geo-coded transport costs. In Section 3 I discuss the agricultural production data, and panel assembling. The spatial framework is developed in Section 4. I calibrate the model to aggregate, and district-level moments from the Ethiopian data in Section 5. Section 6 reports the aggregate and distributional effects from the quantitative experiments. I conclude in Section 7

2 Roads Data

Over the last two decades Ethiopia embarked on an extensive road development program, as a pilar of its growth strategy. Starting in 1997, through the implementation of successive Road Sector Development Programmes there has been substantial improvement in the volume and distribution of the road network, as well as the conditions of the existing roads. The more recent Universal Rural Road Access Program, in particular aimed to extend roads that would connect all lower administrative units in rural areas to all-weather roads.

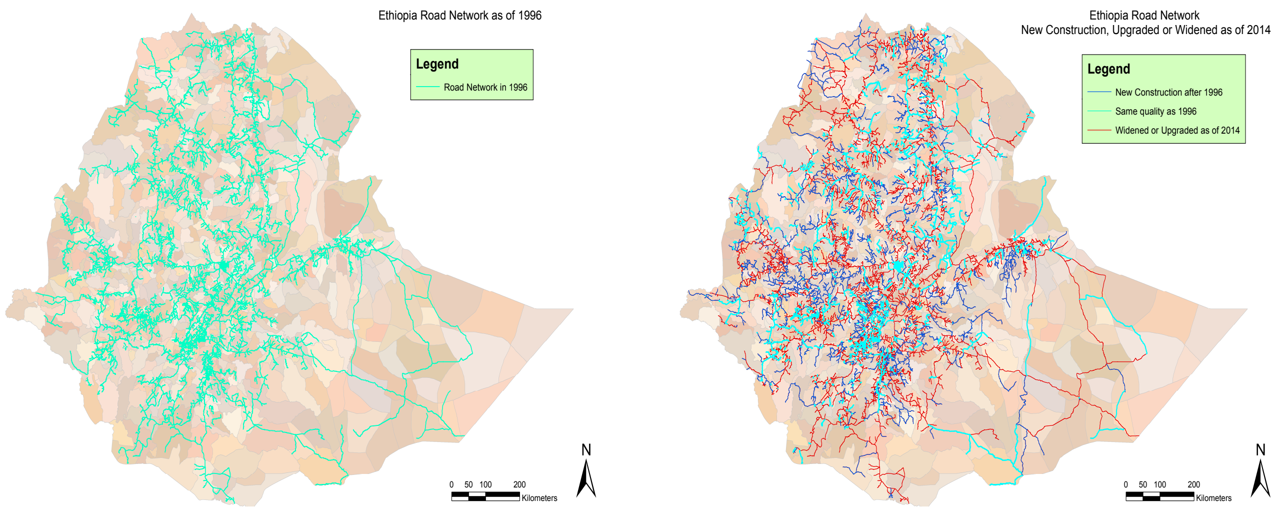
These efforts have had a substantial impact on the extent and quality of the road network in Ethiopia. The volume of the total network increased almost 3-fold, from 24,970 kilometers in 1997 to 69,951 kilometers in 2014. However, the volume increase in the rural road network has been 4.7-fold (from 9,100 in 1997 to 43,094 kilometers in 2014). The road density (including community roads) over 1997-2010 increased from 24 kilometers per 1000 squared kilometers to 136.6 kilometers per 1000 squared kilometers, and from 0.49 kilometers per 1000 people to 1.83 kilometers per 1000 people. In terms of qualitative indicators, the proportion of asphalt roads in good condition increased from 17% in 1997 to 73% in 2010. The proportion of rural roads in good condition increased from 21% to 53% over the same period.

To assess the effect of Ethiopia's road infrastructure expansion program I use detailed GIS data on the universe of roads in Ethiopia, starting in 1996, just before the program began. In particular, the road network data in vector form are obtained from the *Ethiopian Roads Authority* (ERA) for highways and regional roads, and the *Regional Roads Authorities* for regional roads. These data are obtained biennially for the period 1996-2014. The road network data provide information not only on the volume but also on the quality of every link in the network, each year. The data come with information on road class and surface type (e.g., whether a particular road is a highway or a town road, and if a town road whether dirt, or asphalt etc.), year of construction, as well as year of upgrading or rehabilitation.

Panel A in Figure 1 shows the road network in 1996, before the comprehensive road expansion program. Panel B in Figure 1 provides a map of Ethiopia's entire road network as of 2014, indicating both the new links in the network (blue) as well as the links of the pre-1996 network that have been rehabilitated or upgraded by 2014. A casual inspection of the two maps shows a substantial expansion in the volume and quality of the network over the period 1996-2014, especially with respect to feeder roads and roads reaching rural dispersed communities.

Figure 1: Roads in Ethiopia

Panel A: *Before* the Program Panel B: *After* the Program



Notes: With data from the Ethiopian Road Authority (ERA). In Panel B, the network links in blue represent newly constructed roads after 1996; the network links in red represent rehabilitation or quality upgrade of pre-1996 network links.

This roads data is the main ingredient going into the estimation of the geo-coded transportation costs, outlined below.

2.1 Geo-coded Transportation Costs

The goal is to estimate geo-coded transportation costs from agricultural production sites to agricultural markets. The spatial unit of observation is taken to be a district or *woreda*.[5] The main measure of transportation costs I use in the analysis is the travel time in minutes between the district centroids and the nearest destination crop markets. For food crops (cereals), the possible destinations where output can be disbursed are taken to be Ethiopia's 33 major wholesale grain markets (obtained from the Ethiopian Grain Trading Enterprise), which are spread throughout Ethiopia. The food crop travel time for each district is the travel time to the nearest grain market.

[5]Ethiopia is subdivided, in ascending order of disaggregation, into regions, zones, woredas (districts), and kebele (farmer associations).

For cash crops, that are primarily destined for exporting via the capital, the destination market for computing the domestic transportation cost is Addis Ababa.

To estimate a panel of travel times from districts to destination crop markets I overlay the universe of the actual road network data by year described above, with high resolution geographic data on elevation and land use, along with the GPS coordinates of the district centroids and the destination crop markets.[6] The layer of geographic data on land use and elevation is used to obtain as precise geo-coded estimates of travel time as possible by taking into account the topography of the terrain that has to be travelled to reach the relevant market. This captures, the type of land a farmer would have to travel on foot or animal drawn cart before reaching the road, but also accounts for the fact that travel speeds are different on steep roads than on flat surfaced roads.

In order to implement this methodology the entire extent of Ethiopia is formatted into a high resolution grid where the size of a cell (or pixel) in the grid is 250 meters $\times$ 250 meters. All the data are at this fine level of disaggregation in ArcGIS raster files. The travel time (cost) in each cell depends on whether there is a road or not, what type of road there is if a road exists, the type of terrain within the cell if no road exists, and finally the topography (slope) of the terrain. Using Dijkstra's algorithm, I then determine the optimal route for each district center to each destination grain market as the least-accumulative-cost path. The nearest grain market is the one with the lowest cumulative-cost along the set of optimal routes. The measure of geo-coded transport cost for a district is the travel time along the optimal route to the nearest grain market. I note, that the nearest route market is not held fixed over time but is allowed to change in the algorithm. If a different market becomes the nearest one after a given improvement in the road network that involves a particular district, the computed travel time will be the one to the new nearest market. Note that this measure of travel time changes over time as the extent and quality of the network expands.[7]

[6] The topograhical data on slope are from the *United States Geological Survey (USGS)*. The land use data are from the *Geospatial Information System Ethiopia (EthioGIS)*.

[7] The travel time from the centroid better reflects the reality that farmers face, as the center of a district does not

3 Agricultural Production Data

I use household-level data from the Ethiopian Agricultural Sample Survey (AgSS), a nationally representative annual survey administered by the Central Statistical Agency (CSA) in Ethiopia. The data contain information at the field level (a household typically has more than one fields) on what crops are produced, what quantity is produced, how much of the land is allocated to the production of the crop, and information on intermediate input use such as fertilizer. The data I use cover the period from 1995/96 to 2014/15.[8] Given that the AgSS data do not necessarily follow the same households over time and do not contain GPS information on the location of individual households I conduct the analysis at the district (or woreda) level, the lowest level of spatial disaggregation for which a reliable panel could be constructed. See Warner et al. (2015) for a discussion of the challenges involved in assembling a more disaggregate panel.

An issue that arises in merging the AgSS household-level data over the long number of years required for my purposes is that there was redistricting of zones and woredas over time. To address discrepancies of district identifiers that arise from redistricting I homogenize the coding across all years using the 2007 IPUMS zonal and district boundaries and identifiers. While the AgSS waves from 2003/04 and on abide by the IPUMS coding, the earlier years do not. The earlier years were cross checked against IPUMS coding using the names of the districts and zones.

The quantitative analysis focuses on comparing the period before the comprehensive infrastructure program begins (1997) to the end of the period (2014) of the study. In order to have a more representative sample of household observations per district, and to ameliorate any potential noisiness of the household-level data, I pool household data from three years for the earlier period (1995/96,

necessarily fall where a town is located. Nevertheless, I also consider alternative measures of transport costs: the distance from the district centroid to the nearest market through the existing road network; the travel time from the centroid to the nearest market without accounting for terrain and land use; the average distance that can be traveled within an hour from the district centroid given the road network; the travel time from the district capital to the nearest grain market accounting for topography and land use; the travel time to nearest town with population 20, 50, 100, 250 thousand in turn; the travel time to the nearest port. These measures are all highly correlated.

[8] The exceptions are the years 1997/98, 1998/99, 2001/02, and 2002/03 for which data are not available.

1996/97, and 1999/00) and three years for the later period (2012/13, 2013/14, and 2014/15).

The above process allows me to obtain a district-level panel on agricultural production, land allocations across crops, and intermediate input use.[9] The measure of agricultural productivity I focus on at the district-level is the real yield or land productivity, measured as real output per hectare. To construct a real measure of yield over a basket of crops, I aggregate using as a common set of prices across districts, the average prices for each crop over the period 2004-07 in Ethiopia (in local currency units), obtained from the Food and Agricultural Organization (FAOSTAT).

