Road Infrastructure and Enterprise Dynamics in Ethiopia*

February 2015

Admasu Shiferaw[(a)], Måns Söderbom[(b)], Eyerusalem Siba[(b)] and Getnet Alemu[(c)]

Abstract:

We analyse the relationship between road infrastructure quality and location choice and entry size of manufacturing firms in Ethiopia. We use GIS based panel data on town-level measures of road infrastructure and census based panel data on firms. Our dataset covers a period of considerable improvements in road infrastructure as a result of major public investments. We find that local infrastructure is important for entry, while more extensive market connectivity is important for the entry of large firms. We conclude that improved infrastructure has been associated with favourable outcomes with respect to entry patterns and firm size in Ethiopia's manufacturing sector.

Key Words: Road Infrastructure, Firm Entry, Location Choice, Startup-Size, Ethiopian Manufacturing.
JEL code: O14, O18, H54.

---

* The authors thank two anonymous referees and participants at seminars organized by the University of Oxford, University of Växjö, ASWEDE, UNU-WIDER and the International Growth Centre for very helpful comments on previous versions of the paper. Financial support from the International Growth Centre is gratefully acknowledged. Admasu Shiferaw is the corresponding author for the paper. His contact details are as follows: The College of William and Mary, Morton Hall, room 131, Williamsburg, VA 23188. Tel: 757-221-2369. Fax: 757-221-1175. Email: ashiferaw@wm.edu.
[(a)] The College of William and Mary, USA.
[(b)] University of Gothenburg, Sweden.
[(c)] Addis Ababa University, Ethiopia

1. Introduction

Poor infrastructure and high transport costs are often identified as key constraints for industrial development in low-income countries (for example Bloom and Sachs, 1998). As noted by Collier (2000), manufacturing firms are intensive users of transport infrastructure services, so if such services are of poor quality, or high cost, manufacturing will be at a comparative disadvantage. Tybout (2000) argues that poor infrastructure is an important reason why markets for manufactured goods in low-income countries are often small and fragmented. In such an environment, firms start and stay small because they target small, localised product markets. These arguments appear to be consistent with the facts observed for Sub-Saharan Africa, where the infrastructure is underdeveloped, and the industrial sector is small and populated primarily by micro and small enterprises supplying local markets. Several studies of advanced economies have documented positive effects of better transport infrastructure on the average number of entrants in a locality[i], but evidence from developing countries is limited and refers almost exclusively to Asian countries.[ii] In this paper we analyse the relationship between improvements in road infrastructure and the entry decisions and entry size of manufacturing firms in Ethiopia.[iii]

Ethiopia is strongly dependent on road infrastructure for its freight and public transport services. In fact, there are few alternatives to road transportation. The country has been landlocked since Eritrea's secession in 1993, has practically no railways except for the very old single-track connecting Addis Ababa and Djibouti, and only a few of its major rivers are navigable. Following a major public investment programme known as the Road Sector Development Programme (RSDP), Ethiopia's road infrastructure has improved considerably since the late 1990s. Data published by Ethiopian Road Authority (2011) indicate that, between 1997 and 2011, the road network expanded from 26,550 km to 53,997 km, while the fraction of roads in good and serviceable conditions increased from 22% to 57%.

In this paper we use GIS based panel data on the road accessibility of Ethiopian towns, and census based panel data on manufacturing firms, to investigate the relationship between road infrastructure and the entry decision and entry size of firms over the period 1996-2008. We use three measures of road infrastructure: the total distance that can be travelled during a 60 minute drive from a particular locality, the total area accessible during those 60 minutes of drive, and the total travel time from a particular locality to major economic destinations. The former two measures primarily capture local improvements in the road infrastructure, while the latter is a more comprehensive measure of how roads affect the connectivity of firms with respect to local as well as distant markets.

Our empirical results indicate that improvements in road access are significantly associated with a town's attractiveness for manufacturing firms. We find a positive and statistically significant association between the quality of the local road infrastructure and the number of firms present in the locality. This result stays robust when we include controls for unobserved town fixed effects, and when we treat road infrastructure as econometrically endogenous using a system GMM approach. In contrast, we find no relationship between the connectivity measure of road infrastructure and the number of firms present in a locality. We analyse the relationship between infrastructure and the size of new entrants using two approaches: in the first we control for unobserved town fixed effects in the firm size model, in the second we use road density in 1990 as an instrument for contemporaneous infrastructure. Results from the fixed effects regressions indicate a strong relationship between connectivity and start-up size. The instrumental variable results are weaker and less conclusive, but consistent with the notion that entry size is more strongly associated with connectivity than with the quality of the local road infrastructure. Taken together, our results provide some insights into what type of infrastructure different firms need: local infrastructure is important to enable more firms to set up; more extensive market connectivity may be important for the entry of large firms. Our results thus add to the accumulating evidence on how the market structure shapes business decisions in Africa. Our findings are consistent, for example, with the argument advanced by Fafchamps and Söderbom (2014) that fragmented markets and high transport costs enable firms to supply local niche markets with little exposure to competitive pressure.

The paper is organised as follows. Section 2 discusses the conceptual and empirical framework of the analyses. Section 3 presents the policy process and the criteria for road placement in Ethiopia as well as our proxies for those criteria. Section 4 describes the panel data on road accessibility and manufacturing firms and provides descriptive statistics. Section 5 contains our econometric analysis. Conclusions and a brief policy discussion are offered in Section 6.

2. Conceptual Framework and Empirical Approach

The premise of our empirical analysis is that, other factors held constant, better road infrastructure reduces the firm's distribution costs, expands its output market, and reduces the costs of sourcing inputs. [iv] In this paper we investigate if better infrastructure is associated with an increase in the number of firms present in a locality, and with larger new entrants. We now discuss our empirical approach for examining these mechanisms.

To investigate if better infrastructure is associated with an increase in the number of firms present in a town we use a model of the following form:

(1) $$\log(N_{it} + 1) = \beta_1 \cdot roads_{it} + \mathbf{X'}_{\boldsymbol{it}}\boldsymbol{\alpha}_{\mathbf{1}} + \theta_i + \gamma_t + u_{it}$$

where N_{it} is the number of firms in town *i* at time *t*, $roads_{it}$ is a measure of the quality of road infrastructure, β_1 is a slope coefficient, $\mathbf{X}_{it}$ is a vector of control variables, $\boldsymbol{\alpha}_1$ is a vector of slope coefficients, θ_i is a town fixed effect, γ_t is a time effect and u_{it} is an error term. The dependent variable is defined as N_{it} plus one, since N_{it} is sometimes equal to zero. This implies that the elasticity of number of firms with respect to *roads* is equal to $\beta_1(N_{it}+1)/N_{it}$. Our sample average of N_{it} is slightly higher than 10, so the adjustment factor $(N_{it}+1)/N_{it}$ is of marginal importance on average and we will abstract from it when discussing the econometric results below. We estimate (1) using town-level data.

OLS estimates of (1) will shed light on whether our measures of road infrastructure quality correlate with the number of firms operating in a town, conditional on the control variables in the model. For this to be a consistent estimator of the causal effect of road infrastructure on firms' entry decisions, the *roads* variable must be orthogonal to the error term in the model. However, the placement of roads is clearly non-random and road infrastructure may therefore correlate with unobservable factors. As discussed in Section 3, the road placement criteria used by the Ethiopian Roads Authority (ERA) are based on a desire to exploit economic potentials of localities or to benefit from favourable initial conditions that could maximize returns on public investment in roads. Moreover, interest groups such as business associations may use their economic clout to demand better road access. Such influential factors on road placement could also lure potential entrepreneurs to open businesses in a particular location, making it difficult to isolate the impact of better road networks from the effect of road placement or factors that drive road placement.

In our empirical analysis we use a fixed effects approach to ensure that our results are robust to endogeneity bias that would otherwise arise if the road infrastructure is correlated with unobserved time constant determinants of entry and exit decisions of firms across towns. We consider results from the within (fixed effects) estimator and a specification estimated in long differences. Under the assumption that $Cov(roads_{it}, u_{is}) = 0$, for all *s*, *t* (strict exogeneity; see Wooldridge, 2010), these are consistent estimators of the causal effect of road infrastructure, even if $Cov(roads_{it}, \theta_i) \neq 0$.

