Online Appendix to “Spatial Integration and Agricultural Productivity: Quantifying the Impact of New Roads”

Tasso Adamopoulos
York University

December 2023

A Criteria for Road Project Selection

The Ethiopian Roads Authority (ERA) had separate specified criteria for earmarking road projects for rehabilitation and new construction (ERA, 2016).

There were five selection criteria for road *upgrading* projects, with the following assigned weights: (i) roads with high traffic (30 percent weight); (ii) roads that would improve the efficiency of the overall road connectivity (20 percent weight); (iii) roads that were in poor condition (20 percent weight); (iv) roads connecting towns with established or emerging industries (10 percent weight); (v) roads linking Ethiopia to ports in other countries - given that Ethiopia is a land locked country, to promote regional integration to international trade routes (20 percent weight).

The five criteria for the preliminary selection of *new* road projects, with corresponding assigned weights were as follows: (i) roads providing access to areas with untapped natural resources (20 percent weight); (ii) roads that would improve connectivity of areas growing surplus food crops as well as cash crops for export (20 percent weight); (iii) roads providing short access roads (missing links) between towns (20 percent weight); (iv) roads providing access to large isolated rural population centres (30 percent weight); (v) roads for equitable regional development (10 percent weight).

The above criteria should be viewed as broad guidelines for earmarking projects. Subsequently, according to the ERA's *Route Selection Manual* (Manual-ERA, 2013), a route selection step would consider the different shortest route options through the nodal points of the earmarked projects, taking into account engineering, topographical, environmental, social, and rough cost considerations. For example, they would try to minimize the ascends and descends in the road in mountainous terrain as well as the number and span of bridges over river crossings or fans. The route selection also avoids areas prone to landslides and flooding as well as areas of natural beauty and ecological conservation (e.g., national parks, forested areas), always adhering to the technical road design standards set out by engineers. Route selection is mainly dealt with centrally by the ERA. Once a preferred route is chosen an economic feasibility analysis is conducted that considers construction and maintenance costs, and socio-economic benefits. Finally, budgetary constraints determine how many of the set of prioritized projects move forward to the construction stage.

B Data Assembly

B.1 Geo-Coded Transportation Costs

To estimate the panel of travel times 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 × 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.[1] The qualitative differences across the different types roads are captured by differences in average travel times across these roads. The speed of driving on different classes of roads is adopted from the *Ethiopian Road Design Manual* and from the input of an expert consultant on the Ethiopian network. To take the effect of slope of the terrain on travel times into account, I use the slope raster data at the $250m \times 250m$ resolution, which provides a slope (grade) value by cell, and assume an elasticity of travel speed with respect to slope of 0.24 (based on road capacity manuals, such as the World Bank's Highway Development and Management model). The above raster files (layers) are combined to create a "friction layer" which helps calculate travel time per meter on any point of the geographic extent of Ethiopia, taking into account the information on road network, land use, and slope. 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, i.e., as the minimum travel time through the "friction layer." 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

[1] 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)*. This data 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.

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 the years as the extent and quality of the network expands.[2]

B.2 Agricultural Production Data

I use household-level data from the Ethiopian Agricultural Sample Survey (CSA, 1995), 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.[3] 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 in Sections IV and V 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, 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 allo-

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

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

cations across crops, and intermediate input use.[4] 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, 2004).

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 403 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 percent. Over the same period the yield over grain crops increased 2.5-fold, with an annual average growth rate of 5.9 percent.[5] 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. While productivity growth has been ubiquitous across virtually all districts the productivity gains have not been shared equally. Figure 1 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.

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

[5]The average economy-wide yield in each year is calculated as the ratio of the sum of (real) value of gross output across all districts divided by the sum of corresponding land (that produces output) across all districts. The gross growth rate is calculated as the ratio of the yield between 2014 and 1996. The growth rates in Figure 1 are calculated similarly at the district level.