The crops with available output and land data in 1996 and 2014 are all the cereals (barley, maize, millet, oats, rice, sorghum, teff, wheat), and legumes (such as chick peas, dry beans), seeds (such as linseed, sesame, sunflower), spices (such as cardamon, nutmeg), fruit (such as mangoes, papayas, pineapples), vegetables (such as chillies and peppers, garlic, kale), godere, enset, sugar cane, avocados. While coffee has output data at the end of the panel, it does not have output data at the beginning of the panel. Given that the relative price of coffee is high in the price data from FAOSTAT, including coffee only in the later years in the panel, would inflate productivity gains. As a result I exclude coffee from the yield estimates in all periods.

Next, I merge the agricultural productivity data from the AgSS (matched between the pooled 1996 and pooled 2014 periods) with the geo-coded transport cost data, summarized by the travel times from each district centroid to the nearest major grain market, and Addis Ababa. This process allows me to construct a balanced panel of 402 districts with both agricultural production data and transportation cost data between the earlier and later period. Across the districts in the balanced panel, the average yield over all crops across districts increased 4.4-fold, implying an annual average growth rate of 9.7%. Over the same period the yield over grain crops increased 2.5-fold, with an annual average growth rate of 5.9%. This is remarkable growth in real agricultural productivity by any standard. Note, that this growth is not due to price changes since crops have been aggregated using a common set of prices, purging the effects of any possible inflation in prices.

[9]The AgSS data do not report the amount of family or hired labor.

While productivity growth has been ubiquitous across virtually all districts the productivity gains have not been shared equally. Figure 2 shows the histogram of log- growth rates in total real output per hectare across districts. As is clear from the figure, although almost all growth rates are positive there is wide dispersion across districts.

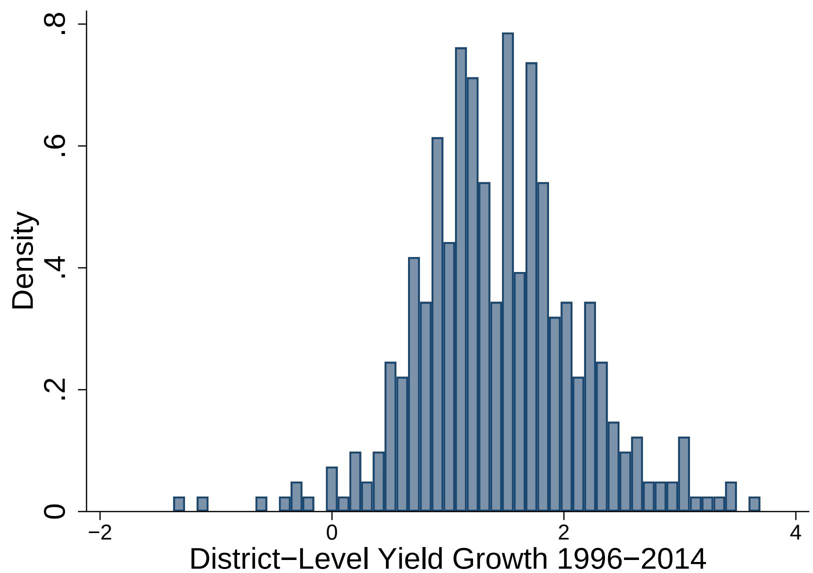


Figure 2: Dispersion of Real Output per Hectare Growth Rates Across Districts

Table 1 presents summary statistics for the estimated geo-coded transport costs. There are two points to note. First, average transport costs from district centroids to grain markets dropped from 345 minutes in 1996 to 220 minutes in 2014, a -36% change. The transport costs from district centroids to Addis Ababa are higher in level, both in 1996 and 2014, dropping by -24% over the period. Second, the dispersion of transport costs across districts dropped, implying better accessibility to markets for more districts. The share of districts within four hours of a major grain market increased from 0.50 in 1996 to 0.71 in 2014. The share of districts within four hours from the capital of Addis Ababa, started from the lower level of 0.11, and increased to 0.20.

Table 1: Summary Statistics of Estimated Transport Costs (Travel Time in Minutes)

	To Nearest Grain Market		To Addis Ababa	
	1996	2014	1996	2014
Mean travel time in minutes	345.4	220.2	575.1	438.9
Median travel time in minutes	241.1	165.0	550.9	419.0
Fraction of Districts < 4 hours	0.498	0.714	0.114	0.204

Source: Author calculations based on estimated district-level geo-coded transportation cost data. Summary statistics are reported for the balanced panel of 402 districts for which both agricultural production (AgSS) data and estimated transport cost data are available. The first two columns report statistics for estimated travel times from the district centroids to the nearest grain markets. The last two columns report statistics for the estimated travel times from the district centroids to the capital of Addis Ababa.

4 A Spatial Model of Agricultural Productivity

I develop a spatial equilibrium two-sector model of agriculture and non-agriculture to assess the effects of changes in transportation costs on agricultural productivity and development, at both the aggregate and local levels. Overall, the model pins down: the allocation of land across crops within locations; the distribution of agricultural production across space on the rural side of the economy; and the distribution of labor across rural locations and sectors. In equilibrium, the economy-wide aggregate measures of interest and their distribution across space are affected by the overall level of transportation costs in the economy, their variation across goods, and the spatial dispersion of these transport costs.

4.1 Environment

Consider a spatial economy with an urban center and a finite number of J rural locations, indexed by $j \in \mathcal{J} \equiv \{1, 2, ...J\}$. The economy produces two agricultural goods, a food crop f, and a cash crop s. In addition, the economy produces a non-agricultural good n. Agricultural production takes place only in the rural locations, while non-agricultural production takes place only in the urban center. Each rural location can produce either of the two crops. The outputs of the same crop across locations are perfect substitutes for each other. The food crop is used only for domestic consumption in the urban center, while the cash crop is fully exported through the urban center.[10] It is assumed that there is unlimited demand abroad at the international price for the cash crop. The non-agricultural good is used only for consumption in the urban center.

Preferences There is a representative household in the urban center with preferences over food and non-agricultural goods,

$$u\left(c_f, c_n\right) = \begin{cases} \overline{f} + \log\left(c_n\right), & \text{if } c_f \geq \overline{f} \\ c_f, & \text{if } c_f < \overline{f}, \end{cases}$$

where $\overline{f}$ is the minimum consumption requirement of food, and c_n is the consumption of the non-agricultural good. These non-homothetic preferences capture Engel's law, whereby when income is low it is fully allocated to the consumption of food but as income rises, and that level of food consumption is achieved, the remaining income is allocated to the consumption of the non-agricultural good. The representative household is endowed with total amount of labor N, that is inelastically supplied to the market. The representative household also owns the productive land in each rural location, L_j. Production of each crop in each location j is undertaken by a representative farm. The farms are also owned by the household, and therefore any profits they make accrue to the household as income.

[10] Qualitatively the results would not change if instead the cash crop was partially consumed domestically. However, because domestic consumption of cash crops is small relative to the domestic consumption of food, as well as the export amount of cash crops, I simplify the model along this dimension.

Production of food crops The food crop in each location j is produced using land, labor and imported intermediate inputs, according to a decreasing returns to scale technology,

$$y_{fj} = \left[z_{fj}^{1-\gamma}\left(n_j^{\alpha}\ell_{fj}^{1-\alpha}\right)^{\gamma}\right]^{\theta} x_j^{1-\theta}, \tag{1}$$

where y_{fj} is output of the food crop, z_{fj} is food crop productivity, and n_j, ℓ_{fj}, x_j are labor, land, and intermediate inputs respectively used in the production of food in location j. In equation (1), $(1-\theta)$ determines the elasticity of final output with respect to intermediate inputs. The object in brackets raised to θ is the production function net of intermediate inputs, with parameter $\gamma < 1$ regulating the extent of returns to scale. Parameter $\alpha < 1$ captures the importance of labor relative to land in food production. Note that decreasing returns to scale imply incomplete specialization and thus the food crop will be produced, at least partly, by every location j.

Production of cash crops The cash crop in each location j is produced according to a constant returns to scale technology that is linear in land,

$$y_{sj} = z_{sj}\ell_{sj},$$

where z_{sj}, y_{sj}, ℓ_{sj} are productivity, output and land under the cash crop technology. The presence of the cash crop technology allows for an alternative use of land, outside food, in rural locations. The decreasing returns to scale in food production and the linearity of the cash crop technology allow to capture the stylized feature of the data that rural locations can be completely specialized in food, but not in cash crops.

Production of non-agricultural good The non-agricultural good is produced by a representative firm in the urban location according to a constant returns to scale technology that is linear in labor,

$$Y_n = AN_n,$$

where A is non-agricultural productivity and N_n is the amount of labor allocated to non-agricultural production.

The total amount of land in location j, L_j, can be allocated to the production of food or cash crops within that location. Labor is used only in the production of food crops within each location and it is perfectly mobile across all rural locations and the urban center.

Goods Prices The non-agricultural good is the numeraire with its price normalized to one. Let p_f be the relative consumer price of the food crop in the urban location, which is endogenous. Note that because food produced in one location is a perfect substitute for food produced in another location, in equilibrium the consumer prices of food in the urban center from different locations will have to be the same and equal to p_f. This small open economy imports all the intermediate inputs from abroad, which are assumed to be inelastically supplied in the international market, and in exchange exports the cash crop it produces. Given that the cash crop is fully exported and the intermediate inputs are fully imported their international prices p_s^*, and p_x^* respectively are taken as given (small open economy assumption). To the extent that there are differences in transportation costs faced by locations in delivering their crops to market, these will show up as differences in the farm-gate (producer) prices for food and cash crops. Similarly, while p_x^* is the price of intermediate inputs upon landing in the urban center, the local prices of intermediate inputs in the different rural locations will differ according to their location-specific transportation costs for delivering intermediate inputs.

Transportation Technology Delivery of crops from each rural location to the urban center for consumption (food crop) or export (cash crop), as well as the delivery of imported intermediate inputs from the urban center to the rural locations is subject to origin-good-specific transportation costs of the iceberg form. In particular, to sell 1 unit of crop $i \in \{f, s\}$ to the urban center, farms in location j have to ship $\tau_{ij} \geq 1$ units of the crop. Similarly, in order for one unit of imported intermediate inputs to arrive in rural location j, τ_{xj} units have to be shipped. Given that the consumer price of food has to be the same in the urban center regardless of origin, the transport technology implies that the farm-gate producer prices of food will differ across locations at origin according to the transport costs involved in delivering their output to the market, p_f/τ_{fj}. Similarly the farm-gate price of cash crops will be p_s^*/τ_{sj} and the farm-gate price of imported intermediate inputs $p_{xj} = p_x^* \tau_{xj}$ in location j. In other words, transport costs reduce the price farms receive for their goods, and raise the prices they pay for their intermediate inputs.