We recognize that the strict exogeneity assumption is potentially strong, however. Road infrastructure investments may be correlated with time varying unobservable factors determining the entry decisions of firms. Moreover, the *roads* variables may be measured with error which would result in a negative correlation between the error term and observed road infrastructure, and attenuation bias in the estimator (Wooldridge, 2010, p.81). To address these concerns we consider results from a system GMM estimator (Blundell and Bond, 1998) which, under certain assumptions, is consistent even if $roads_{it}$ is correlated with the unobservable factors θ_i and u_{it}. This approach involves forming a system of equations consisting of the levels equation (1) and a differenced equation of the form:

$$(2) \qquad \Delta\log(N_{it} + 1) = \beta_1 \cdot \Delta roads_{it} + \Delta\mathbf{X'}_{\boldsymbol{it}}\boldsymbol{\alpha_1} + \gamma_t + \Delta u_{it},$$

and using lagged variables in levels as instruments for the differenced equation (2), and lagged variables expressed in differences as instruments for the levels equation (1). In our most general specification we include a lag of the dependent variable in the set of control variables, resulting in a dynamic panel data model. The purpose of this specification is to capture dynamic agglomeration effects. If such effects are important, one would expect the current entry of firms to spur future entry of firms, generating persistence in the number of firms in the locality conditional on other determinants of location choice.

Estimates of β_1 in the model above shed light on the relationship between net entry and the quality of road infrastructure. The dependent variable in (1) clearly depends on firm exit as well as entry. In order to establish whether better infrastructure specifically triggers a larger inflow of new firms, we also model gross entry separately. In this part of the analysis our outcome variable is the number of start-ups in a town at a particular point in time, which we model as a function of location specific factors including road infrastructure. Because the dependent variable only takes non-negative integer values and the proportion of towns with zero entrants is quite large (approximately 70%), we use a hurdle count data model in which the probability of observing *y* new entrants is specified as:

$$(3) \qquad g(y) = \begin{cases} f_1(0) & \text{if } y = 0 \\ \frac{1-f_1(0)}{1-f_2(0)} f_2(y) & \text{if } y > 0 \end{cases}$$

where $f_1(0)$ is the probability that there are exactly zero new entrants, and $f_2(y)/\left(1 - f_2(0)\right)$ is the probability that there are *y* new entrants, conditional on there being at least one new entrant (see e.g. Winkelmann, 2008). We model positive *y* assuming a negative binomial distribution which corrects for possible overdispersion, and the participation decision using a logit model. Estimation thus involves estimating a truncated negative binomial model for positive entry and a logit model for zero vs. non-zero entry.[v] This is quite a flexible approach since the two processes generating the zeros and the positives are not constrained to be the same. A disadvantage is that it is not feasible to allow for town fixed effects. We therefore include in the specification a set of control variables, to be discussed in Section 3.

The second hypothesis that we wish to test is that there is an association between the size of new entrants and the quality of the road infrastructure. Underlying this hypothesis is the idea that, if demand for manufactures is small and the existing markets are localised due to inadequate infrastructure, optimal firm size is small. As the scope of the market broadens due to better road connectivity, entrants will be larger. We use a model of the following form:

$$\text{(4)} \qquad \log(S_{kit}) = \beta_2 \cdot roads_{it} + \mu_i + \omega_t + v_{kit}$$

where S_{kit} is employment at start-up firm *k* in town *i* at time *t*, β_2 is a slope coefficient to be estimated, μ_i is a town fixed effect, ω_t is a time effect and v_{it} is an error term. If better road infrastructure results in larger new entrants, other factors held constant, the slope parameter β_2 will be positive. Since entry is a one-time event, we have only one observation per firm. Estimation of (4) is thus based on pooled firm-level cross-sectional data. We are able to control for town-level fixed effects, but since there is no panel dimension at the level of the firm we cannot control for firm fixed effects. We investigate the robustness of our results using a two-stage least squares approach in which data on road density in 1990 at the woreda (district) level are used to instrument road accessibility since 1999. Since road density in 1990 is a time constant variable, we cannot control for town fixed effects with this approach. We therefore include in the specification various control variables (see Section 3). For this estimator to be consistent, the district level road density in 1990 may not correlate with the start-up size of firms after 1999 other than through the quality of infrastructure at the time of entry. This means that the 1990 road density must be uncorrelated with unobserved town fixed effects.

3. Road Placement

The Ethiopian government implemented three RSDPs during the period 1997-2010. The estimated total cost of the RSDP during the 14 years was about US$ 7.08 billion, and it was financed partly by various donors including the World Bank, European Union, ADB, NDF, BADEA, OFID, Governments of Japan, Germany, U.K., and Ireland. The first RSDP ran from 1997-2001 and the second from 2002-2007. This is undoubtedly the largest infrastructure development programme in the country's history and probably one of the largest in the region. According to the ERA, the total road network expanded from 26,550 km in 1997 to 53,997 km in 2011 while the fraction of roads in good and serviceable conditions increased from 22% to 57% (Ethiopian Road Authority, 2011). Major activities of the RSDP include rehabilitation of 17 trunk roads, upgrading of 26 trunk and 32 link roads (roads that link trunk roads) and construction of 73 link roads. Table 1 shows additional summary statistics published by the ERA, for 1997 and 2011. For all the infrastructure measures listed in the table we observe considerable improvements over the period.

The ERA applies five criteria for the preliminary selection of new road projects that are proposed by regional states (Appendix A). Priority in road placement is given to areas with high economic potential and surplus food and cash crop production. ERA also takes into account population distribution as well as regional equity in economic development. Road projects that pass the preliminary selection will go through feasibility studies allowing ERA to refine its selection of projects and the proposed budget. Once a five-year plan is approved by government, the number and type of road projects remain intact except for minor adjustments to accommodate unanticipated high priority road projects. The five-year RSDP is implemented through annual action plans. For the assignment of road upgrading projects, ERA follows slightly different criteria. More weight is given for existing roads with high traffic densities and better connectivity with other road networks, both of which are strongly correlated with economic potentials and market size.

Despite having a set of criteria for road placement, it is not clear what specific variables ERA uses to operationalise them. For instance, it is not clear how exactly economic potentials of different geographic locations are assessed or how regional inequality in economic development is evaluated. There is also lack of clarity about the process and criteria by which regional states prioritise their road projects for submission to ERA. It seems that the above mentioned criteria serve as broad guidelines rather than strict rules for road placement. One way of controlling for endogenous road placement would be to include control variables representing the placement criteria. We have obtained woreda (district) level data on population and food self-sufficiency, which can serve this purpose to some extent. The road placement criteria (Appendix A) imply that population will be taken into account in the selection process. Further, since agriculture is the mainstay of the Ethiopian economy, it would seem reasonable to proxy economic potential by the agricultural potential of a location. Agricultural potential is captured by a dummy variable indicating whether the woreda (district) in which a town is located is either food self-sufficient or food surplus. Information on food self-sufficiency is obtained from the Productive Safety-Nets Programme (PSNP) which has been implemented since 2005 by the Ethiopian government and a consortium of donors to provide transfers to the food insecure population in chronically food insecure woredas in a way that prevents asset depletion at the household level and creates assets at the community level (Government of Ethiopia, 2004). Woredas participating in the PSNP are therefore considered to be of low agricultural potential. Unfortunately, data on population and food security are available for one year only. While they still may serve as useful proxies for the entire time period (population and food security status are likely slow changing), they will be redundant in regressions with controls for town fixed effects included. We will use these proxies in regressions in which town fixed effects are not included. In such regressions we will also include region dummies and a variable measuring the average number of manufacturing firms in a town during 1996 to 1998. The latter variable captures initial conditions (including physical and institutional infrastructure) that propagate agglomeration benefits for firms. Such initial conditions, while serving as indicators of economic potential for road placement, would sustain the attractiveness of historical centres of manufacturing for potential entrants. Regressing our road infrastructure variables on the above mentioned control variables, we find that most of the variation in road accessibility is captured by the year dummies and the region fixed effects (results available on request).

Notwithstanding our efforts to address the problems posed by endogenous road placement, we are also reasonably optimistic that the endogeneity problem is not overly serious. As already indicated, RSDPs are rolled out by ERA at a five-year interval and local authorities cannot change the plan after the roll out. Moreover, instead of road project assignments and budget allocations, our regression models use actual improvements in road accessibility at the town level every other year. Measuring improvements in road accessibility at a higher frequency than government's decision on road placement mitigates the latter's feedback with respect to our annual measures of entrepreneurial decisions. The timing of project placement and improvements in road access may also differ because of idiosyncrasies in the implementation strategy and performance of road contractors. Finally, the manufacturing sector is relatively small in Ethiopia, accounting for approximately 5% of GDP. This suggests that infrastructure investment decisions are not strongly dependent on the performance of the manufacturing sector.

4. Data and Descriptive Statistics

4.1 Data

We use accessibility analysis to construct indicators of the quality of road networks measuring travel time and area accessible by road. The accessibility analysis relies on the ‘Network Analysis’ tools of GIS (Geographic Information System) such as Service Coverage and O-D (Origin-Destination) matrix. We prefer these indicators of infrastructure quality to other widely used proxies such as total spending on road projects or the stock of infrastructural capital. One advantage of our measures is that they reflect the type and quality of roads, in addition to the stock of roads. ERA’s 2011 report on the 14 years of RSDP implementation provides project level data on roads that have been rehabilitated, upgraded or newly constructed between 1997 and 2010.[vi] Roads were also identified by their pavement type and condition such as asphalt or gravel roads in order to estimate improvements in travel time. Table B1 in Appendix B shows the expected speed per hour on each type of road which is used as conversion factor for our accessibility analysis. Travel time and area accessible are calculated at the town level every other year from 1996 to 2008, and data for the remaining years are generated through linear interpolation. After discarding observations for which data on the road infrastructure is either incomplete or of poor quality, we obtain a sample for analysis consisting of 90 towns spanning the period 1996-2008.