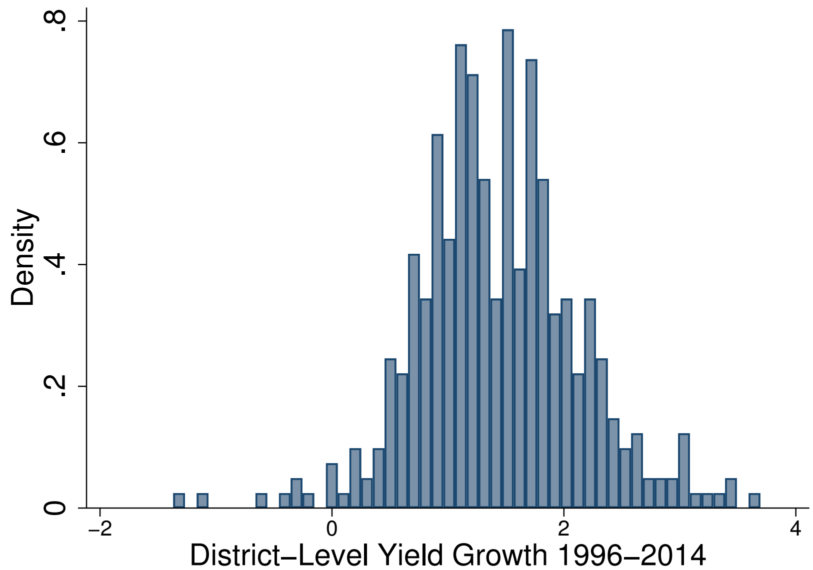

Figure 1: Dispersion of Real Output per Hectare Growth Rates Across Districts

C Difference-in-Differences Estimation

C.1 Pre-treatment Variation

Table C.1 sorts districts into quintiles of the overall change in travel time to nearest grain market over 1996-2014, with $Q1$ denoting the bottom 20 percent and $Q5$ the top 20 percent. There is no systematic pattern across quintiles in 1996 in terms of their distance to the nearest grain market (column 2) or their travel time to Addis Ababa (column 3). In addition, as reported in columns 4-6, there are also no systematic differences in terms of the outcome variables (grain yield, fertilizer use, specialization in grains) in 1996.

Table C.2 shows that there are no strong differences in the timing that districts receive new roads. According to Panel A, the most distant districts from grain markets appear to be treated slightly more first (by 2004) relative to the districts that are closest to the grain markets. Similarly, according to Panel B, by 2004 districts seem to be treated fairly uniformly with respect to their distance from Addis Ababa, with a slight tendency for the furthest districts from Addis Ababa being treated less by 2004, and more by 2012.

Table C.1: Pre-treatment variation by intensity of treatment

	1996 Values (Pre-treatment)				
Quintile of Change in Market Access	Kilometers to Grain Market	Travel Time to Adis Ababa	Grain Yield (log)	Fertilizer Use (No. of fields)	Grain Specialization (Share of Output)
Q1	91.0	620.5	7.45	213	0.84
Q2	73.8	548.0	7.42	351	0.85
Q3	114.7	596.7	7.59	253	0.85
Q4	98.6	542.4	7.43	246	0.83
Q5	84.6	564.8	7.28	304	0.85

Note: To capture the intensity of treatment, districts are ordered according to their overall drop in travel time to nearest grain market over 1992-2014. The first column denotes the quintile of the distribution of overall change in this market access measure, with Q1 being the lowest 20 percent and Q5 the highest 20 percent. Columns 2-3 report the 1996 pre-treatment values in the market access measures that are subsequently affected by the new roads. Columns 4-6 report the 1996 pre-treatment values of the outcome measures in the difference-in-differences specification.

C.2 Difference-in-Differences Robustness

In the baseline empirical difference-in-differences results, the change in district-level market access exploits both the change in transport costs (travel times) and the penetration of transport costs to districts, captured by their initial specialization. Here I estimate the same specification but with a $\Delta Access_i$ that captures only the change in transport costs (travel times to grain markets), but abstracts from differences in their penetration across districts. Table C.3 shows that the results in Table 2 are robust to this alternative measure of market access.