Market Structure All domestic goods and factor markets are perfectly competitive. The economy-wide market clearing condition for food is,

$$c_f = \sum_{j=1}^{J} c_{fj}, \tag{2}$$

where $c_{fj} = \frac{y_{fj}}{\tau_{fj}}$ is the amount of food (consumption) delivered to destination in the urban center originating from location j, and y_{fj} is the amount of the food crop produced and shipped from rural location j. Note that while the only source of demand for food from any location is the consumers of the city center the amount of consumption is not equal to the amount of food produced in each rural location, since part of the output "melts" in transit. So c_{fj} is also the amount of net output of the food crop from location j. Within each location there is barrier μ to the allocation of land between cash and food crops, such that the rental price of land under food crops is a fraction of the rental price of land under cash crops,

$$q_{fj} = (1 - \mu)\, q_{sj}, \tag{3}$$

where q_{ij} is the rental price of land under crop i in location j. The barrier μ is introduced for quantitative purposes, in order to match the ratio of the aggregate yield in cash relative to food crops in the data. The market clearing condition for land in location j is,

$$\ell_{fj} + \ell_{sj} = L_j. \tag{4}$$

The labor market clearing condition requires that the total amount of labor used in all rural locations and the urban location is equal to the total amount of labor in the economy,

$$N_a + N_n = N,$$

where N_a is the total amount of labor devoted to agricultural production across all rural locations,

$$N_a = \sum_{j=1}^{J} n_j.$$

Since the non-agricultural good is produced and consumed in the urban center, the market clearing condition is,

$$Y_n = c_n N.$$

The entire amount of cash crop production from each location j is shipped to the urban center for export, with the export value upon arrival at the urban center being $ex_j = p_s^* \frac{y_{sj}}{\tau_{sj}}$. All intermediate inputs are imported, with a value upon reaching their destination in each rural location j of $im_j = p_x^* \tau_{xj} x_j$. The small open economy's total exports are $EX = \sum_j ex_j$ and imports are $IM = \sum_j im_j$. The economy's net exports are then given by, $NX = EX - IM$.

To summarize, rural locations are heterogeneous with respect to: (a) crop-location-specific productivities $\{z_{fj}, z_{sj}\}$; (b) the total amount of productive land L_j; and (c) the vector of location-good-specific transportation costs $\{\tau_{fj}, \tau_{sj}, \tau_{xj}\}$.

Definition of equilibrium A competitive equilibrium is a set of prices $\left\{p_f, w, (q_{sj}, q_{fj}, w_j)_{j=1}^{J}\right\}$, an allocation for each food crop farm $\{y_{fj}, \ell_{fj}, n_j, x_j\}$ and each cash crop farm $\{y_{sj}, \ell_{sj}\}$ in location j, an allocation for the non-agricultural firm $\{Y_n, N_n\}$, a consumption allocation $\{c_f, c_n\}$, such that: (a) the consumption allocation for urban consumers $\{c_f, c_n\}$ maximizes their utility subject to their budget constraint, given prices and the land allocation barrier μ; (b) the production allocation for each food crop farm in location j, $\{y_{fj}, \ell_{fj}, n_j, x_j\}$ maximizes profits given prices, transportation costs $\{\tau_{fj}, \tau_{xj}\}$, and land L_j; (c) the production allocation for each cash crop farm in location j, $\{y_{sj}, \ell_{sj}\}$ maximizes profits given prices and transportation cost τ_{sj}; (d) the non-agricultural production allocation $\{Y_n, N_n\}$ maximizes the profits of the non-agricultural representative firm, given prices; and (e) the markets for labor, land, food crops, and non-agricultural goods clear.

4.2 Analysis

The profit maximization problem of the food crop farm in rural location j is given by,

$$\max_{\{n_j, \ell_{fj}, x_j\}} \left\{\pi_j = \frac{p_f}{\tau_{fj}} \left[z_{fj}^{1-\gamma} \left(n_j^{\alpha} \ell_{fj}^{1-\alpha}\right)^{\gamma}\right]^{\theta} x_j^{1-\theta} - w_j n_j - q_{fj}\ell_{fj} - p_{xj}x_j\right\},$$

subject to the constraint that the total land allocated to food crop production in a given location cannot exceed the total amount of land in that location, $\ell_{fj} \leq L_j$. The location j wage rate is w_j.[11]

For a given price of food p_f, if the land constraint is binding in j, the optimal choice of land involves a corner solution $\ell_{fj} = L_j$. In this case, rural location j is completely specialized in the production of food. If the solution to the above problem for location j is at an interior optimum then,

$$\ell_{fj} = \left[p_f \gamma \theta \varphi_{fj} \left(\frac{x_j}{y_{fj}}\right)^{\frac{1-\theta}{\theta}}\right]^{\frac{1}{1-\gamma}} \left(\frac{1-\alpha}{q_{fj}}\right)^{\frac{1-\alpha\gamma}{1-\gamma}} \left(\frac{\alpha}{w_j}\right)^{\frac{\alpha\gamma}{1-\gamma}} < L_j, \tag{5}$$

[11]Standard non-linear optimization techniques can be used to solve this problem numerically for every location, given a relative price for food p_f.

and the location is incompletely specialized in the production of food, i.e. produces cash crops as well. I denote by $\varphi_{fj} = z_{fj}^{1-\gamma}/\tau_{fj}$ the "effective" productivity of location j in food crops, adjusting for transport costs.

For every location j, regardless of the extent of specialization, the intensity with which food crop farms apply intermediate inputs depends on the relative cost of intermediate inputs to the producer price of food,

$$\frac{x_j}{y_{fj}} = (1-\theta)\,\frac{p_f}{\tau_{fj} p_{xj}}, \tag{6}$$

and food farm labor demand is a function of the food land input in that location,

$$n_j = \left[\left(\frac{\alpha}{w_j}\right)\gamma\theta p_f \varphi_{fj} \ell_{fj}^{(1-\alpha)\gamma}\left(\frac{x_j}{y_{fj}}\right)^{\frac{1-\theta}{\theta}}\right]^{\frac{1}{1-\alpha\gamma}}. \tag{7}$$

The cash crop farm in each location j solves a simple problem,

$$\max_{\ell_{sj}}\left\{p_s^* \frac{z_{sj}}{\tau_{sj}}\ell_{sj} - q_{sj}\ell_{sj}\right\},$$

where the first order condition pins down the rental price of land in each location j,

$$q_{sj} = \varphi_{sj}, \tag{8}$$

with $\varphi_{sj} \equiv p_s^* \frac{z_{sj}}{\tau_{sj}}$ being "effective" productivity in cash crop production, accounting for iceberg transportation costs, and inclusive of the fixed international price of cash crops.

The profit maximization problem of the non-agricultural firm in the urban center is,

$$\max_{N_n}\left\{AN_n - wN_n\right\},$$

where w is the urban wage rate. The first order condition implies that the wage rate is determined by non-agricultural productivity $w = A$. Given that labor is perfectly mobile across the urban and all rural locations the wage rate in each rural location will be equal to this wage rate, $w_j = w = A$.

Household income consists of labor income, the total return to land from all rural locations and the profits from producing the food crop in each rural location,

$$I = wN + \sum_j (\widehat{q}_{fj} \ell_{fj}) + \sum_j q_{sj} (L_j - \ell_{fj}) + \sum_j \pi_j,$$

where $\widehat{q}_{fj}$ is the adjusted rental cost of food land in location j, which is equal to q_{fj} for incompletely specialized districts (where the land constraint is not binding) and greater than q_{fj} by the marginal benefit of additional food land in districts completely specialized in food (binding land constraint). Given the nature of the preferences the household will consume an amount of food $c_f = \overline{f}$ and allocate the residual income to the consumption of non-agricultural goods.

The relative price of food crops in the urban center, p_f, must clear the market for food (2). Consumers in the urban center consume a fixed amount of food $\overline{f}$. Each rural location produces food y_{fj}, which upon delivery to the urban center is $c_{fj} = y_{fj}/\tau_{fj}$, due to the incurred transport costs. Then the market clearing condition for food crops that implicitly determines p_f is,

$$\overline{f} = \sum_{j \in \mathcal{S}} c_{fj} + \sum_{j \notin \mathcal{S}} c_{fj} \tag{9}$$

where $\mathcal{S}$ is the set of locations completely specialized in food crop production, $\mathcal{S} = \{j \in J : \ell_{fj} = L_j\}$.

Spatial Distribution of Food Production When some locations are completely specialized, and others are incompletely specialized, the relative price of food and the spatial distribution of production cannot be determined analytically. However, for illustrative purposes we can obtain an analytical solution for the relative price of food and the land allocation if all locations were assumed to be incompletely specialized, i.e., there was an interior solution for every j. In this case it can be shown that the equilibrium land allocation in food production in each location j is,

$$\ell_{fj} = \kappa \frac{\left[\frac{\varphi_{fj}}{\varphi_{sj}^{1-\alpha\gamma}}\right]^{\frac{1}{1-\gamma}} \left[\frac{1}{\tau_{fj}\tau_{xj}}\right]^{\frac{1-\theta}{\theta(1-\gamma)}}}{\left[\sum_{k=1}^{J} \left[\frac{\varphi_{fk}}{\varphi_{sk}^{\gamma(1-\alpha)}}\right]^{\frac{1}{1-\gamma}} \left[\frac{1}{\tau_{fk}\tau_{xk}}\right]^{\frac{1-\theta}{\theta(1-\gamma)}}\right]^{\frac{1}{1-\theta(1-\gamma)}}}, \tag{10}$$

where κ is a constant that summarizes parameters of the model. How much of the economy's food a location produces depends on the relative effective productivity and the (inverse) level of transportation costs of that location relative to those of all other locations. To understand the spatial distribution of food production implied by the model consider the ratio of equilibrium land allocations between any two incompletely specialized locations j and k,

$$\frac{\ell_{fj}}{\ell_{fk}} = \left[\frac{\varphi_{fj}/\varphi_{sj}^{1-\alpha\gamma}}{\varphi_{fk}/\varphi_{sk}^{1-\alpha\gamma}}\right]^{\frac{1}{1-\gamma}} \left[\frac{\tau_{fk}\tau_{xk}}{\tau_{fj}\tau_{xj}}\right]^{\frac{1-\theta}{\theta(1-\gamma)}}. \tag{11}$$

According to the first term in equation (11), if location j is relatively more productive in food crops than cash crops in comparison to location k, the relatively more food j will produce compared to k. In other words, comparative advantage in effective productivity matters, which captures comparative advantage in relative actual productivities (z_{ij}'s) as well as comparative advantage in (inverse) relative transport costs (τ_{ij}'s). Locations that face relatively high transport costs in producing food relative to cash crops will allocate less of their land to the production of food crops. The second term in equation (11) indicates that the inverse ratio in the levels of food and intermediate good transport costs is also relevant. The channel through which the levels of transport costs matter is the intermediate input use, as higher transport costs deter the use of intermediate inputs. Note that, according to equation (11), a uniform drop in transport costs across goods and locations will not alter the relative land allocation across rural areas.