Service Coverage Analysis

Improvements in travel time and distance at the town level were calculated using the expected travel speed in Table B1. GIS tools allow us to overlay road projects with the location coordinates of towns. This was done for the towns included in the 2007 census of manufacturing firms carried out by the Central Statistical Agency (CSA) in Ethiopia. The GIS analysis uses a 60 minutes cutoff to observe the change in travel time using all roads that serve a town. Two alternative measurements emerge from this exercise. The first captures the total distance traveled during a 60 minute drive from the center of town while the second captures the total road-accessible area during a 60 minute drive. The latter uses a buffer zone (area of influence) of 5 km on both sides of a road. Figure 1 compares the total area accessible during a one hour travel from Addis Ababa in 1996 and 2008.

Origin-Destination Matrix

Origin-Destination (OD) matrix is another GIS tool to determine the impact of road projects on travel time. It measures the travel time from a town in our sample (the origin) to major economic destinations in Ethiopia. Travel time could decline as a result of road projects that do not necessary pass through the town as long as they help improve its connection with major destinations. The O-D matrix based measure of infrastructure quality is therefore more comprehensive than those based on service coverage analysis discussed above, which capture only the impact of road projects in the vicinity of the selected nodes. The OD-matrix uses 15 regional capital cities and other urban centres as major economic destinations (see Table B2, Appendix B). The destinations relevant to a town are decided based on a 10 hour travel time threshold at the beginning of the study period, that is, all destinations that take more than 10 hours in the base year (1996) are excluded.

Data on Manufacturing Firms

The town level panel data on road accessibility are matched with establishment level panel data on Ethiopian manufacturing firms for the period 1996-2008. These data have been collected by the CSA of Ethiopia through annual censuses of manufacturing establishments that employ at least 10 workers. The CSA updates its list of establishments annually in consultation with the federal and regional departments of trade and industry. All establishments are required by law to cooperate with the CSA's data collection efforts. All manufacturing industries except the arms industry are covered by this survey following the same sampling procedure.[vii]

4.2 Descriptive Statistics

Table 2 highlights the extent of the improvements in road networks in and around the towns covered in our dataset due to the RSDP. Column 1 shows the total area that can be accessed by road per hour of travel from the sample towns. This area has expanded on average by about 260 km^2 between 1996 and 2008. Column 2 shows substantial increase in total distance that can be travelled from a town (using all roads serving a town) in 60 minutes, that is, allowing vehicles to travel 46 km further on average in 2008 as compared to 1996. Similarly, Column 3 shows that the average travel time to major economic destinations has declined by about 5 hours per annum during the sample period. Table 2 also shows very little, if any, improvement in road accessibility during the first few years of the RSDP. In fact it is only in 2002 that noticeable changes in road accessibility began to emerge.[viii] It is also interesting to note that while road accessibility improved significantly after 1999, the variation across towns in the quality of local infrastructure increased at the same time. This is shown by the increase in the standard deviation and coefficient of variation in area accessible and travel distance during the latter half of the sample period.

Data from the CSA show that the total number of manufacturing firms increased from 617 in 1996 to 1713 in 2009 with annual average growth rate of about 7.8%. Figure 2 shows the distribution of manufacturing firms across groups of towns. There has been a significant decline in the share of manufacturing firms located in the top five historically important manufacturing hubs, from 77% in 1996 to 55% in 2009. The importance of the capital city in particular has declined significantly from 65% to 42% during the same period. While the historic hubs still host the majority of manufacturing firms in Ethiopia, it is quite clear that most of the recent increases in the number of manufacturing firms have taken place in previously less important towns. "Large" towns other than the historic hubs raised their share of manufacturing firms rapidly since 2002 followed by "other" relatively small and remote towns whose share increased since 2007.[ix]

5. Econometric Results

We start by investigating the cross-sectional relationship between the total number of firms in a locality (town) and our infrastructure measures. We compute town-level averages of the number of firms and the infrastructure variables over the 1999-2008 period, treat each town as one data point, and use a simple logarithmic specification with control variables for food security, woreda population, initial conditions and region dummies. We discard observations for which there is missing information on the variables used in the regressions, resulting in an estimation sample consisting of 69 towns here. Results are shown in Table 3. The estimated coefficients on area accessible (denoted *areacc*) and travel distance (*trvdist*) during a one-hour travel are positive and these variables are highly statistically significant. The estimated coefficient on the travel time to major economic centres (*trvtime*) is negative, but not statistically significant. We further note that initial conditions, proxied by the number of manufacturing firms during 1996 to 1998, tend to have a persistent influence. Woreda population and the dummy variable for food surplus are statistically insignificant. Overall, these results are in line with our working hypothesis, which will be examined further in the subsequent sections.

5.1 Road Infrastructure and Firm Establishment

We now utilise the panel nature of our data in order to estimate (1) whilst controlling for town fixed effects. Our estimation sample consists of 84 towns.[x] Columns (1)-(3) of Table 4 show results from the fixed effects estimator.[xi] The t-statistics in this table, and in all subsequent tables, are based on town-level clustered standard errors, hence they are robust to arbitrary heteroskedasticity and serial correlation. The estimated coefficients on area accessible and travel distance during a one-hour journey are positive and the estimated coefficient on travel time to major economic centres is negative. Travel time to major economic centres is statistically significant at the 10%, but the other two measures of road infrastructure are statistically insignificant. In columns (4)-(6) we show results obtained from estimating (1) in long differences. That is, rather than using the within transformation to eliminate the unobserved town effects, we difference the equation such that the dependent and independent variables are measured as changes over a twelve year period, 1996-2008:

(5) $$\Delta_{12}\log(N_{it}) = \beta \cdot \Delta_{12} roads_{it} + \Delta_{12} u_{it},$$

where $\Delta_{12}\log(N_{it}) = \log(N_{it}) - \log\left(N_{i,t-12}\right)$, and so on. If the *roads* variables are measured with error and these measurement errors are serially uncorrelated while true road infrastructure is highly persistent, the measurement error bias in the long-differences estimator may be more moderate than that of the within estimator (Griliches and Hausman, 1986). Consistent with this hypothesis, the long-diff estimates of the roads coefficients are larger in absolute terms than their fixed effects counterparts. Moreover, the roads variables are statistically significant at the 10% level or lower. We thus obtain evidence from these regressions that, conditional on unobserved time constant town-level factors, there is a positive association between improvements in the quality of road infrastructure and the number of firms in a particular area. The estimates in fact imply that this relationship is economically quite significant.

We now relax the strict exogeneity assumption and instead treat the road infrastructure variables as econometrically endogenous. Using the system GMM estimator proposed by Blundell and Bond (1998), we use $roads_{i,t-2}$, $roads_{i,t-3}$, and $roads_{i,t-4}$ as instruments for the differenced equation (2), and $\Delta roads_{i,t-1}$ as an instrument for the levels equation (1). A full set of time dummies is included in the regressions, but the time invariant control variables are not (since we control for town fixed effects). Estimation results are shown in Table 5. Columns (1)-(3) show results from a non-dynamic specification, and Columns (4)-(6) contain estimates from a dynamic specification in which a lag of the dependent variable is included in the set of explanatory variables. In the specifications without the lagged dependent variable the estimated coefficients on area accessible and travel distance are positive and these variables are statistically significant at the 1% level. The estimated coefficients imply that a 1% improvement in the roads measures is associated with a 1.1-1.2% increase in the number of firms in the locality. These estimates, which are economically quite important, are very similar to those obtained from the long differenced estimator shown in Table 4. The fact that they are larger than the OLS and fixed effect estimates suggests that endogenous road placement does not result in an upward bias. The estimated coefficient on the travel time to major economic centres is negative, but the effect is not quite statistically significant.

There is evidence of first-order and second-order serial correlation in the differenced error term (m1 and m2). Second-order serial correlation suggests serial correlation in the level error term, which may pose a problem given that lags of the explanatory variables are used as instruments. However, based on the Hansen test, we do not reject the null hypothesis that the overidentifying restrictions are valid. Based on the difference-Hansen test, we do not reject the moment conditions used specifically for the levels equation. Overall, the outcomes from these statistical tests suggest that the system GMM model is well specified. We carry out criterion based tests for endogeneity of the *roads* variables (see

Wooldridge, 2010, pp. 226-227, for details on hypothesis testing based on the GMM criterion function). Implementing such a test involves estimating an additional model with $E(\Delta roads_{it}\Delta u_{it}) = 0$ added to the set of moment conditions, and then comparing the resulting criterion value to that obtained with $E(\Delta roads_{it}\Delta u_{it}) = 0$ excluded. Under the null hypothesis that $E(\Delta roads_{it}\Delta u_{it}) = 0$, the simple difference in the criterion value has a chi-square distribution with one degree of freedom. We report *p*-values in Table 5. In no case do we find evidence that the *roads* variables are endogenous.