D Estimating the Transport Elasticity

In the model, how differences in effective travel times (that encapsulate differences in road infrastructure) map into differences in transport costs depends crucially on the elasticity of transport costs with respect to effective travel time. Given that this elasticity may vary across countries and may depend on the level of infrastructure development of the country, I estimate it here for the case of Ethiopia. To do so, I use price gaps between regional wholesale grain markets and the capital city of Addis Ababa along with effective travel times between regional markets and Addis Ababa. In

Table C.2: Variation in timing of treatment by pre-treatment market access

Panel A: By 1996 Distance to Grain Market			
	Percent of districts first treated in year:		
Quintile	2004	2008	2012
Q1	14.8	22.7	28.4
Q2	21.9	18.5	16.2
Q3	19.5	24.4	13.5
Q4	21.9	16.8	20.3
Q5	21.9	17.6	21.6

Panel B: By 1996 Travel Time to Addis Ababa			
	Percent of districts first treated in year:		
Quintile	2004	2008	2012
Q1	20.7	23.5	13.5
Q2	24.9	12.6	20.3
Q3	20.7	20.2	14.9
Q4	16.6	27.7	13.5
Q5	17.2	16.0	37.8

Note: Panel A groups districts by quintile of the 1996 kilometers to nearest grain market, while Panel B groups districts by quintile of the 1996 travel time to Addis Ababa. Columns 2-4 report the percentage of districts from a quintile first treated by the year 2004, 2008, 2012 respectively over the total number of districts treated that year.

Table C.3: Estimated Effects of New Roads on Productivity without Penetration

	Dependent Variable (in logs):		
	Yield	Fertilizer Use	Specialization
	(1)	(2)	(3)
$(\Delta Access \cdot NewRoad_{it})$	0.089**	0.145*	0.141**
	(2.27)	(1.93)	(2.15)
Intercept	7.423***	4.815***	-0.013
	(45.75)	(15.67)	(-0.05)
District FE	yes	yes	yes
Year FE	yes	yes	yes
Observations	2299	2274	2299
Adjusted R^2	0.93	0.74	0.58

Note: All columns contain estimates from OLS regressions of district log outcomes on the product of treatment amount (extent of market access) and a time varying dummy for whether the district is treated in a particular period, including time fixed effects, and district fixed effects in each case. The outcome variable is the average yield over all grains in column (1), the number of fields in the district that use fertilizer in column (2), and the share of grains output in total output in column (3). The sample is an balanced panel of districts with six different time periods, 1996, 2004, 2006, 2008, 2010, 2012. t-statistics are in the parentheses, ***, **, and ** represent significance at the 1% ($p < 0.01$), 5% ($p < 0.05$), and 10% ($p < 0.10$) level respectively.

particular, I use wholesale price data in local currency unit (birr) from the *Ethiopian Grain Trade Enterprise* (EGTE) over the period March 2014 - October 2017.[6] I use the reported prices for all grain crops[7] and multiple times for the same crop over the same month for the same market. The EGTE prices are reported for crops that are intended for the international market and as a result have to be traded through the capital of Addis Ababa. To the extent that these crops are actually transported from the regional markets to Addis Ababa, the observed price gaps are a reasonable proxy for transport costs. To capture effective distances between the regional wholesale markets and Addis Ababa, I use my GIS estimated travel times (effective distances) between the districts in which those markets are located and Addis Ababa for the last year in my sample 2014.