Outcome measures The spatial distribution of good-specific transport costs impacts several metrics of productivity at both the local and aggregate level. In particular, transport costs affect the labor-land ratio in food production across incompletely specialized locations,

$$\frac{n_j/\ell_{fj}}{n_k/\ell_{fk}} = \frac{z_{sj}}{z_{sk}}\frac{\tau_{sk}}{\tau_{sj}}.$$

In equilibrium the intensity with which farmers use intermediate inputs depends on the transport costs that farmers have to pay for delivering their crops to markets τ_{fj} and the transport cost involved in having intermediate inputs delivered to their farm from the urban center τ_{xj},

$$\frac{x_j/y_{fj}}{x_k/y_{fk}} = \frac{\tau_{fk}\tau_{xk}}{\tau_{fj}\tau_{xj}}.$$

The yield for food crops and the total yield in location j are given by,

$$\frac{y_{fj}}{\ell_{fj}} = z_{fj}^{1-\gamma}\left(\frac{n_j}{\ell_{fj}}\right)^{\alpha\gamma}\ell_{fj}^{\gamma-1}\left(\frac{x_j}{y_{fj}}\right)^{\frac{1-\theta}{\theta}},$$

$$\frac{Y_j}{L_j} = \frac{y_{fj}}{\ell_{fj}}\frac{\ell_{fj}}{L_j} + \frac{y_{sj}}{\ell_{sj}}\left(1-\frac{\ell_{fj}}{L_j}\right).$$

Finally note that average farm size for each rural location is the total amount of agricultural land in that location over the total number of (food) farm workers,

$$AFS_j = \frac{L_j}{n_j}.$$

Labor productivity in rural location j is the product of the total yield and land per farmer,

$$\frac{Y_j}{n_j} = \frac{Y_j}{L_j}\frac{L_j}{n_j}.$$

In the next section the model economy is calibrated to 1996 district-level and aggregate data for

Ethiopia, and then the effect of transport infrastructure improvements is assessed through the model by changing only the transportation costs in each district and for each good to their 2014 levels.

5 Calibration

The spatial unit of observation of a "rural location" in the model is a district (or woreda) in the Ethiopian data. This is the most disaggregate level for which a reliable panel of agricultural production and geographic data could be constructed. The strategy is to calibrate the benchmark economy to the Ethiopian district-level and aggregate data for 1996, just before the comprehensive transport infrastructure program was initiated.

The parameters that need to be determined, in order to calibrate the model to match the spatial agricultural production structure of the Ethiopian economy, are: (a) the $J \times 2$ matrix of crop-specific productivities across the different locations $\{z_{fj}, z_{sj}\}_{j=1}^{J}$; (b) the $J \times 3$ matrix of iceberg transportation costs for each of the crops, as well as the intermediate inputs between the different rural locations and the urban center $\{\tau_{fj}, \tau_{sj}, \tau_{xj}\}_{j=1}^{J}$; (c) the $J \times 1$ vector of total agricultural land for each location $\{L_j\}_{j=1}^{J}$; (d) non-agricultural productivity in the urban location A; (e) food crop technology parameters (γ, α, θ); (f) the preference parameter $\overline{f}$; (g) the barrier for allocating land to cash crops μ.

My calibration approach does not rely on parametric assumptions about the distributions from which transportation costs (τ_{ij}) and productivities (z_{ij}) could be drawn from for each crop-location pair. Instead, transportation costs before and after are estimated from geographic measures of travel times using GIS software, and productivities by crop for each rural location are backed out from the model by matching district-level targets in the data. I describe this procedure in detail.

In the data $J = 402$, which includes the districts for which agricultural production data and transport cost data are available in both the 1996 and 2014 periods. The food crop in the model corresponds to cereals in the data, which account for 84% of the land allocation overall in the economy. Cash crops are taken to include all other crops. The beginning and end of the period are 1996 and 2014 using the pooled data for each period, as described in Section 3. The world prices of cash crops p_s^*, and intermediate inputs p_x^* are normalized to one, as they do not vary in the quantitative experiment.

Agricultural land by location The total amount of land for each rural location $\{L_j\}_{j=1}^{J}$ is taken directly from the data to be the sum of agricultural land allocated to any crop, food or cash, across all households for that district in the 1996 agricultural production data (AgSS).

Total labor The total amount of labor in the economy N is taken directly from aggregate data for the Ethiopian economy in 1996, from the Groningen Growth and Development Center (GGDC) 10-sector database (Timmer et al., 2015).

Transportation costs by location and good τ_{ij} In the benchmark economy, transportation costs for the two crops and intermediate inputs are estimated from travel times from district centroids to destination markets within Ethiopia, through the existing road infrastructure network, measured from the geographic analysis for 1996. When estimating transportation costs for food crops (cereals) the travel times used are those from district centroids to the nearest grain market. Note that in the model there is only one agricultural market in the urban center while in the data there are multiple. I use the travel time to the closest grain market as the measure of travel time to the central market in the model. While the regional grain markets in Ethiopia are appropriate for estimating food transport costs they are unlikely to be a good approximation for the costs incurred for selling cash crops and purchasing intermediate inputs. Given that cash crops are primarily exported, and exports run through the capital of Addis Ababa, the transportation cost for cash crops is estimated from the travel time from a district centroid to Addis Ababa. Given that the distribution of intermediate inputs is centralized, the travel time between the district centroid and Addis

Ababa is also used for intermediate inputs. Note that the model has iceberg transport costs, which use up resources, while the data involve travel times. As a result the travel times or "time" costs, associated with transportation, have to be mapped into iceberg transport costs or "goods" costs. To map travel times to iceberg transport costs I posit a transport cost function of the following form,

$$\tau_{ij} = 1 + \psi_i \cdot (tt_{ij})^{\eta},$$

where tt_{ij} is the travel time (in minutes) for good i from rural location j to the market. The parameter η captures the sensitivity of transport costs with respect to travel time, and with $\eta < 1$ the transport cost - travel time relationship is concave. $\psi_i > 0$ is a scale parameter that controls the units, in particular, regulating how far from one the implied transport costs are (with $\psi_i = 0$ there are no iceberg costs and $\tau_{ij} = 1$). Next, I explain how I calibrate the parameters of the transport cost function. Combes and Lafourcade (2005) estimate generalized transport costs for road transportation across French districts and report an elasticity of their transport cost measure with respect to real road distance of 0.8 in 1998. This implies that transport costs rise in a concave fashion with distance. While Combes and Lafourcade (2005) do not report the elasticity with respect to real time, the correlation of real time and real distance in their data is 0.986 in 1998 (Table 5). I impose the same concavity on my transport cost measure with respect to travel time by setting $\eta = 0.8$. Then given the value for η, I calibrate the units parameter ψ_i for crops so that the total amount of resources devoted to transport as a share of consumer value of output in the model matches the share of transportation costs in the sales value of food in the data for Ethiopia in 1996. Based on a survey of grain wholesale traders across grain markets in Ethiopia for 1996, Gabre-Madhin (2001) shows that for grain "exporting" regions 26% of the sale price is accounted for by marketing costs of various kinds and the profit margin of the transporter. Direct transport costs, including road stops, during the transportation of grains accounted for 58% of the overall marketing costs, implying 13.2% of the final sale price is transport. This provides a lower bound on the transport cost share. However, given that some of the other marketing costs such as handling, sacking, storage, commission of brokers, travel cost of transporter, and profit of the transport company can arguably be attributed to "transportation," I target a transport cost share of the final sale price of 18%, which is between the lower bound of 13% and the upper bound of 26%. This implies $\psi_i = 0.00258$ for cereals, which I use for both food and cash crops in the model. The transport cost share of the delivered farm-gate price of fertilizer is higher. Minten et al. (2013) using data from Northwestern Ethiopia show that transportation costs, accounting for "last mile" costs, can raise the effective price of chemical fertilizer by up to 50%. I set $\psi_x = 0.0041$, which implies a share of transport costs in the farm-gate cost of intermediate inputs of 36% to be conservative.

Food crop technology parameters (γ, α, θ) The elasticity of output with respect to intermediate inputs is chosen to match the intermediate input cost share in the value of final output. Aggregate estimates agree that the value of intermediate inputs in gross output for the agricultural sector in Ethiopia is 10% (Prasada Rao, 1993; Ahmed et al., 2009), which implies $\theta = 0.9$. In the food crop production function γ and α regulate the extent of decreasing returns to scale and the income share split between land and labor respectively in the non-intermediate input part of the production function in equation (1). I calibrate γ to match the return to the farm operator's contribution to production and α to the land income share. Given that these are technological parameters, and factor shares for the Ethiopian economy may be distorted, I calibrate them to developed economy estimates. In the model economy $(1 - \gamma)$ is the share of profits in agricultural value added. Using data from Statistics Canada on the agricultural value added accounts for Canada, I compute an average share of unincorporated operator returns, and corporate profits in gross value added in agriculture of 19.3% over 1996-2000.[12] This implies a value for γ of 0.807. Given the value of γ I choose α to match a land income share in agriculture of 18% for the United States (Valentinyi and Herrendorf, 2008). This choice is also consistent with the aggregate factor cost share of capital-land in value added for cereal crops in Ethiopia, which is also 18% (Ahmed et al., 2009). The resulting income share for labor in value added is 63%.

[12] https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3210004801

Barrier to land reallocation The parameter μ that deters the reallocation of land towards cash crops is calibrated so that at the economy-wide level the ratio of average yield for cash crops to average yield for food crops is equal to the value for this statistic from AgSS data in 1996. The cash-food yield ratio is 1.33 in the data.