Extended specifications with the lagged dependent variable included are shown in Columns (4)-(6). Area accessible and travel distance are statistically significant at the 5% level, and 10% level, respectively, and the implied long-run effects are 1.29 and 1.01, respectively.[xii] The results with respect to the estimated effects of road infrastructure in these dynamic specifications are thus similar to those in Columns (1)-(3), and to the long differenced results. Similar to our previous findings, the estimated coefficient on the travel time to major economic centres is negative but not statistically significant. As expected there is a non-negligible degree of persistence in the dependent variable, which is picked up in these regressions by the lagged dependent variable: estimated coefficients on the lagged dependent variable range between 0.6 and 0.77. This result is consistent with the hypothesis that there are dynamic agglomeration effects, such that an increase in the number of firms in town *i* at time *t* spurs additional firm entry in period *t* + 1. Clearly alternative interpretations based on the significance of the lagged dependent variable are possible too. We further observe that, as a result of including a lagged dependent variable in the model, there is no strong evidence of second order serial correlation in the differenced error term, suggesting that the level error term in these dynamic specifications is serially uncorrelated. There is no evidence from the Hansen and diff-in-Hansen tests that the system GMM model is mis-specified.

As part of our effort to identify the impact of road infrastructure on firms' location choices, we investigate whether different types of firms respond to different types of road quality improvements. Specifically, we estimate our model distinguishing between small and large firms. That is, rather than counting all firms in the town, we now count the number of firms with less than 50 employees, and the number of firms with more than 50 employees, and construct two different dependent variables accordingly.[xiii] We consider the dynamic specification, and use system GMM to estimate the parameters. The results are reported in Table 6. Area accessible and travel distance have somewhat larger effects on small firms than on large firms, and the effects are statistically more significant. On the other hand, there is stronger persistence in the dependent variable for large firms, so the implied long run effects of road infrastructure (i.e. the roads coefficient divided by one minus the coefficient on the lagged dependent variable) are actually quite similar for small and large firms. The estimated coefficients on travel time suggest a more important effect for large than for small firms, but in neither case is the effect statistically significant. The specification tests are mostly satisfactory. The two exceptions are the diff-in-Hansen test for travel time for small firms (p-value 0.03) and the Hansen test for travel distance for large firms (p-value 0.037).

Number of entrants

The preceding discussion examines the change in the total number of firms in a town which combines the effects of both firm entry and exit. In this sub-section we focus only on entrants. We consider a firm to be an entrant if it appears for the first time in the CSA census of manufacturing firms during the period 1996-2008. Firms that re-enter the sample after a temporary exit or slippage below the CSA cut-off point will not be considered as entrants. As indicated above, we use a hurdle count data model for the analysis of the number of manufacturing start-ups. Our dependent variable is annual number of entrants in a town from 1999 to 2008. In addition to road infrastructure, the covariates include our proxies for ERA's road placement criteria, time dummies and region dummies (as far as we know it is not feasible to control for town fixed effects due to the nonlinear nature of the model). The results, shown in part (A) of Table 7, indicate that road infrastructure is strongly associated with whether there is any entry at all in a town at a particular time period. Conditional on there being at least one entrant, the association between road infrastructure and the number of entrants is considerably weaker. The initial number of manufacturing firms has a positive and significant effect both on the probability that there is some entry and on the number of entrants conditional on there being some entry. Woreda population and food surplus status, in contrast, are statistically insignificant throughout.

The hurdle count data model (3) implies that the expected number of entrants can be written

(6) $$E(y) = \frac{1-f_1(0)}{1-f_2(0)} \sum_{j=1}^{\infty} j \cdot f_2(j).$$

(see Winkelmann, 2008, p.139). Part (B) of Table 7 shows the estimated marginal effects of road infrastructure on the expected number of entrants, computed by differentiating the sample analogue of eq. (6) with respect to the roads variables and evaluated at sample averages of the explanatory variables. The t-statistics are based on standard errors that were estimated by means of a bootstrapping procedure.[xiv] Similar to the results discussed above, we obtain positive and statistically significant marginal effects of area accessible and travel distance on the number of entrants, and negative but insignificant effects of travel time to major economic centres. Clearly the statistical significance of the marginal effects with respect area accessible and travel distance are driven primarily by the statistical significance of these roads variables in the logit regressions. We also note that the estimated marginal effects and the standard errors are quite similar to those obtained from simple OLS regressions with the dependent variable specified as the number of entrants (results are available on request).

5.2. Road Infrastructure and the Size of Entrants

We now investigate whether better road accessibility of towns is associated with larger firm size at start-up. If demand for manufactures is small and the existing markets are fragmented due to inadequate infrastructure, then firms will start small because of the confined scope of the potential market. As the scope of the market broadens due to better road access, entrants would start with larger firm size as compared to a town with poor road connectivity. As discussed in Section 3 above, entry size is measured at the firm level, and since entry is a one-time event, our dataset in this part of the analysis is thus a pooled cross-section of new entrants. Since there is no panel dimension at the level of the firm we cannot take time differences of the data or control for firm fixed effects.

We use two alternative approaches to control for unobservable factors. The first is to include in the specification a full set of town dummies in order to control for unobserved town fixed effects. It follows that other time constant variables cannot be used within this framework, neither as control variables nor as instrumental variables. We estimate this model using OLS. The second approach is an instrumental variable (IV) approach in which we use data on road density in 1990 (length of roads per thousand square meters) at the woreda level to instrument road accessibility after 1999. The basic idea is that while the Road Sector Development Programme builds upon road networks that already existed in 1990, the latter may not directly affect start-up size after 1999 once initial number of firms is controlled for.[xv] Because road density in 1990 is a time constant variable it will not be possible to include in the specification controls for town fixed effects, and a necessary condition for consistency of the IV estimator is therefore that road density in 1990 is uncorrelated with unobserved town fixed effects. We include in this specification the proxies for ERA's road placement criteria and region dummies. We further allow for time fixed effects, to account for macroeconomic factors (e.g. GDP growth) and improved access to credit, which could increase entry size across all towns.

The OLS results, shown in Table 8, Columns (1)-(3), indicate quite strongly that there is a positive association between the average size of entrants and the quality of the road infrastructure. The estimated coefficients are large, especially for the travel time measure of road infrastructure. Under the assumption that the roads variables are strictly exogenous, in which case these are consistent estimators of causal effects, the estimated elasticity of firm size with respect to area accessible and travel distance is marginally larger than unity. In contrast, the elasticity with respect to travel time is close to three; that is, a 1% reduction in travel time to the major destinations in Ethiopia would imply an increase in the size of new entrants by nearly 3%.

The IV estimates, shown in Table 8, Column (4)-(6), have the same sign as their OLS counterparts, thus consistent with a positive association between the average size of entrants and the quality of the road infrastructure. However, the IV estimates are closer to zero and, the infrastructure variables are not statistically significant. The 1990 road density variable is a reasonably strong instrument, statistically significant in the first stage at the 5% level for the specifications in Columns (4) and (6) and at the 10% level for the specification in Column (5).

Absent an exogenous time varying instrument, we cannot determine by means of statistical tests which of the assumptions underpinning the two approaches above is best supported by the data. To shed some light on whether the IV approach has helped us overcome an endogeneity problem, we implemented Hausman-Wu exogeneity tests. There is no evidence from these tests that road infrastructure is in fact endogenous (see Table 8). In contrast, the town fixed effects are highly statistically significant in the OLS regressions, and there is strong evidence of a non-zero correlation between the fixed effects and the road infrastructure variables. These results suggest that neglecting the town effects may lead to misleading results. We tentatively conclude that the fixed effects estimator is more reliable, although we note again that this conclusion is only valid under the assumption of exogeneity.

We have also investigated whether the start-up size of new entrants relative to the average size of incumbents in a town varies with infrastructure quality. We estimate regressions similar to those shown in Table 8 (either with town dummies included or adopting the same IV approach as above) but with the dependent variable specified as the ratio of entry size to mean firm size in a town at a given point in time (where the average excludes the entrants). There is no evidence whatsoever from these regressions that the size of new entrants *relative* to that of incumbents varies with road infrastructure. That is, the firm size response of incumbents appears to be similar to that of new entrants. These regression results are available on request from the authors.