Given that the price data start in 2014 (the last year for which I have the infrastructure data), I focus on the cross-sectional spatial variation in prices by computing an average price across all months for each crop in each location, including Addis Ababa. I then compute average price gaps between Addis Ababa and each regional market by crop and estimate the elasticity of price gaps with respect to effective travel time using the following regression,

$$\log\left(\frac{P_{AA,c}}{P_{j,c}} - 1\right) = \beta_0 + \beta_1 time_{AA,j} + \mu_c + \varepsilon_{AA,j,c}$$

where $P_{AA,c}/P_{j,c} - 1$ is the price gap between Addis Ababa and market j, where the -1 is due to the fact that transport costs are greater or equal to 1. The effective travel time is $time_{AA,j}$ and the key elasticity I am after is β_1, with μ_c capturing crop fixed effects since multiple crops are included in the estimation. The results in Table D.4 show that the sensitivity of price gaps to the road infrastructure implied travel time between regional markets and Addis Ababa is 0.79 for all grain crops. In the calibration of the model in Section IV, I use the estimated coefficient of 0.79 for all grains as the elasticity parameter for mapping effective travel times into iceberg transport costs.

E Robustness of Quantitative Results

E.1 Robustness - Parameter Values

In the baseline economy I calibrated the food production elasticity parameters to those in Ethiopia, given that the production technology in the agricultural sector in developing countries could potentially be quite different from that in developed economies. However, one argument against using

[6] http://www.egte-ethiopia.com/en/in-t-market/market-statistics.html (accessed December 10, 2017).

[7] Maize, barley by variety (mixed, white), sorghum by variety (white, red, mixed), teff by variety (red, white, mixed), wheat by variety (mixed, white).

Table D.4: Estimates of the effect of travel time on price gaps in Ethiopia

	All Grains
log-effective travel time ($\widehat{\beta}_1$)	0.79^{*}
Standard error	(0.43)
95% confidence interval	$[-0.06, 1.64]$
Intercept ($\widehat{\beta}_0$)	-5.75^{**}
Standard error	(2.49)
95% confidence interval	$[-10.72, -0.77]$
Number of Observations	71
R-squared	0.18

Estimates are based on an OLS regression of log-price gaps between Addis Ababa and regional markets on log-effective travel time between the regional market and Addis Ababa, using the 2014 road network and accounting for the topography of the terrain. The estimation includes crop fixed effects. Statistical significance at the 10, and 5 percent levels are reflected in the exponents $*$, and $**$ respectively. Standard errors are in parentheses and 95% confidence intervals in square brackets.

factor income shares from developing countries is that the estimated shares are biased due to factor market distortions. While my analysis explicitly considers land market frictions, I still examine the robustness of my results to the factor elasticities. In Table E.5 I show the key results of the main quantitative experiment when I recalibrate to different income shares keeping all other data targets as in the baseline economy. The first row repeats the results for the baseline calibration to Ethiopia.

The second row shows the results when calibrating the income shares to an advanced economy. In the United States the share of intermediate inputs in gross agricultural output is 40 percent, which implies a θ of 0.6. In the model the returns to scale parameter determines the profit share in agriculture. 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 percent over 1996-2000.[8] This implies a value for γ of 0.807. Finally, given the value of γ, I choose α to match a land income share in agriculture of 18 percent for the United States (Valentinyi and Herrendorf, 2008). The structural transformation effects from improving transport infrastructure are more pronounced under this alternative calibration. The share of labor in agriculture drops by 7.7 percentage points (vs. 5.5 in the baseline economy), while

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

Table E.5: Effects of Transport Costs under Alternative Parameterizations

Model	Agr. Empl. Share	Food Land Sh.	Yield	VA per Worker	Inter. Input Inten.	AFS	GDP p.c.
Baseline	-5.5	-11.0	14.7	23.4	0.7	6.8	22.0
Developed Economy Shares	-7.7	-18.2	25.5	43.9	1.0	9.8	44.3
Land share of 0.13	-6.3	-12.4	16.7	27.0	0.6	7.8	24.7
Land share of 0.3	-4.9	-10.1	13.3	20.9	0.8	6.0	20.1
DRS of 0.9	-4.3	-14.9	20.9	28.4	0.8	5.3	27.3
DRS of 0.65	-6.3	-10.6	13.6	23.5	0.6	7.9	21.1