Rural productivity parameters z_{ij}; urban productivity A For each district j the productivity terms of the food crop and the cash crop technologies $\{z_{fj}, z_{sj}\}$ are chosen to match two targets for that district: (i) the land allocated to food production ℓ_{fj}, and (ii) the actual yield for food crops y_{fj}/ℓ_{fj} in that location (which is equivalent to targeting output y_{fj} since ℓ_{fj} is also targeted). The district-level total yield over food crops aggregates crops using a common set of prices. These targets along with the equilibrium equations are sufficient to recover all the variables of interest in the model. I outline the key steps here.

I first normalize the consumer price of food to 1 in the benchmark economy. Then from the food and cash crop farms' first order conditions with respect to land, and the net land rate equalization across crops within districts (3), the effective productivity parameter for cash crops can be backed-out,

$$\varphi_{sj} = \frac{(1-\alpha)\,\gamma\theta}{(1-\mu)} \frac{p_f}{\tau_{fj}} \frac{y_{fj}}{\ell_{fj}}.$$

The effective productivity term for the food crop technology can be recovered residually from the food crop production function in each location,

$$\varphi_{fj} = \frac{y_{fj}/\tau_{fj}}{\ell_{fj}^{\gamma} \left(\frac{n_j}{\ell_{fj}}\right)^{\alpha\gamma} \left(\frac{x_j}{y_{fj}}\right)^{\frac{1-\theta}{\theta}}},$$

where I have used (7) and (6) to compute the labor-land ratio and the intermediate input intensity for each district. The non-agricultural productivity parameter A, which is calibrated to match a target for the share of labor in agriculture of 86% based on aggregate data for the Ethiopian economy,

Table 2: Calibrated Common Parameters

Parameter	Description	Value
γ	share of land and labor	0.81
α	share of labor	0.78
θ	non-intermediate input share	0.90
μ	barrier to land reallocation	0.86
η	sensitivity of transport costs to travel time	0.80
ψ_f, ψ_s	transport cost scale parameter - crops	0.00258
ψ_x	transport cost scale parameter - fertilizer	0.00410
A	urban non-agricultural productivity	775.4
N	total number of workers	24806
$\overline{f}$	subsistence food consumption (000s)	29315

Table 3: Calibrated District-specific Parameters

Parameter	Description	Target Data (1996)
$\{L_j\}_{j=1}^J$	total agricultural land	total land from AgSS Data
$\{\tau_{fj}\}_{j=1}^J$	food crop iceberg transport cost	travel time to nearest grain market
$\{\tau_{sj}\}_{j=1}^J$	cash crop iceberg transport cost	travel time to Addis Ababa
$\{\tau_{xj}\}_{j=1}^J$	inter. input iceberg transport cost	travel time to Addis Ababa
$\{z_{fj}, z_{sj}\}_{j=1}^J$	productivity parameters by district	food yield and land share from AgSS data

from the GGDC 10-sector database (Timmer et al., 2015). To see this note that $N_a = \sum_j n_j$ can be used to solve for A as,

$$A = \frac{\alpha}{1-\alpha} \frac{1-\mu}{\hat{N}_a} \sum \varphi_{sj} \ell_{fj},$$

where $\hat{N}_a$ is the target for the overall employment in agriculture.

Food consumption requirement $\overline{f}$ Given that food production data and transport costs are targeted explicitly in the calibration as described above, the market clearing condition for food (2) pins down the value of the subsistence food consumption term $\overline{f}$ as the total production of food from all districts net of transport costs.

The economy-wide calibrated parameters of the model, that are common across all locations, along with their descriptions are provided in Table 2. Table 3 provides a description of the location-specific parameters that are mapped into actual district-level data, along with their data targets. The calibrated model does well in replicating aggregate and spatial features of the Ethiopian economy. The values of key variables of interest in the calibrated benchmark economy are provided in the first column of Table 4. However, the model also does well in matching district-level statistics that are not targeted in the calibration. Panel A in Figure 3 compares the district-level yields for cash crops implied by the model against their counterparts in the 1996 pooled AgSS data, that were not explicitly targeted. Panel B in Figure 3 compares the spatial distribution across districts of food farm labor (share in total economy-wide labor engaged in food production) to the 1996 distribution of households engaged in cereal production across districts (as a share of the total households engaged in cereals).[13] There is a strong positive correlation between model and data for both the cash crop yields and the food crop labor shares.

The AgSS data provide reliable information on whether any given field operated by a household uses fertilizer. In addition, the AgSS contains information on the amount of fertilizer applied. However, this data is more sparse and less reliable for time series comparisons. Using the AgSS data I construct a district-level measure of intermediate input use as the share of all fields that have any amount of fertilizer applied to them. On average this has increased from 32% in 1996 to 52% in 2014, indicating a significant increase in the use of fertilizer. In the model, there are no fields, so an intermediate input intensity district-level measure can only be constructed as the share of intermediate inputs in final output in each district. The spatial distribution of district-level intermediate input intensities is not targeted in the calibration. Nevertheless, as Panel C in Figure 3 shows the intensive margin intermediate input measure from the benchmark economy in the model is strongly positively correlated with the extensive margin intermediate input intensity measure from the data, with a correlation coefficient of 0.48.

[13] The AgSS does not provide information on labor. I use the number of households engaged in cereal production as a proxy for food labor in this comparison.

6 Quantitative Experiment

The experiment involves studying the effects from reducing geographic transport costs across all districts from their actual 1996 levels to their actual 2014 levels. In order to isolate the effects of transport cost changes alone I keep all other parameters to their 1996 levels. In other words, I ask what would be the aggregate and spatial micro-level effects on the Ethiopian economy if the only change between 1996 and 2014 had been the change in the transportation network and the associated changes in transportation costs? I then compare these changes to the actual changes in the variables of interest that occurred in the data over the same period. The model allows me to assess directly the effects of transport cost changes irrespective of the other changes that may have occurred over the same period and which may have also contributed to changes in the variables of interest.

The iceberg transport costs for 2014 are obtained from the same transport cost function as above,

$$\tau_{ij,2014} = 1 + \psi_i \cdot (tt_{ij,2014})^{\eta},$$

for $i \in \{f, s, x\}$, where the travel times for 2014, $tt_{ij,2014}$, are the ones estimated in Section 2 from the road infrastructure network present in 2014. The associated changes in transport costs alter the connectivity of districts with markets, and do so in a heterogeneous fashion, since the volume and quality of the road network did not expand for all districts at the same rate. As a result there is a change in both the level and the dispersion of good-specific transport costs across districts. Keeping all other parameters (including productivity) in all rural and urban locations to their benchmark economy levels, I feed the 2014 iceberg costs $\{\tau_{fj,2014}, \tau_{sj,2014}, \tau_{xj,2014}\}_{j=1}^{J}$ into the model and solve for the new equilibrium.

6.1 Aggregate Effects

The aggregate outcomes in the new equilibrium, associated with the 2014 transport costs, are presented in the second column of Table 4, and the percentage changes relative to the 1996 benchmark economy (first column) are presented in the third column. I note that in the new equilibrium I aggregate across crops and goods using a common set of prices before and after the change in transport costs, just as statistical agencies measure "real" changes. The common set of prices I use are the ones from the benchmark economy net of transport costs.

There are substantial aggregate effects when transport costs alone are reduced to their 2014 levels. Productivity increases substantially as captured by several metrics in Table 4. The aggregate economy-wide real yield in agriculture, measured as real value of final output per unit of land, increases by 14.2%. This is achieved through an increase in within-crop real yields, 11.5% for food and 7.1% for cash crops, as well as a reallocation of land from food crop production to cash crop production in the economy overall. The share of land in food production in the economy drops from almost 84% in 1996 to under 72% after the transport cost changes.

The within-crop increases in the yields are achieved through the spatial reallocation of production across districts according to changes in relative comparative advantage implied by the changes in transportation costs. To see this recall that in the model comparative advantage across districts is determined not only by relative actual TFP (z_{ij}) but also by relative transportation costs (τ_{ij}), i.e., by relative "effective" productivity. After the drop in transport costs, food production is undertaken increasingly by relatively more "productive" districts. Given the inelastic demand for food in the country, captured by the subsistence requirement $\overline{f}$, as the economy becomes more productive in producing food, districts do not need to devote as many resources to food production. The land that is freed up from food production is allocated to the production of cash crops. Given that labor is used only in food production, when land allocated to food falls, the amount of labor needed to produce that same amount of food also goes down, being now reallocated to non-agricultural production in the urban center. The share of labor in agriculture in the model drops from 86% in the benchmark economy to 80.7%, a drop of roughly five percentage points. With the drop in the overall engagement in food production, the demand for imported intermediate inputs falls. However, the use of intermediate inputs on the intensive margin increases. The overall share of intermediate inputs in final output for the economy increases by one percentage point. This is because the relative cost of intermediate inputs depends not only on the domestic transport costs of delivering those inputs to districts but also on the relative price of food. While transport costs for intermediate inputs and food fall, the relative price of food also falls by 6%, counteracting part of the transport cost savings depending on the district.

Note that the iceberg transport costs are resource costs that show up in the model as goods "melting" in transit. As a result, part of the output of each crop (food and cash) constitute payments to the transportation sector. The net amount of output of crop i delivered to the urban location (for consumption in the case of food, and for export in the case of cash crops) is y_{ij}/τ_{ij}. An alternative real productivity metric to consider is the real net yield, that nets out transport costs, and thus takes into account the direct resource savings from lower transport costs. The real net yield in the model increases by 18% when transport costs drop to their 2014 levels. Again this is the result of within-crop net yield increases (16.9% for food crops and 13.7% for cash crops) as well as the reallocation of land to cash crops at the economy-wide level. The indirect productivity gains achieved through the mechanism of the model account for 80% of the overall gains ($\log(1.142)/\log(1.181)$), implying that the direct savings from lower transport costs are the residual 20%.

Real value added per unit of land overall in agriculture increases 14.8%. Real value added is computed using a common set of prices not only for crops but also for intermediate inputs. The most common measure of productivity looked at, real value added per worker in agriculture increases by 22%. This takes into account not only changes in the real value added yield but also the induced equilibrium changes in agricultural labor, described above. The economy-wide agricultural farm land per worker, or average farm size in the model, increases by 6.6%. A simple decomposition of the total aggregate gain in value added per worker, as shown in equation (12), reveals that the real value added yield gain accounts for 69% (= log(1.148)/ log(1.223)), while the average farm size gains for the remaining 31% (= log(1.066)/ log(1.223)).

$$\underbrace{\frac{VA_a}{N_a}}_{1.223} = \underbrace{\frac{VA_a}{L}}_{1.148} \cdot \underbrace{\frac{L}{N_a}}_{1.066}, \tag{12}$$

These findings indicate that lower transport costs not only raise productivity in farming but also lead to a restructuring of the agricultural sector, characterized by a shift of production towards more export oriented cash crops, lower employment in agriculture, and larger farm sizes.