6. Conclusions

This paper focuses on the relationship between improved road infrastructure and firm dynamics in Ethiopia. It combines census based firm level panel data from the manufacturing sector and GIS based town level panel data on road accessibility. The response variables include the total number of firms, the number of entrants, and the size of new entrants. Obtaining credible estimates of causal effects of road infrastructure on firm-level decisions is generally a difficult task, since road placement is likely correlated with unobserved determinants of such decisions. We exploit the time dimension in our data to control for unobserved town fixed effects, and, where possible, the panel dimension to construct instrumental variables.

Our results indicate that better road accessibility is associated with a town's desirability for manufacturing firms. System GMM results and estimates from regressions specified in long differences suggest that improved local road infrastructure – measured in our data set as the total distance that can be travelled during a 60 minute drive from the centre of town and the total area accessible during those 60 minutes of drive - is associated with a larger number of firms in that locality. The results suggest that the RSDP has contributed to the reduction in the high concentration of firms in historical hubs of manufacturing. Arguably our best measure of how roads affect the connectivity of firms outside the local market is the travel time from a particular town to major economic destinations. This variable is not associated with more firms or a larger number of entrants, but it is quite strongly correlated with the size of new entrants as indicated by regressions controlling for town fixed effects. This result suggests that connectivity with distant markets is important for large firms.

While the primary purpose of the public road infrastructure investment programme studied in this paper was not to spur the manufacturing sector specifically, our results suggest that the manufacturing sector nevertheless benefited from it. However, despite the significant boost to the road infrastructure resulting from the RSDP, much remains to be done, in Ethiopia as well as in other African countries. Structural change and development in rural areas remain important priorities (Page, 2012). In Ethiopia, the vast majority of firms are very small. Such firms typically pay low wages and record low levels of investment and labour productivity, and finding ways of facilitating for the establishment of more successful large firms is therefore a priority (e.g. Page and Söderbom, 2012; Sutton and Kellow, 2010). Our findings suggest that, in order to facilitate for the establishment of large firms, a carefully planned holistic approach for improving infrastructure can be more effective than initiatives at the regional level that are not coordinated across regions or implemented with a view to improving overall connectivity. Regional development, the pace of structural change and modernization of the economy may thus depend on the design of future infrastructure investment programs.

References

Arauzo, J. M. (2005). Determinants of Industrial Location. An Application for Catalan Municipalities. *Papers in Regional Science* 84, 105–120.

Arauzo, J. M., Liviano, D., and Manjón, M. (2010). Empirical Studies in Industrial Location: An Assessment of their Methods and Results. *Journal of Regional Science* 50, 3, 685–711

Arauzo, J. M. and Viladecans, E. (2009). Industrial Location at the Intra-metropolitan Level: The Role of Agglomeration Economies. *Regional Studies*, 43, 545–558.

Baum-Snow, N., Brandt, L., Vernon Henderson, J., Turner, M. A., Zhang, Q. (2012). Roads, Railways and Decentralization of Chinese Cities. Unpublished mimeo, Brown University.

Bloom, D., and Sachs, J. (1998). Geography, Demography, and Economic Growth in Africa. *Brookings Papers on Economic Activity* 2,207-295.

Blundell, R., and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87,1, 115–143.

Buys, P., Deichmann, U., and Wheeler, D. (2010). Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa. *Journal of African Economies* 19, 3, 399–432

Binswanger, P.H., and Khandker, S.R. (1993). How infrastructure and financial institutions affect agricultural output and investment in India. *Journal of Development Economics* 41, 337-366.

Chen, C. (1996). Regional Determinants of Foreign Direct Investment in Mainland China. *Journal of Economic Studies* 23, 2, 18-30.

Cieslik, A. (2005). Regional Characteristics and the Location of Foreign Firms within Poland. *Applied Economics* 37,863–874.

Collier, P. (2000). Africa's Comparative Advantage. In H. Jalilian, M. Tribe, and J. Weiss (Eds.), *Industrial Development and Policy in Africa: Issues of De-Industrialization and Development Strategy*. Cheltenham: Edward Elgar.

Datta, S. (2012). The impact of improved highways on Indian firms. *Journal of Development Economics* 99,1, 46-57.

Dercon, S., Gilligan, D.O., Hoddinott, J., Woldehanna, T. (2008). The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages. *IFPRI Discussion Paper* No. 00840

Donaldson, D. (2010). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. NBER Working Paper No. 16487, National Bureau of Economic Research.

Duflo, E., Banerjee, A., and Qian, N. (2012). On the Road: Access to Transportation Infrastructure and Economic Growth in China. NBER Working Paper No. 17897. National Bureau of Economic Research.

Escribano, A., Guash, J. L., and Pena, J. (2010). Assessing the impact of infrastructure quality on firm productivity in Africa: Cross-country comparisons based on investment climate surveys from 1999 to 2005. *World Bank Policy Research Working Paper 5191*. Washington D.C: The World Bank.

Ethiopian Road Authority. (2011). Assessment of 14 Years of Performance of Road Sector Development Programme. Addis Ababa, Ethiopia .

Fafchamps, M., and Söderbom, M. (2014). Network Proximity and Business Practices in African Manufacturing. *World Bank Economic Review* 28(1), 99-129.

Griliches, Z. and J.A. Hausman (1986). Errors in Variables in Panel Data. *Journal of Econometrics* 31, 93-118.

Government of Ethiopia. (2004). *Productive Safety Net Programme: Programme Implementation Manual*. Addis Ababa: Ministry of Agriculture and Rural Development.

Helpman, E. (1998). The Size of Regions. In D. Pines, E. Sadka, and I. Zilcha (Eds.), *Topics in Public Economics: Theoretical and Applied Analysis*(pp. 33-54). Cambridge: Cambridge University Press.

Holl, A. (2004a). Manufacturing Location and Impacts of Road Transport Infrastructure: Empirical Evidence from Spain. *Regional Science and Urban Economics* 34, 341–363.

Holl, A. (2004b). Transport Infrastructure, Agglomeration Economies, and Firm Birth. Empirical Evidence from Portugal. *Journal of Regional Science* 44, 693–712.

Jedwab, R. and Moradi, A. (2011). Transportation Infrastructure and Development in Ghana. Working Paper No.2011-24, Paris School of Economics.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy* 99, 483–499.

Limão, N., and Venables, A. J. (2001). Infrastructure, Geographical Disadvantage, Transport Costs, and Trade. *The World Bank Economic Review*, 15(3), 451-479.

List, J. (2001). US County-Level Determinants of Inbound FDI: Evidence from a Two-Step Modified Count Data Model. *International Journal of Industrial Organization* 19, 953–973.

McPherson, M. (1995). The Hazards of Small Firms in Southern Africa. *Journal of Development Studies* 32, 1,31-54.

Mills, E. (1967). An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review* 57, 197–210.

Naude, W. and Matthee, M. (2007). The Significance of Transport Cost in Africa. Policy Brief No. 5, UNU-WIDER.

Page, J. (2012). Can Africa Industrialize? Journal of African Economies 21, Special Issue 2, ii86-ii124.

Page, J. and M. Söderbom. (2012). Is Small Beautiful? Small Enterprise, Aid and Employment in Africa. UNU-WIDER Working Paper No. 2012/94. Helsinki, Finland: UNU-wider.

Renkow, M., Hallstrom, D.G., and Karanja, D.D. (2004). Rural infrastructure, transactions costs and market participation in Kenya. *Journal of Development Economics* 73, 349– 367

Rothenberg, A. (2011). Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways. Unpublished Paper

Smith, F. D., and Florida, R. (1994). "Agglomeration and Industrial Location: An Econometric Analysis of Japanese-Affiliated Manufacturing Establishments in Automotive-Related Industries. *Journal of Urban Economics* 36, 23–41.

Sutton, J., and N. Kellow. 2010. An Entreprise Map of Ethiopia. London: International Growth Centre.

Tybout, J. R. (2000). Manufacturing Firms in Developing Countries: How Well Do They Do, and Why? *Journal of Economic Literature* 38, 11–44.

Vesterinen, J. E. Pouta, A. Huhtala, M. Neuvonen (2010). "Impacts of changes in water quality on recreation behavior and benefits in Finland," *Journal of Environmental Management* 91:984-994.

Winkelmann, R. (2008). *Econometric Analysis of Count Data*. Heidelberg, New York: Springer.

Wei, Y., Liu, X., Parker, D., Vaidya, K. (1998). The Regional Distribution of Foreign Direct Investment in China. *Regional Studies* 33,9, 857-867.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts and London, England: The MIT Press.

Table 1: Improvements in Road Infrastructure

Indicator	1997	2011
Asphalt roads in good condition	17%	74%
Gravel roads in good condition	25%	55%
Rural roads in good condition	21%	54%
Total road network in good condition	22%	57%
Road Density/ 1000 sq. km	24.1 km	49.1 km
Road Density/ 1000 Population	0.46 km	0.66 km
Area more than 5km from all weather road	79%	61%
Average distance to all weather road	21.4 km	10.2 km

Source: Table 3, Ethiopian Road Authority (2011).