Notes: Each row shows the percentage changes in the counterfactual economy (with reduced transport costs) relative to the benchmark for each model version. For variables reported in shares (agricultural employment share, food land share, intermediate input intensity) the reported results are the absolute change in the share between the pre- and post- transport costs change. "Baseline" refers to the main model in the text. All other rows refer to the main quantitative experiment of the drop in transport costs in recalibrated models with alternative parameterizations.

the yield and labor productivity in agriculture increase by 25.8 and 43.9 percent respectively (in the baseline economy the corresponding changes are 14.7 and 23.4 percent).

The other rows show the sensitivity of the baseline quantitative results to individual parameter values. In the baseline calibration the calibrated land share is 22 percent. I consider alternative land shares of 13 and 30 percent in turn (third and fourth rows in Table E.5).[9] The results are robust, being slightly larger under the smaller land income share and slightly smaller under the higher land income share. The results are also robust to reasonable variations in the extent of returns to scale from the baseline 0.76. When the returns to scale are 0.9 (fifth row), the productivity effects of the 2014 transport costs are larger, while they are slightly lower when the returns to scale are 0.65 (sixth row).

E.2 Robustness - Extensions

I consider two extensions of the main model: (a) intermediate input use in cash crop production; and (b) barriers to the mobility of labor across rural locations.

Intermediate Inputs in Cash Crops In the main model, only food production uses intermediate inputs, while the cash crop production technology is linear in land. Here I allow for intermediate

[9] The 13 percent land share is the lower bound of estimates for Ethiopia in 2004/5 based on national accounts data, see (Bachewe, 2012).

Table E.6: Effects of Transport Costs in Model Extensions

Model	Agr. Empl. Share	Food Land Sh.	Yield	VA per Worker	Inter. Input Inten.	AFS	GDP p.c.
Baseline	-5.5	-11.1	14.7	23.4	0.7	6.8	22.0
Interm. Inputs in Cash							
$\kappa = \theta$	-5.1	-12.0	17.0	23.9	0.7	6.4	21.3
$(1-\kappa) = 2(1-\theta)$	-4.6	-13.2	20.4	24.9	2.8	5.7	20.5
Barriers to Labor Mobility							
Drop in transport costs	-3.2	-11.0	14.7	19.9	0.7	3.8	20.9
+ Drop in Barriers	-5.8	-12.4	16.6	26.0	0.6	7.2	23.8

Notes: Each row shows the percentage changes in the counterfactual economy (with reduced transport costs) relative to the benchmark for each model version. For variables reported in shares (agricultural employment share, food land share, intermediate input intensity) the reported results are the absolute change in the share between the pre- and post- transport costs change. "Baseline" refers to the main model in the text. "Interm. Inputs in Cash" refers to an extension of the main model, in which cash crop production also features intermediate inputs. "Barriers to Labor Mobility" refers to the extension with barriers to the mobility of labor across rural districts.

input use in cash crop production under constant returns to scale. In particular, the production function for cash crops in location j is given by, $y_{sj} = (z_{sj}\ell_{sj})^{\kappa} x_{sj}^{1-\kappa}$, where x_{sj} is the quantity of intermediate inputs in cash crops and $\kappa < 1$ is the intermediate input share. I assume that transport costs for intermediate inputs to a given location are the same for food and cash crops. I recalibrate this extended version of the model, to the same 1996 targets as in the benchmark calibration, under two alternative parameterizations for κ. In the first, I assume that cash crops are as intensive in the use of intermediate inputs as food crops, $(1-\kappa) = (1-\theta) = 0.13$. Given that food production has a low share in the baseline model and cash crops tend to use more intermediate inputs, I also consider a calibration where the cash intermediate input share is double that in food, $(1-\kappa) = 0.26$. Under each calibration, I then run the same quantitative experiment of reducing transport costs to their 2014 levels. With identical intermediate input shares for the two crops, I find that the productivity effects of lower transport costs are slightly larger, with the yield and value added per worker in agriculture increasing by 17 and 23.9 percent respectively (vs. 14.7 and 23.4 in the baseline model). With a higher intermediate input share for cash crops, productivity effects are amplified and the effect on overall intermediate input use is more pronounced (increase of 2.8 percentage points vs. 0.7 in the benchmark economy).