Real GDP per worker in the economy, which also takes into account the output of the non-agricultural sector in the urban center, increases by 20.9%. This is a substantial increase due to the heavy reliance of the economy on developments in the agricultural sector.

6.2 Spatial Distribution of Effects

While the economy-wide aggregate gains capture the overall effect of changes in transportation costs because they take into account the gains from the spatial reallocation of production, it is also important to understand the spatial patterns across districts that the changes in transport costs impart. The spatial distributional consequences of the transport cost changes further help shed light on the mechanism of the model.

In Panel A in Figure 4 I plot the change in the log–real total yield by district against the change in log–level of transport costs in food by district. There is a U-shaped pattern across districts between changes in yields and changes in the level of food transport costs. This implies that for one set of districts the smaller the changes in the level of their food transport costs the larger the gains (upward sloping part of the U-curve), while for another set of districts the larger the changes in the level of food transport costs the larger the gains (downward sloping part of the U-curve).

In order to understand what accounts for this relationship note that the U-shaped pattern across districts is also present when examining the change in log–real yield in food crops alone against log–changes in the level of food transport costs, as illustrated in Figure 4, Panel B. This is to be expected as in the model the cash crop production technology is linear in land, and therefore there are no within-district changes to the real yield in cash crops. As a result, the U-shaped pattern of the food crop yield carries over to the district-level total yield.

At a proximate level, the reason for the U-shaped pattern is due to the degree of specialization in food crops: the yield growth-transport cost change relationship is negative among districts that are completely specialized in food production, but positive for the districts that are incompletely specialized in food (i.e., those that also produce cash crops). This can be shown analytically through the lens of the model. The yield growth in food crops between two periods t and $t-1$, for districts completely specialized in food crop production ($\ell_{fj} = L_j$) and incompletely specialized ($\ell_{fj} < L_j$), are respectively,

$$\frac{y_{fj,t}/\ell_{fj,t}}{y_{fj,t-1}/\ell_{fj,t-1}} = \left[\frac{n_{j,t}/\ell_{fj,t}}{n_{j,t-1}/\ell_{fj,t-1}}\right]^{\alpha\gamma} \left[\frac{\ell_{fj,t-1}}{L_j}\right]^{1-\gamma} \left[\frac{x_{j,t}/y_{fj,t}}{x_{j,t-1}/y_{fj,t-1}}\right]^{\frac{1-\theta}{\theta}}, \tag{13}$$

$$\frac{y_{fj,t}/\ell_{fj,t}}{y_{fj,t-1}/\ell_{fj,t-1}} = \left(\frac{p_{f,t}}{p_{f,t-1}}\right)^{-1} \left(\frac{\tau_{fj,t}/\tau_{sj,t}}{\tau_{fj,t-1}/\tau_{sj,t-1}}\right). \tag{14}$$

For districts completely specialized in food the total yield for the district is the yield over food crops. For these districts changes in relative food-to-cash transport costs are not relevant for determining the change in the yield. The only thing that matters is changes in the levels of transport costs, as can be seen from equation (13). The larger the drop in food transport costs τ_{fj} and intermediate input transport costs τ_{xj} the larger the increase in intermediate input intensity, with an elasticity of $\frac{1-\theta}{\theta}$. The larger the drop in cash transport costs τ_{sj} the larger the increase in the labor-land ratio, with an elasticity of $\alpha\gamma$. So the larger the drops in the level of transport costs the larger the yield increase among completely specialized districts, which accounts for the downward sloping part of the

U-curve. Note that for districts that are completely specialized in food production the yield growth depends positively on the initial share of land allocated to food. In other words, districts with an initial comparative advantage in food production either because of a higher relative productivity or lower relative transport costs, would tend to have a higher yield growth, other things equal. Lower transport costs allow these districts to better exploit the initial comparative advantage they have in food crop production.

For districts that are incompletely specialized in food only the relative food-to-cash crop transport costs $\frac{\tau_{fj}}{\tau_{sj}}$ matter for the food yield, as can be seen from equation (14). Note that the level of transport costs also affects the labor-land ratio and intermediate input use for these districts too, but because these two factors impact the yield directly and indirectly through the land allocation, the direct and indirect effects cancel out. As a result, the larger the increase in the relative food transport cost the larger the drop in the share of land allocated to food, and by diminishing returns the larger the increase in the food yield, accounting for the upward sloping part of the U-curve. The total yield increases even more for these districts because land is allocated towards the relatively more productive cash crops.

Which districts completely specialize and which incompletely specialize in food crops? The degree of specialization of a district depends on the extent to which changes in the level of food transport costs τ_{fj} or "absolute advantage," are aligned with changes in the relative food-to-cash transport costs $\frac{\tau_{fj}}{\tau_{sj}}$ or "comparative advantage." While food transport costs drop for all districts, for districts that completely specialize in food, the drops are very large, and they are completely aligned with changes in relative food-to-cash transport costs, i.e., both the level and relative transport costs exhibit large drops and are virtually perfectly correlated. For districts that are incompletely specialized the level of food transport costs (which decreases) and relative food-to-cash transport costs (which tend to increase for these districts) are misaligned, and they are more weakly correlated.

In Figure 5, I plot the within-district change in the share of land allocated to food production, implied by the model after the fall in transport costs, against the ratio of productivities (TFP) between the food and cash crop technologies in 1996 (both in logs). The positive relationship indicates that there was a reallocation of land towards food production, following the transport cost changes, in districts that were relatively more productive in food crops in 1996. As districts become more integrated with markets, and trade becomes freer, there tends to be a spatial reallocation of production according to comparative advantage in inherent productivity.

Changes in Spatial Inequality It is natural to ask, to what extent did the building of new roads mitigate spatial inequality across districts? Before the road expansion program in Ethiopia began, transport costs where high particularly for isolated regions, detached from the capital center of Addis Ababa. In Table 5, Panel A I order districts according to their distance (travel time) from Addis Ababa in 1996, and group them into quintiles of the distribution of distances from Addis Ababa. As seen in the first column, the closest districts (Q1) were on average 3.5 hours away from the capital in 1996, while the furthest districts (Q5) were more than 16.5 hours away on average. The second column shows the average increase in the total yield for the districts in each of the 1996 distance quintiles, implied by the model when all transport costs change to their actual 2014 levels. While all districts benefited from the drops in transport costs, the districts furthest away in 1996 (Q5) exhibited higher yield growth on average, 15.9%, than the closest districts in 1996 (Q1), which had 10.6% growth. The highest growth however was exhibited by the districts in the fourth quintile, the second most distant group from Addis Ababa. These growth differences contributed to the partial convergence of the more distant districts from Addis Ababa to the nearby ones.

Panel B in Table 5 orders and groups districts into quintiles according to the level of their 1996 total yield, with Q1 being the lowest productivity 20% in 1996, and Q5 the highest productivity 20%. The first column displays the average total yields for each group in 1996. The second column shows the average yield growth for each group following the drops in transport costs to their 2014 levels. While the average yield growth was similar across all quintiles, the least productive ones in 1996 (Q1) exhibited higher growth than the most productive in 1996 (Q5). This process again was not entirely monotonic.

In sum, there is some evidence of convergence and a dent in spatial inequality, with more distant and lower productivity districts benefiting more from the new infrastructure.

6.3 Comparison to Data Changes

In the above quantitative experiment the only object changed relative to the benchmark economy was the matrix of good-district-specific transportation costs. It is of interest to see how the changes in the allocations induced by the transport cost changes alone compare to the actual changes observed in the Ethiopian economy over the period 1996-2014.

Table 6 compares the aggregate changes from the model (first column) to the ones in the data (second column) for key variables of interest. The real aggregate yield of final output increases 14.2% in the model, accounting for 9% ($\log(1.142)/\log(4.419)$) of the overall increase in the same metric in the data 341.9%. If however the direct resource savings from the transport cost reductions are included then the aggregate net yield increases by 18.1%, accounting for 11.2% of the overall yield gain in the micro-level agricultural production data. In other words, the model with only transport cost changes can account for about 1/10 of the yield gains in the data. The gross yield in food crops in the model accounts for 11.7% of the one observed in the data.

In terms of other statistics, the drop in the share of land allocated to food crops and the increase in average farm size are in the neighborhood of these changes in the data over the period 1996-2014. The model also accounts for less than half of the drop in the share of labor employed in agriculture. Finally, the model also generates an increase in GDP per worker that is about half of the actual change in the data.

Next, I compare the spatial pattern of the yield gains produced by the model with the ones in the data. Figure 6 compares (log) changes in the total district-level yield in the micro data to (log) changes in the total district-level yield implied by the model, against the actual (log) changes in the level of transport costs over the period 1996-2014. While in the data there is more noise and the magnitudes of the changes in the district-level yields are larger than those produced by the model with only changes in transport costs, the U-shaped pattern of the district-level gains with transport costs changes is present in both the data and the model. In Figure 7, I show the same relationship but only for the districts completely specialized in food crops, i.e., those with a decrease in the *relative* food to cash crop transport costs. For these districts there is a negative relationship between aggregate yield gains and transport cost reductions, implying that on average the districts that gain the most are ones that experience the largest drops in their transport costs.

7 Conclusions

This paper has studied a particular episode of a large-scale infrastructure intervention, undertaken in Ethiopia starting in 1997. To measure the effects of the road expansion program I combined a quantitative spatial framework with novel panel data on agricultural production and transportation costs. I find that the changes in transport costs implied by the expansion of the road network have had a sizable impact on productivity and the structure of the agricultural sector in Ethiopia. The gains in real output per worker are about 1/10 of the overall gains observed in the data. I also find that "closer" markets contribute to a structural transformation of the agricultural sector, with more export-oriented cash crop production, fewer farmers, and higher average farm size as employment shifts to other sectors of the economy. These effects are sizable and in the neighborhood of what occurs in the data.

At the individual district level the gains are not uniform. The model produces a U-shaped relationship between district-level gains and transport cost changes, that is similar in nature to the corresponding one in the data. In particular, the districts that experience the largest yield gains are not necessarily only those that experience the largest drops in the *level* of their transport costs. There are other factors driving heterogeneity in the responses of localities even after controlling for transport costs, such as relative transport costs changes across crops, and the relative initial productivities. The implication is that one should not expect a uniform response across regions to lowering transport costs across the board, in the face of inherent heterogeneity.