Table 2

Trends in the Road Accessibility of Firms

Year		(1) Area accessible (km2)	(2) Travel Distance (km)	(3) Travel Time to Major Destinations (hours)
1996	Mean	1116	215	384
	Std Dev	497	118	106
	c.v.	0.45	0.55	0.28
1998	Mean	1122	215	384
	Std Dev	497	118	106
	c.v.	0.44	0.55	0.28
2000	Mean	1132	217	381
	Std Dev	510	120	105
	c.v.	0.45	0.55	0.28
2002	Mean	1182	224	365
	Std Dev	578	127	105
	c.v.	0.49	0.57	0.29
2004	Mean	1212	228	362
	Std Dev	604	132	103
	c.v.	0.50	0.58	0.28
2006	Mean	1290	240	348
	Std Dev	675	139	101
	c.v.	0.52	0.58	0.29
2008	Mean	1377	260	329
	Std Dev	737	157	96
	c.v.	0.54	0.60	0.29
1996-2008	Growth	261.3	45.7	-54.4
	p-value(1)	0.000	0.000	0.000
	Towns	90	90	90

Source: Authors' computation based on data obtained from ERA.
Std Dev = Standard Deviation; c.v. = Coefficient of Variation.
(1) H_0: Zero growth between 1996 and 2008.

Table 3

Road Accessibility and Number of Firms: OLS Estimates

	(1)	(2)	(3)
log *areaacc*	0.840		
	(3.36)***		
log *trvdist*		0.679	
		(3.17)***	
log *trvtime*			-0.154
			(0.70)
log $\text{firms}_{1996\text{-}98}$	0.837	0.832	0.985
	(8.78)***	(8.48)***	(10.51)***
log woreda pop.	0.113	0.102	0.111
	(0.71)	(0.64)	(0.57)
Food surplus area	0.409	0.403	0.460
	(1.22)	(1.21)	(1.16)
R^2	0.75	0.75	0.70
Towns	69	69	69

Note: The dependent variable is the log of the average number of firms in a town over the 1999-2008 period. The numbers in parentheses absolute t-values. Standard errors, clustered at the level of the town, are robust to heteroskedasticity and serial correlation. Statistical significance at the 10 percent, 5 percent, and 1 percent levels is indicated by *,**, and ***, respectively. State (region) dummies are included in all specifications. *areaacc* = area accessible (km2); *trvdist* = travel distance (km); *trvtime* = travel time to major destinations (hours); $\text{firms}_{1996\text{-}98}$ refers to the average number of firms in the locality during 1996 to 1998.

Table 4
Road Accessibility and Number of Firms:
Estimates from Fixed Effects and Long Differenced Specifications

	Fixed Effects			Long Difference: 1996-2008		
	(1)	(2)	(3)	(4)	(5)	(6)
log *areaacc*	0.465			1.017		
	(1.46)			(1.85)*		
log *trvdist*		0.345			1.173	
		(1.10)			(2.27)**	
log *trvtime*			-0.984			-2.158
			(1.79)*			(1.96)*
R^2	0.25	0.25	0.26	0.03	0.05	0.05
Observations	1,008	1,008	1,008	84	84	84

Note: The dependent variable is the log of the number of firms in town *i* at time *t*, expressed in deviations from town-level means (within transformation) in col. (1)-(3) and in long differences in col. (4)-(6). The numbers in parentheses absolute t-values. Standard errors, clustered at the level of the town, are robust to heteroskedasticity and serial correlation. Statistical significance at the 10 percent, 5 percent, and 1 percent levels is indicated by *,**, and ***, respectively. Year dummies are included in the fixed effects specifications. An intercept is included in the long differenced specifications.

Table 5

Road Accessibility and Number of Firms: System GMM Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
log *areaacc*	1.225			0.512		
	(4.16)***			(2.19)**		
log *trvdist*		1.093			0.382	
		(2.94)***			(1.92)*	
log *trvtime*			-2.297			-0.421
			(1.50)			(1.18)
log firms_{t-1}				0.603	0.623	0.772
				(6.09)***	(5.94)***	(9.60)***
Specification tests						
m1 (p-value)	0.006	0.007	0.006	0.000	0.000	0.000
m2 (p-value)	0.008	0.008	0.008	0.117	0.119	0.139
Hansen test (p-value)	0.760	0.330	0.751	0.444	0.333	0.246
Diff-in-Hansen test (p-value)	0.984	0.764	0.875	0.113	0.123	0.242
Roads exogeneous (p-value)	0.173	0.921	0.637	0.276	0.265	0.309
Observations	1,008	1,008	1,008	840	840	840

Note: The dependent variable is the log of the number of firms in town *i* at time *t*. The numbers in parentheses absolute t-values. Standard errors, clustered at the level of the town, are robust to heteroskedasticity and serial correlation. Statistical significance at the 10 percent, 5 percent, and 1 percent levels is indicated by *,**, and ***, respectively. Year dummies are included in all specifications. The instruments are as follows. Col. (1)-(3): differenced equation, the roads variable lagged 2, 3 and 4 periods; levels equation, the first differenced roads variable lagged 1 period; and year dummies. Col. (4)-(6): differenced equation, the roads variable and the firms variable lagged 2, 3 and 4 periods; levels equation, the roads variable and the firms variable differenced and lagged 1 period; and year dummies.

Specification tests: m1 - first order serial correlation in the differenced error term; m2 – second order serial correlation in the differenced error term; Hansen test – validity of the overidentifying restrictions; Diff-in-Hansen test – validity of the overidentifying restrictions related to the levels equation; Roads exogenous – criterion based test of H_0: E(Δ*roads*Δerror term) = 0, against H_1: H_0 not true.

Table 6
Road Accessibility and Number of Firms: System GMM Estimates, Small and Large Firms

	Small Firms			Large firms		
	(1)	(2)	(3)	(4)	(5)	(6)
log *areaacc*	0.662			0.255		
	(2.72)***			(2.27)**		
log *trvdist*		0.487			0.182	
		(2.26)**			(1.61)	
log *trvtime*			-0.395			-0.506
			(1.07)			(1.57)
log $firms_{t-1}$ (small or large)	0.466	0.485	0.629	0.741	0.749	0.723
	(5.18)***	(4.80)***	(6.65)***	(7.77)***	(8.08)***	(7.33)***
Specification tests						
m1 (p-value)	0.000	0.000	0.000	0.001	0.001	0.001
m2 (p-value)	0.408	0.434	0.504	0.425	0.424	0.422
Hansen test (p-value)	0.398	0.225	0.168	0.108	0.037	0.053
Diff-in-Hansen test (p-value)	0.271	0.153	0.030	0.624	0.546	0.381
Roads exogeneous (p-value)	0.075	0.146	0.184	0.986	0.441	0.742
Observations	837	837	837	837	837	837

See Table 5 for notes.

Table 7

Road Accessibility and the Number of New Entrants: Hurdle Count Data Model Estimates

	(1)		*(2)*		(3)	
	Logit	Truncated Neg. Binomial	*Logit*	Truncated Neg. Binomial	Logit	Truncated Neg. Binomial
A. Parameters						
log *areaacc*	1.051	0.289				
	(2.75)***	(1.08)				
log *trvdist*			0.930	0.204		
			(2.69)***	(0.84)		
log *trvtime*					-0.506	0.012
					(1.46)	(0.05)
log $firms_{1996\text{-}98}$	0.913	0.378	0.884	0.381	1.063	0.395
	(6.34)***	(4.59)***	(6.13)***	(4.53)***	(7.97)***	(4.27)***
log woreda pop.	-0.106	-0.023	-0.128	-0.010	-0.056	0.035
	(0.35)	(0.12)	(0.44)	(0.05)	(0.19)	(0.19)
Food surplus area	0.619	0.087	0.624	0.124	0.638	0.124
	(0.98)	(0.18)	(0.98)	(0.26)	(1.06)	(0.26)
B. Marginal Effects						
d E[entrants]/d log(*areaacc*)	0.686					
	(2.02)**					
d E[entrants]/d log(*trvdist*)			0.584			
			(1.92)*			
d E[entrants]/d log(*trvtime*)					-0.269	
					(0.82)	
Observations	690	293	690	293	690	293

Note: The dependent variable in the logit models is a dummy variable that is equal to 1 if there was any new entry in town *i* at time *t* and 0 otherwise. The dependent variable in the truncated negative binomial regressions is the number of new entrants, if any new entry. The numbers in parentheses are absolute t-values. Standard errors, clustered at the level of the town, are robust to heteroskedasticity and serial correlation. Standard errors for the marginal effects were obtained by means of a bootstrap procedure. Statistical significance at the 10 percent, 5 percent, and 1 percent levels is indicated by *,**, and ***, respectively. Year and state (region) dummies are included in all specifications. Observations from 1999-2008 are used in the regressions.