Barriers to Labor Mobility The baseline model features frictions to local land markets within districts, as the key distortion to factor allocation. I consider here an extension of the model, where in addition there are barriers to mobility of labor across rural districts. In particular, the wage rate in location j is now given by, $w_j = \xi_j w$, where ξ_j captures the location-specific tax to labor, and w is the non-agricultural wage rate. I assume that all generated tax revenue is rebated lump-sum to consumers. I recalibrate the model to match the same set of 1996 targets as in the benchmark economy. The labor mobility barriers $\{\xi_j\}_{j=1}^J$ are identified from differences in average farm size across rural districts (given that land is also targeted in the calibration, this is equivalent to targeting the labor distribution across districts). Then I run the same main quantitative experiment of reducing transport costs to their 2014 levels. The results are reported in Table E.6, row "Drop in transport costs." The overall effect on the yield is the same as in the baseline economy (14.7 percent), but the effect on the share of labor in agriculture is slightly lower, at 3.2 percentage point drop (vs. 5.5 in the baseline economy). Intuitively, under imperfect labor mobility, despite the transport costs "shock," workers are not as free to move across space and sectors. As a result, labor productivity increases less than in the baseline economy (19.9 vs. 23.4 percent). It is possible that part of the measured barriers to labor mobility that rationalize the 1996 labor distribution across space, are due to high transport costs. I consider a counterfactual experiment, where improved roads reduce not only transport costs for goods, but also the barriers to labor mobility across space (in proportion to the drop in intermediate input transport costs). The results of this counterfactual are reported in row "+ Drop in Barriers" in Table E.6. When labor barriers drop in addition to goods transport costs, the productivity and structural transformation outcomes are magnified relative to the baseline.

F Other Model Metrics

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 j and k,

$$\frac{n_j/\ell_{fj}}{n_k/\ell_{fk}} = \frac{z_{sj}}{z_{sk}}\frac{\tau_{sk}}{\tau_{sj}}\left(\frac{1-\mu_j}{1-\mu_k}\right).$$

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,

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

G Local Labor Markets

In the baseline model, all consumption takes place in the urban center, by a representative household. I consider an extension of the model with spatially segmented local labor markets, whereby individuals reside and consume in the location that they work in. Let N_j be the total fixed population in location j, that is taken as given. Individuals in each location have the same preferences over food and non-agricultural goods as in the baseline model, $U_j = \overline{f} + \log(c_{nj})$. Their labor is inelastically supplied in the district they reside in, either in agriculture for the production of food, or in non-agriculture. Non-agricultural production in each district takes place according to the same constant returns to scale production function as in the baseline model, $Y_{nj} = AN_{nj}$, but with local labor in each location. Non-agricultural productivity is the same across all locations. Note that while the individuals are immobile across space, they are mobile across sectors within locations. Local labor market clearing, requires that within each location labor to food production and non-agriculture exhausts total local labor, $N_{fj} + N_{nj} = N_j$. Non-agricultural consumption in each location is met either by local non-agricultural production or by imports from other domestic locations, free of transport costs. Land and production units in each location are owned by the consumers in that location. The rest of the model is the same as the baseline. In addition to the aggregate and spatial productivity implications of changes in transport costs, analyzed in the baseline model, this setup also has implications for the separation of sectoral and spatial reallocation, and rural household inequality.