There is some evidence of spatial convergence across districts with relatively more distant and lower productivity districts tending to exhibit higher productivity growth following the drop in transportation costs.

For an economy like Ethiopia that is heavily skewed towards agriculture, any productivity gains in this sector will translate to aggregate productivity benefits. This is a characteristic shared by many other developing countries, particularly in Sub-Saharan Africa. I note that while the drops in transport costs have been large, Ethiopia started from a very high base, and their level as well as dispersion still remain high. The implication of the analysis here is that, further investments in infrastructure expansion, can have real productivity benefits for the economy. Finally, I note that the analysis has focused on quantifying the effects of transport infrastructure improvements for a given (1996) distribution of crop-specific TFPs across districts. If changes in the distribution of transport costs induce changes in the district-crop-specific TFPs, through for example the adoption of modern agricultural technologies and mechanization or the adoption of high yield varieties, then the gains from transport infrastructure improvements can be potentially larger. A quantification of such gains would be a fruitful avenue for future research.

References

Adamopoulos, T. (2011). Transportation costs, agricultural productivity, and cross-country income differences. *International Economic Review*, 52(2):489–521.

Adamopoulos, T. and Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, 104(6):1667–97.

Adamopoulos, T. and Restuccia, D. (2018). Geography and agricultural productivity: Cross-country evidence from micro plot-level data. Technical report, National Bureau of Economic Research.

Ahmed, H., Amha, M., Amogne, A., Beyene, H., Dorosh, P., Fisseha, E., McDonald, S., Robinson, S., et al. (2009). Ethiopia: Input output table and social accounting matrix.

Alder, S. (2019). Chinese roads in india: The effect of transport infrastructure on economic development. Unpublished manuscript, Univ. North Carolina, Chapel Hill.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Allen, T. and Arkolakis, C. (2019). The welfare effects of transportation infrastructure improvements. *Manuscript, Dartmouth and Yale*.

Asher, S. and Novosad, P. (2020). Rural roads and local economic development. *American Economic Review*, 110(3):797–823.

Asturias, J., García-Santana, M., and Ramos, R. (2019). Competition and the welfare gains from transportation infrastructure: Evidence from the golden quadrilateral of india. *Journal of the European Economic Association*, 17(6):1881–1940.

Banerjee, A., Duflo, E., and Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in china. Technical report, National Bureau of Economic Research.

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

Brooks, W. and Donovan, K. (2020). Eliminating uncertainty in market access: The impact of new bridges in rural nicaragua. *Econometrica, Forthcoming*.

Caselli, F. (2005). Accounting for cross-country income differences. *Handbook of economic growth*, 1:679–741.

Combes, P.-P. and Lafourcade, M. (2005). Transport costs: Measures, determinants, and regional policy implications for france. *Journal of Economic Geography*, 5(3):319–349.

Costinot, A. and Donaldson, D. (2016). How large are the gains from economic integration? theory and evidence from us agriculture, 1880-1997. Technical report, National Bureau of Economic Research.

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

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A market access approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Donovan, K. (2018). Agricultural risk, intermediate inputs, and cross-country productivity differences. *Working Paper, Yale University.*

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

Fajgelbaum, P. D. and Schaal, E. (2019). Optimal transport networks in spatial equilibrium. *Econometrica, Forthcoming.*

Felbermayr, G. J. and Tarasov, A. (2015). Trade and the spatial distribution of transport infrastructure.

Gabre-Madhin, E. Z. (2001). Market institutions, transaction costs, and social capital in the ethiopian grain market. Technical report, Research Report 124, International Food Policy Research Institute (IFPRI), Washington D.C.

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). Agricultural Productivity Differences across Countries. *American Economic Review*, 104(5):165–70.

Gollin, D., Parente, S., and Rogerson, R. (2002). The Role of Agriculture in Development. *American Economic Review*, 92(2):160–164.

Gollin, D. and Rogerson, R. (2014). Productivity, transport costs and subsistence agriculture. *Journal of Development Economics*, 107:38–48.

Herrendorf, B., Schmitz, Jr, J. A., and Teixeira, A. (2012). The role of transportation in us economic development: 1840–1860. *International Economic Review*, 53(3):693–716.

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

Minten, B., Koru, B., and Stifel, D. (2013). The last mile (s) in modern input distribution: Pricing, profitability, and adoption. *Agricultural Economics*, 44(6):629–646.

Prasada Rao, D. (1993). International comparisons of agricultural output and productivity. *FAO Economic and Social Development Paper*, (112).

Restuccia, D., Yang, D. T., and Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, 55(2):234–250.

Schultz, T. W. (1953). *The economic organisation of agriculture*. McGraw-Hill.

Sotelo, S. (2019). Domestic trade frictions and agriculture. Journal of Political Economy, Forthcoming.

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

Timmer, M., de Vries, G. J., and de Vries, K. (2015). Patterns of structural change in developing countries. In *Routledge handbook of industry and development*, pages 79–97. Routledge.

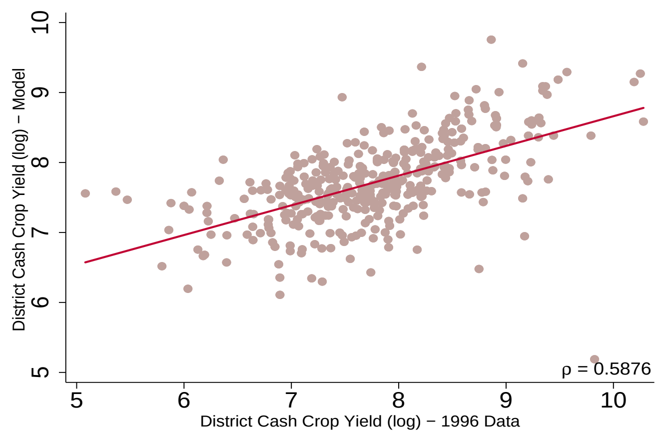
Tombe, T. (2015). The missing food problem: Trade, agriculture, and international productivity differences. *American Economic Journal: Macroeconomics*, 7(3):226–58.

Valentinyi, A. and Herrendorf, B. (2008). Measuring Factor Income Shares at the Sector Level. *Review of Economic Dynamics*, 11(4):820–835.

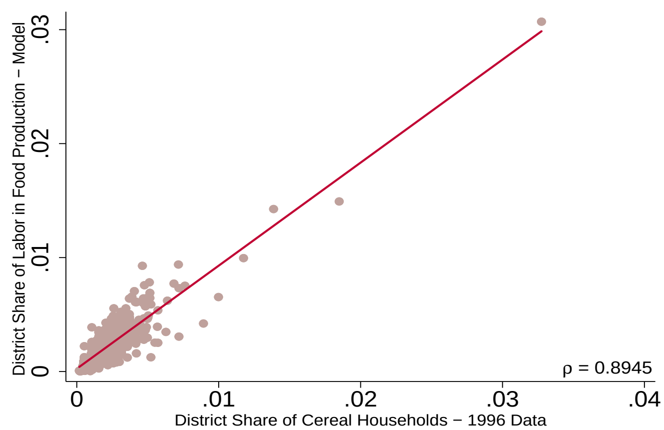
Warner, J., Stehulak, T., and Kasa, L. (2015). Woreda-level crop production rankings in ethiopia: A pooled data approach. *Addis Ababa, Ethiopia*.

Figure 3: Cross-District Comparison of Model to 1996 Data

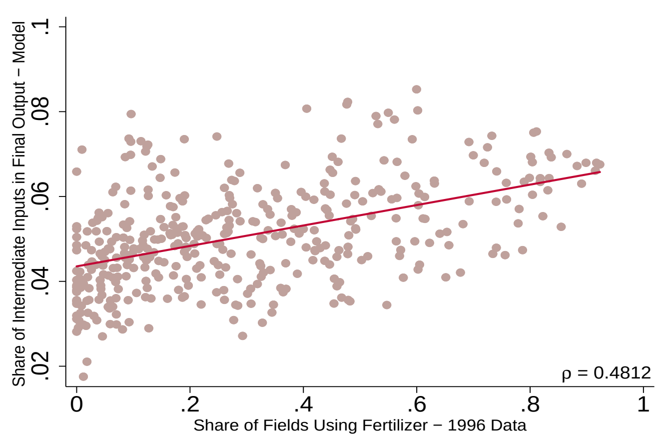
Panel A: Cash Crop Yield



Panel B: Food Farm Labor Share



Panel C: Intermediate Input Use



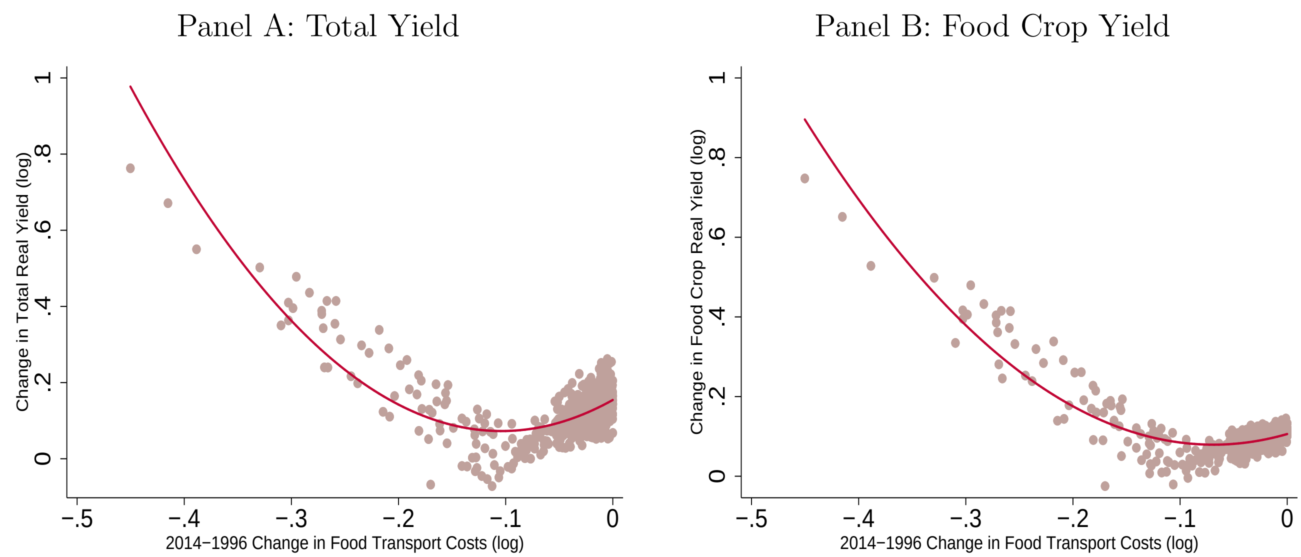
Notes: "Model" refers to the value in the calibrated benchmark economy. "1996 Data" refers to the district value from the 1996 pooled data from the Ethiopia Agricultural Sample Surveys.