Table 8
Road Accessibility and Size of New Entrants

	OLS			Instrumental variable estimation		
	(1)	(2)	(3)	(4)	(5)	(6)
log *areaacc*	1.087			0.202		
	(1.81)*			(0.85)		
log *trvdist*		1.272			0.219	
		(2.16)**			(0.83)	
log *trvtime*			-2.867			-0.523
			(3.37)***			(1.02)
log $\text{firms}_{1996\text{-}98}$				-0.045	-0.047	0.013
				(0.73)	(0.77)	(0.19)
log woreda pop.				0.246	0.240	0.324
				(2.24)**	(2.12)**	(2.67)***
Food surplus area				-0.119	-0.094	-0.109
				(0.74)	(0.62)	(0.461)
Town dummies	Yes	Yes	Yes	No	No	No
R^2	0.12	0.12	0.12			
Observations	2,296	2,296	2,296	1,825	1,825	1,825

Note: The dependent is the log of the number of employees of firm *k* in town *i* at the time of entry. The numbers in parentheses are absolute t-values. Standard errors, clustered at the level of the town, are robust to heteroskedasticity and serial correlation. Statistical significance at the 10 percent, 5 percent, and 1 percent levels is indicated by *,**, and ***, respectively. Year dummies are included in all specifications. State (region) dummies are included in the models shown in col. (4)-(6).

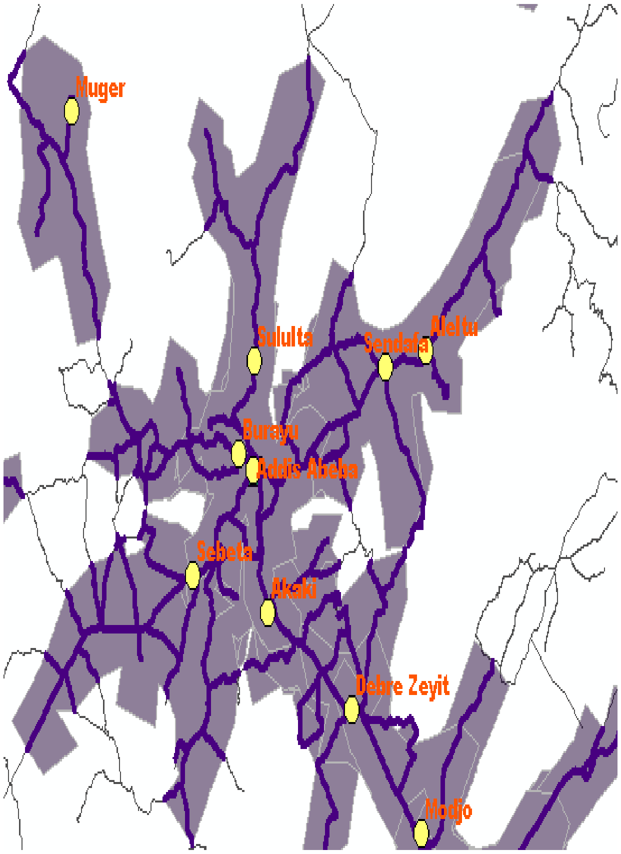

Figure 1a: Area accessible during a 1hr Drive from Addis Ababa in 1996

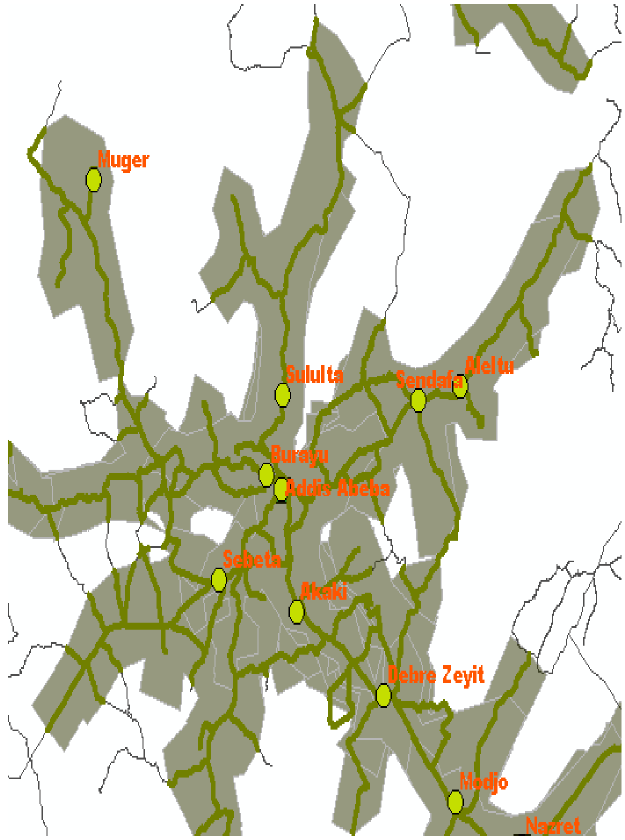

Figure 1b: Area accessible during a 1hr Drive from Addis Ababa in 2008

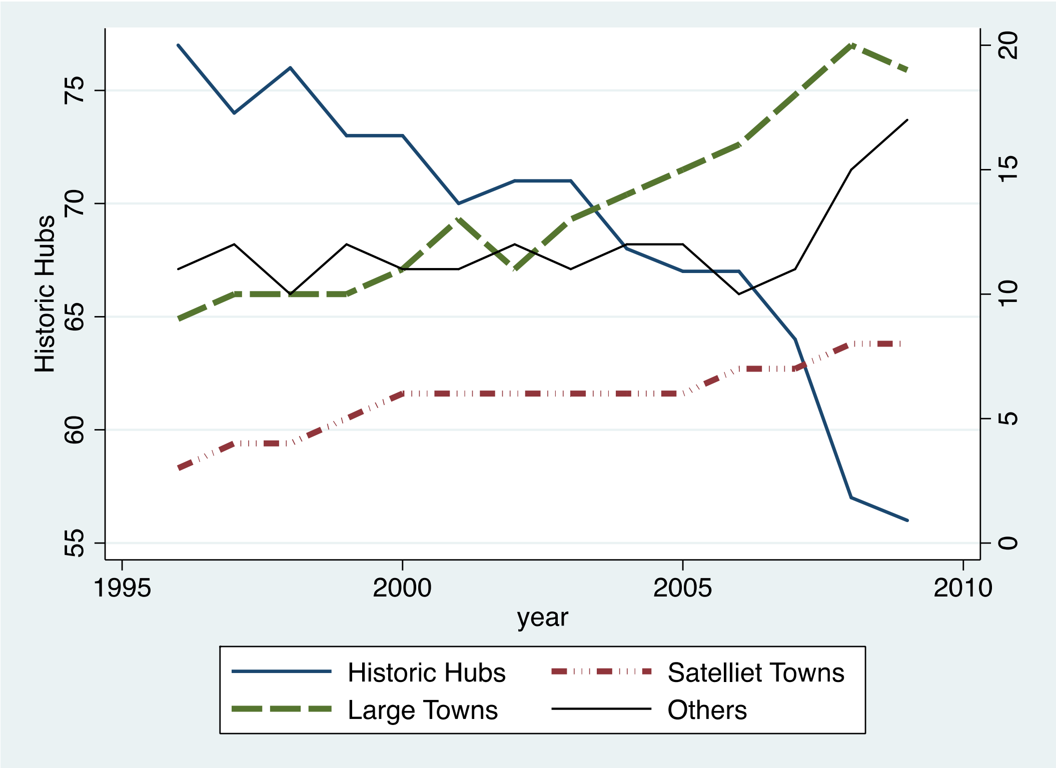

Figure 2: The Distribution of Manufacturing Firms Across Towns (share in %)

Note: The top five towns in 1996 in terms of number of firms are Addis Ababa (65.5%), Dire Dawa (4.1%), Bahir Dar (2.6), Hawassa (2.4%) and Nazreth (2.1%), referred here as "Historic Hubs" of manufacturing. "Satellite Towns" refers to 12 towns close to the capital city (less than 75 miles) while "Large Towns" includes capitals of regional states and large districts (woredas) which until recently have not been important centers of manufacturing. "Others" refers to the remaining 74 typically small and remote towns. The left-hand-side indicates the share of "Historic Hubs" while the shares of all other groups are on the right-hand-side scale. The data for 2005 should be interpreted with caution since a full census was not completed for that year. No data for 2005 are included in the regressions in this paper.