The model is calibrated for 1996 to the same parameters, data targets, and transportation costs as in the baseline model. The distribution of individuals across districts is taken as given from the baseline calibration, consisting of those working in agriculture locally and an equal number of individuals working in non-agriculture in each location. The aggregate statistics of the calibrated model with local labor markets are identical to those in the baseline model. I then run the same main experiment, of reducing all transportation costs across locations and goods to their 2014 levels. The results of this experiment are presented in Table G.7, second row (with the first row repeating the baseline model results). The effects of improvements in transport costs are robust to this alternative formulation, with slightly lower effects on productivity and slightly larger effects on structural change. Given that individuals are now immobile across space, this imposes a limit on how much particularly productive regions can expand their share of food production. However, because individuals can still reallocate to non-agriculture locally, this induces somewhat stronger structural change effects (as seen by the share of employment in agriculture and average farm size). In addition, the lower transport costs lead to lower income inequality across previously most and least distant districts from markets. Relative to the 1996 economy, the income gap between these districts now narrows by 6 percent. The third and fourth row display the effects of reducing transport costs only to Addis Ababa and only to regional grain markets respectively. Roughly, access to Addis Ababa relies more on main national and regional roads, while rural feeder roads play a larger role for access to grain markets. The effects on the yield and rural inequality are roughly of equal magnitude, but the improved roads to Addis Ababa have a larger effect on intermediate input use, while the improved roads to local grain markets have a larger effect on structural change.

Table G.7: Effects of Transport Costs with Local Labor Markets

Model	Agr. Empl. Share	Food Land Sh.	Yield	VA per Worker	Inter. Inp. Inten.	AFS	GDP p.c.	Narrowing Rural Inequ.
Baseline	-5.5	-11.1	14.7	23.4	0.7	6.8	22.0	–
2014 Trans. Costs	-5.8	-11.6	13.7	22.7	0.9	7.3	20.9	0.94
2014 Addis Ababa	0.1	-4.5	7.2	7.1	0.9	-0.1	7.1	0.95
2014 Grain Markets	-5.8	-7.8	7.6	16.1	0.0	7.2	14.2	0.96

Notes: Each row shows the percentage changes in the economy with reduced transport costs relative to 1996. For variables reported in shares (agricultural employment share, food land share, intermediate input intensity) the numbers are the absolute changes in the shares between the pre- and post- transport costs change. "Baseline" refers to the main model. "2014 Addis Ababa" ("2014 Grain Markets") reduces only the transport costs to Addis Ababa (Grain Markets), keeping the costs to grain markets (Addis Ababa) to their 1996 levels. "Narrowing Rural Inequality" is the income gap between the most and least distant (75/25) districts, relative to the 1996 income gap.

References

Bachewe, Fantu Nisrane, "Growth in total factor productivity in the Ethiopian agriculture sector: growth accounting and econometric assessments of sources of growth," *Ethiopia Strategy Support Program II (ESSP II) Working Paper*, 2012, (37).

CSA, "Agricultural Sample Survey, 1995/96-2014/15," *Central Statistical Agency (CSA) of Ethiopia, Addis Ababa, Ethiopia*, 1995.

ERA, *Road Sector Development Program: 19 Years Performance Assessment*, Ethiopian Roads Authority (ERA), Government of Ethiopia., 2016.

FAOSTAT, "Price Statistics," *Food and Agricultural Organization (FAO)*, 2004, (Producer Prices – Annual (LCU/tonne), 2004-2007). http://www.fao.org/faostat/en/#data.

Manual-ERA, *Route Selection Manual*, Ethiopian Roads Authority (ERA), Government of Ethiopia., 2013.

Valentinyi, Akos and Berthold Herrendorf, "Measuring Factor Income Shares at the Sector Level," *Review of Economic Dynamics*, October 2008, *11* (4), 820–835.

Warner, James, Tim Stehulak, and Leulsegged Kasa, "Woreda-Level Crop Production Rankings in Ethiopia: A Pooled Data Approach," *Addis Ababa, Ethiopia*, 2015.