Table 4: Effects of Reducing Transport Costs to 2014 Levels

Statistic	Benchmark Economy BE	2014 Transport Costs	Percentage Change (%)
Aggregate Statistics			
Real Aggregate Yield (Y_a/L)	1676.09	1913.78	14.2
Yield in Food Crops (Y_f/L_f)	1591.11	1774.48	11.5
Yield in Cash Crops (Y_s/L_s)	2117.80	2267.26	7.1
Real Net Aggregate Yield (C_a/L)	1337.79	1579.70	18.1
Net Yield in Food Crops (C_f/L_f)	1304.23	1524.86	16.9
Net Yield in Cash Crops (C_s/L_s)	1512.23	1718.86	13.7
Real Value Added Yield (VA_a/L)	1608.25	1845.52	14.8
Real Value Added per worker (VA_a/N_a)	2020.45	2470.92	22.3
Share of Employment in Agriculture (N_a/N) (%)	0.86	0.81	-5.3
Total Share of land in food (L_f/L) (%)	0.84	0.72	-12.1
Intermediate Input Intensity (X_f/Y_f) (%)	0.08	0.09	0.6
Consumer price of food (p_f)	1.00	0.94	-6.2
Average Farm Size (L/N_a)	1.26	1.34	6.6
Real GDP per Worker (GDP/N)	1435.76	1736.12	20.9
District-level Statistics			
STD of log–Food Yield	0.59	0.59	–
STD of log–Aggregate Yield	0.59	0.60	–
CORR of log–(Food Yield, Trans. Costs)	-0.14	-0.08	–
CORR of log–(Aggregate Yield, Trans. Costs)	-0.17	-0.12	–

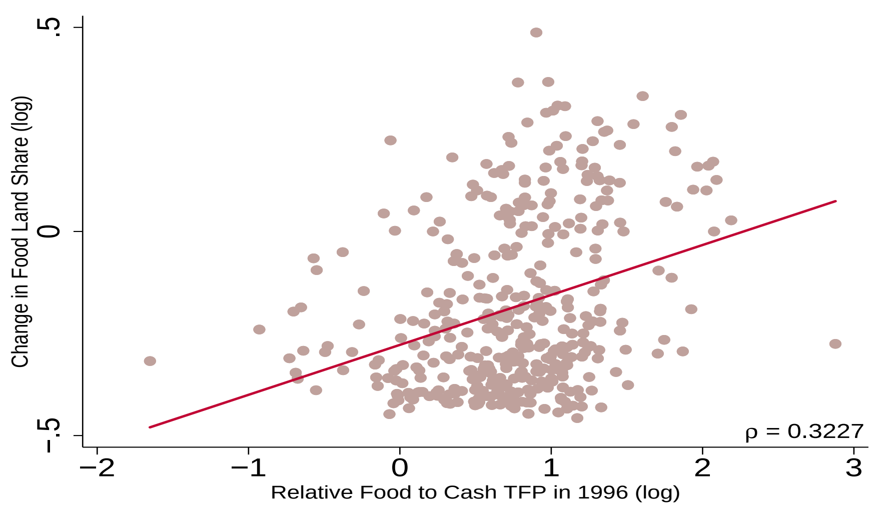
Notes: The column "Benchmark Economy (BE)" displays the values for each variable in the baseline calibrated economy. The column "2014 Transport Costs" displays the values of each variable when transport costs are reduced to their 2014 levels relative to the benchmark economy. The percentage changes in the counterfactual economy (with reduced transport costs) relative to the benchmark are in the last column. All aggregate variables, except for those reported in shares, are reported as the percentage change in the counterfactual relative to the benchmark economy. For variables reported in shares, the last column displays the difference between the pre- and post- transport costs change. "District-level statistics" are reported in levels, and no percentage changes are reported.

Figure 4: Model Changes in Yields with Transport Cost Changes



Notes: "Change in Total Yield" and "Change in Food Crop Yield" refer to the change in the real overall yield (valued at a common set of prices) and food crop yield respectively, in the model, after the reduction in transport costs. The x-axis represents the log-change of food transport costs over 1996-2014.

Figure 5: Change in Food Land Share (model) vs. Relative TFP in 1996



Notes: "Food Land Share" refers to the share of land within each district allocated to food production. The "Change" in this share refers to the change in the model after the reduction in transport costs. "Relative TFP" refers to the ratio of TFP in food production relative to TFP in cash crop production in the benchmark economy.

Table 5: Changes Across Space and Productivity

Panel A: Ordered According to 1996 Distance From Addis Ababa		
Quintile	Travel Time to Addis Ababa (min)	Model % Change in Yield (with transport cost changes)
Q1	210.9	10.6
Q2	406.2	13.6
Q3	549.2	14.6
Q4	713.2	17.8
Q5	996.5	15.9
Panel B: Ordered According to 1996 Yield		
Quintile	Total Yield	Model % Change in Yield (with transport cost changes)
Q1	853	16.8
Q2	1333	13.1
Q3	1719	13.8
Q4	2302	14.8
Q5	4616	14.1

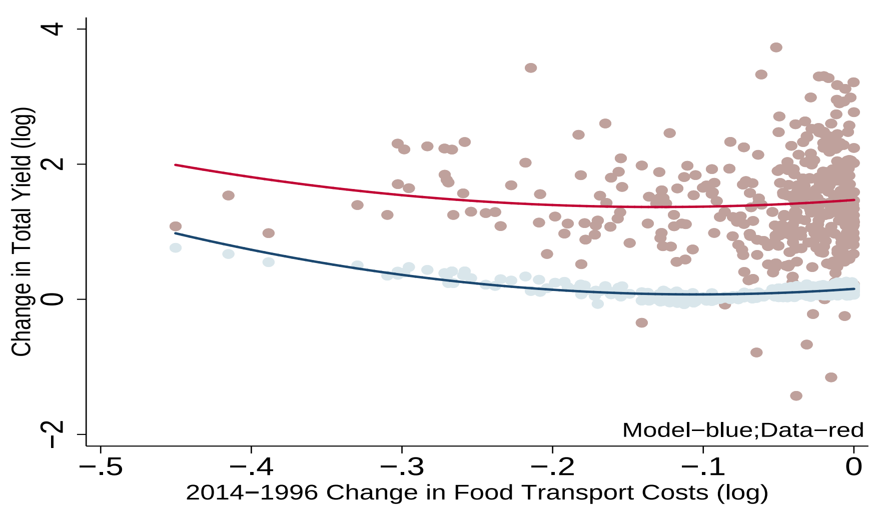
Note: In Panel A, districts are ordered according to their 1996 distance from Addis Ababa, and grouped into quintiles of their distance distribution. In Panel B, districts are ordered according to their total yield in 1996, and grouped into quintiles of the yield distribution. The second column shows the average yield growth rate across districts within each quintile.

Table 6: Comparison of Model and Data Changes (Aggregate Statistics)

Statistic	Changes due to Transport Cost Reductions (%)	Changes in Data Over 1996-2014 (%)
Real Gross Aggregate Yield	14.2	341.9
Real Net Aggregate Yield	18.1	341.9
Real Yield in Food Crops	11.5	153.9
Total Share of land in food (change in %)	-12.1	-8.5
Average Farm Size	6.6	6.2
Share of Employment in Agriculture (change in %)	-5.3	- 12.6
Real GDP per Worker	20.9	67.1

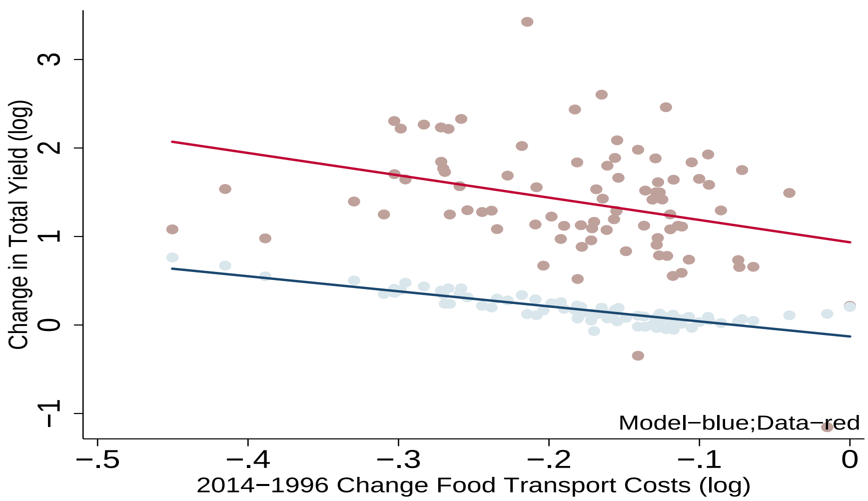
Notes: The first column shows changes relative to the benchmark economy, implied by the model, when all transport costs are reduced to their 2014 levels. All the changes in the data are computed from the Ethiopian Agricultural Sample Surveys data over 1996-2014, with the exception of the "Share of Employment in Agriculture" and "Real GDP per Worker" values which are computed from the GGDC 10-sector database as changes over 1996-2011.

Figure 6: Model vs. Data: Total Yield - Transport Cost Relationship (All Districts)



Notes: "Change in Total Yield" refers to the change in the real overall yield for each district (valued at a common set of prices). In the model this is the change relative to the benchmark economy after the reduction in transport costs. In the AgSS data this is the actual change over 1996-2014. The x-axis represents the log-change of food transport costs over 1996-2014.

Figure 7: Model vs. Data: Total Yield - Transport Cost Relationship (Specialized Districts)



Notes: "Change in Total Yield" refers to the change in the real overall yield (valued at a common set of prices). In the model this is the change relative to the benchmark economy after the reduction in transport costs. In the AgSS data this is the actual change over 1996-2014. The x-axis represents the log-change of food transport costs over 1996-2014.