Appendix A: Road Assignment Process

The Ethiopian Road Authority uses the following five criteria during the preliminary selection of new road projects.

i) Roads providing access to areas with economic development potential (20%)

ii) Roads leading to areas with surplus food and cash crop production (20%)

iii) Roads that link existing major roads (20%)

iv) Roads providing access to large and isolated population centers (30%)

v) Roads that bring balanced development amongst the regions in the country and that provide access to emerging regions (10%)

This shows that economic potentials account for about 40% of the weights for new road placement while another 40% weight is given primarily to social equity concerns (criteria iv and v) that could redress existing inequality in road accessibility. ERA uses different inputs to determine the weight for each criterion. The planning department of ERA undertakes the so-called Transport Poverty Observatory on a regular basis which involves “corridor analysis” and “network studies”. Weights for the above mentioned criteria are determined by a committee on the basis of these studies and additional information provided by regional states and government ministries.

Proposals for new roads come mainly from regional states. Each regional state submits its proposal to ERA with its own prioritization and justification. Some government ministries also put forward proposals for new roads. ERA evaluates all the proposals against the five selection criteria. The next step is to see how many new roads can be funded. By aligning road projects with the budget, ERA will present the proposal to the Federal Ministry of Construction and Urban Development. The approved proposal will be presented to the Prime Minister as well as other relevant ministries. Such meetings involve the Governor of National Bank of Ethiopia and the Federal Ministry of Finance and Economic Development (MoFED) administers the federal budget). At this level the overall framework/criteria and fund will be approved. Specific roads are to be selected by ERA based on the agreed framework.

Following the above mentioned preliminary selection process, all selected roads will go through a feasibility study based on which a final project selection will be made. The estimated budget at preliminary level will be adjusted after the feasibility studies. The final budget is determined when the road design is completed by engineers. Once this is done the budget will be submitted to the Ministry of Finance and Economic Development (MoFED). Because of the priority given to the road sector, MoFED often approves the budget with only minor adjustments. For instance for the 2011/12 fiscal year ERA asked for Birr 17 billion and got Birr 15.4 billion.

Upgrading projects also go through a similar process. Most of the proposals for road upgrading come from regional states. The preliminary selection and prioritization of road upgrading projects by ERA is slightly different from that of new roads. The criteria and their respective weights are as follows:

i) Roads with high traffic density - 30%
ii) Roads with better network connectivity -20%
iii) Roads that are in poor condition - 20%
iv) Roads that link import/export and regional integration corridors - 20%
v) Roads connecting investment routs (potential areas) - 10%

The reason why roads with high traffic are given priority for upgrading is that traffic flows that go beyond the designed capacity could cause severe damage to the road at which level routine maintenance may not be economical.

Despite having a set of criteria for road placement, it is not clear what specific measures ERA uses to operationalize them. For instance, it is not clear how exactly economic potentials of different geographic locations are assessed or how regional inequality is evaluated. There is lack of clarity also about the process and criteria by which regional states prioritize their road projects for submission to ERA. From our discussion with ERA, it seems that the above mentioned criteria serve as broad guidelines rather than strict rules for road placement.

Appendix B: Technical Appendix

Table B1: Expected Improvement in Speed of Travel

Pavement Type and Condition	Average Travel Speed	
	Before Rehabilitation/upgrading	After Rehabilitation/upgrading
Asphalt Roads	50km/hr	70km/hr
Federal Gravel Road	35km/hr	45km/hr
Regional Gravel Road	25km/hr	35km/hr
Earth Surfaced Roads	20km/hr	35km/hr
Federal Gravel or regional rural roads to Asphalt Roads	25km/hr to 35km/hr	70km/hr

Source: ERA (2011).

Table B2: Regional Capitals and other Urban Centers as Destination for O-D matrix

Town	POINT_X	POINT_Y
Addis Abeba	472656.04	998453.60
Arba Minch	338197.89	664536.16
Asosa	10524.93	1115450.18
Awasa	441088.20	779102.38
Bahir Dar	324514.77	1281398.44
Dessie	568955.98	1229367.46
Dire Dawa	814860.19	1063118.29
Gambela	12264.77	913864.40
Harer	842893.99	1030414.93
Jigjiga	917315.59	1035495.53
Jima	260937.87	848508.63
Mekele	551884.95	1492540.45
Nazret	528918.44	943849.55
Nekemte	230291.32	1005545.02
Semera	717990.30	1300962.68

Endnotes

[i] Evidence from the US is provided in Smith and Florida (1994) and List (2001); Arauzo and Viladecans (2009), Holl (2004a) and Arauzo (2005) provide evidence from Spain; Cieslik (2005) and Holl (2004b) focus on Poland and Portugal, respectively. Arauzo et al. (2010) provide a recent review of the empirical studies in developed countries.

[ii] Binswanger and Khandker (1993) estimate the response of aggregate private investment and aggregate output in rural India using district level data on road networks. Rothenberg (2011) investigates the effect of highways on the location choice of new firms in Indonesia. Chen (1996) and Wei et al. (1998) examine the location choices of FDI firms in China and find a positive effect of infrastructure. Datta (2012) evaluates the impact of upgrading Indian highways and finds a significant reduction in the stock of intermediate inputs. Other studies focusing on infrastructure and various aspects of development in Asia include Donaldson (2010; India), Duflo et al. (2012; China), and Baum-Snow et al. (2012; China).

[iii] Most studies related to infrastructure in Sub-Saharan Africa focus on the implications either for international trade (e.g. Limão and Venables, 2001; Buys, Deichmann and Wheeler, 2010; Naude and Matthee, 2007), or the decisions and outcomes of rural households (Dercon et al., 2008; Renkow et al.; 2004; McPherson, 1995). One exception is the analysis by Escribano, Guash and Pena (2010) which uses firm level data from 26 African countries to document the effects of a set of infrastructure indicators on aggregate productivity. Another is that by Jedwab and Moradi (2011), which provides historical evidence that railway lines in the early 20th century contributed to the boom in cocoa production and urbanization in Ghana.

[iv] The precise mechanisms through which road networks boost firm performance have been extensively discussed in the literature. Some contributions to the literature on economic geography, such as Krugman’s (1991) core-periphery model, focus on agglomeration benefits.

Such benefits are reinforced by reduced transport costs, since with better road connectivity firms can agglomerate in some location while supplying the rest of the market through efficient transport networks. Some contributions to the literature on urban economics, on the other hand, point out that lower transport costs may incentivise firms to locate in the periphery of the market rather than at the core in order to avoid congestion costs (Mills, 1967; Helpman, 1998). Since lower transport costs may change the relative attractiveness of established economic centres (large cities) versus peripheries, determining their effect on firms' location is an area of active empirical research.

[v] A similar approach has recently been used by Vesterin et al. (2010).

[vi] Since ERA's report does not contain the completion period for some of the projects, other documents were consulted for detailed information on project specific physical accomplishment and budget disbursement.

[vii] Our sample does not include manufacturing establishments with less than 10 workers. According to the recent Survey of Small Scale Manufacturing carried out by the CSA, there were 43,338 manufacturing firms with less than 10 workers in 2007 as compared to 1339 firms in our sample with at least 10 workers. These are very small firms with average firm size of about 2.75 persons including proprietors. Total employment in 2007 by such small firms was 138,951 of which only 32 per cent are paid jobs. The results in this paper therefore refer to formal sector firms which account for the bulk of manufacturing value added. Unlike the census data we are using in this paper, surveys on firms with less than 10 workers are not collected regularly.

[viii] Only a few road projects were launched at the beginning of the RSDP and most of those projects were upgrading and rehabilitation of existing roads. Moreover, road projects normally take several years to complete.

[ix] As indicated earlier 2002 marks the beginning of the second RSDP while the third round began in 2007.

[x] Compared to the full sample of 90 towns, we lose 6 towns due to incomplete data on the number of firms. Since the control variables are redundant in regressions controlling for town fixed effects, we can use observations even if data on the control variables are missing. This is the reason we can use a larger sample of towns for regressions controlling for town fixed effects than for regressions with the control variables included.

[xi] As discussed in Section 3, our proxies for endogenous road placement are time invariant. These variables are therefore eliminated from the equation by the within transformation and the differencing.

[xii] Long run effects are calculated as the coefficient on the roads variable divided by one minus the coefficient on the lagged dependent variable.

[xiii] We would like to thank a referee for suggesting this approach.

[xiv] For each replication $j=1,2,\ldots J$, we draw a bootstrap sample of towns and estimate the marginal effect of the road infrastructure variable on E(y), as defined by eq. (5). The standard deviation of the estimated marginal effects across the *J* bootstrapped samples is our estimate of the standard error of the marginal effect for our actual sample. The procedure ensures the standard errors are robust to arbitrary heteroskedasticity and autocorrelation within towns.

[xv] The variable based on historical data on road density is likely less suitable as an instrument for current road access in a model where the dependent variable is either the total number entrants or the total number of firms at the town level. This is because manufacturing firms

are still concentrated in major urban centers, which have long been connected to Addis Ababa via trunk roads and continue to attract new firms.