The Role of Road Quality Investments on Economic Activity and Welfare: Evidence from Indonesia's Highways*

Paul J. Gertler
University of California, Berkeley

Marco Gonzalez-Navarro
University of Toronto

Tadeja Gracner
University of California, Berkeley

Alexander D. Rothenberg
RAND Corporation

August 2014

Abstract

This paper provides new evidence of the effects of road quality investments on economic activity and welfare. A uniquely detailed nationwide panel dataset of road surface roughness allows us to construct instruments for road quality from temporal variation in budgets for different road maintenance authorities. We find that higher road network quality causes improvements in household consumption and income. In terms of mechanisms, we first show that roads cause job creation in the manufacturing sector. Second, we show evidence of an occupational shift from agriculture into manufacturing and higher profits for those who stay in agriculture. The gap in average income between agriculture and manufacturing employment is reduced with road quality but not eliminated. Because wages in the manufacturing sector do not change with road quality the results are consistent with dual labor markets.

*The authors thank seminar participants at Berkeley, Millenium Challenge Corporation, University of Toronto, and University of Ottawa for helpful comments and suggestions. This work was generously supported by a USAID Economic Research Partnership Grant.

1 Introduction

Road networks account for a substantial proportion of government expenditures yet little is known about their effects in very poor countries. Recent contributions to this growing literature have used instances of major road works such as China's national trunk system to identify effects of roads in small jurisdictions along the path connecting two large cities. This paper adds to the existing literature by providing evidence on the effects road maintenance within the existing road network.

We are able to accomplish this using an unusually long and comprehensive road quality database collected by Indonesia's road authorities over a 20 year period. The richness of the road quality data allow us to propose a new instrumental variable identification strategy based on a two stage budgeting process. The empirical results show large effects of road quality on household consumption and income levels. We next investigate mechanisms that explain these effects of road quality on household welfare. First, we use manufacturing plant census data to show that road quality allows for manufacturing job growth. We find that roads increase the total number of firms but not average firm size, consistent with a competitive firm environment. For agricultural firms, we find that road quality allows for higher profits.

Because manufacturing jobs are on average better paying than agriculture, the local expansion of manufacturing is possible due to an occupational shift from a proportion of individuals previously engaged in agriculture. In fact, road quality helps close the gap between agriculture and manufacturing but does not eliminate it. We interpret this as evidence of a dual labor market, not only because there is a wage gap between occupations in the same place (controlling for individual fixed effects) but also because wages in the manufacturing sector do not change with road quality.

There are two key problems to tackle in evaluating the impacts of changes in transport infrastructure. First, measuring changes in regional transport costs is difficult. There are no readily available datasets that document the evolution of the quality of transport infrastructure in developing countries.[1] In this sense, the road quality data we use represent a substantial improvement because they provide a very direct measure of transport infrastructure quality.

The second challenge is that road improvements are typically not randomly assigned. This means that estimates of the effects of changes in transport infrastructure may be confounded by the fact that areas receiving improvements were selected by policymakers for economic growth or expected growth reasons, creating simultaneity bias. We address the simultaneity

[1] Previous approaches, widely used in the trade literature, involve *inferring* transport costs from a gravity equation, using data on regional trade flows (e.g. Anderson and Van Wincoop, 2004) or price differences (e.g. Donaldson, 2010) which are not widely tracked and hence rarely available.

problem by making use of a new instrument which has the advantage of being replicable in most contexts. It is constructed from data routinely collected by national road maintenance authorities: the International Roughness Index (IRI) - defined as the the ratio of a vehicle's accumulated suspension motion (in meters), divided by the distance travelled by the vehicle during measurement (in kilometers). IRI data provide information of the road quality of every road segment in Indonesia over a 20 year period. The instrument construction also requires that there exist different road funding authorities in charge of different parts of the road network. In the case of Indonesia, road funding from general sources is allocated to national, provincial and local authorities. Each road authority has a set of roads it is responsible for and selects which roads to upgrade within its purview with an exogenously allocated but time varying budget. In relatively high budget years more roads can be maintained so average road quality increases in the whole budgeting authority's set of roads. Note that the selection problem occurs only when the road authority selects *which* roads to maintain with its exogenous budget. We can take advantage of this two stage budgeting procedure to instrument the quality of a given road segment using the average IRI of the specific road budgeting authority in that year. Consider the case of a person living in a given village: In a trajectory to another location it will use a proportion of roads maintained by local, provincial or national authorities. This generates variation across villages and time in the instrument. The effects identified with this strategy come from the set of roads that get maintained when road budgets allow it, but which get less maintenance when road budgets are tight, ie the set of marginal roads for policy making purposes. There are roads that are such high (low) priority that they are are always well (never) maintained and which experience little or no variation regardless of changes in the budget. We argue that the set of marginal roads are those of policy interest.

In section 2 we provide a theoretical framework, section 3 describes the evolution of road quality and argues in favor of the instrumental variable strategy we pursue, Section 4 presents the results and section 5 concludes.

2 Theoretical Framework

We conceptualize road quality as a village productive amenity, as in Jacoby (2000). The model we propose embeds a standard, prototypical model of the agricultural household (e.g. Singh et al., 1986; Benjamin, 1992; Bardhan and Udry, 1999) in a spatial general equilibrium model typically used in labor and urban economics (e.g. Rosen, 1979; Roback, 1982).

The economy is a collection of M discrete villages. Villages are each endowed with a single productive amenity, which we denote by A. In our application, we assume that villages with better roads have larger values of A. Without loss of generality, we can order these villages by their values of this productive amenity, as follows:

$$\underline{A} \equiv A_1 < A_2 < ... < A_M \equiv \overline{A}$$

There are two types of agents in the model: firms, which make use of land and labor in constant-returns-to-scale production, and household-workers, which decide how to allocate land and labor between farming and off-farm employment. We describe the objectives and constraints of each of these agents in the next subsections. The objective of this section is to provide a conceptual framework within which to interpret the empirical analysis.

2.1 Firms

Firms produce a composite good, X, under a constant-returns-to-scale production environment. The good is freely traded over space, and without loss of generality, we can normalize output prices to 1 (the model's numeraire). Firms use capital, labor, and land for production. For simplicity, we assume that capital is freely traded over space at rate ρ, and we also assume that each firm inelastically demands a single unit of land for production. Both rents, r, and wages, w, vary across locations and are determined in equilibrium. The firm's problem is to choose a cost-minimizing quantity of labor, L^M, to satisfy production requirements.

Crucially, we assume that there is some value of the productive amenity, $\widetilde{A}$, such that if a village has $A < \widetilde{A}$, no firm production will take place. At values of $A < \widetilde{A}$, firms will not be able to produce output with *any* amounts of capital or labor. This amenity threshold introduces non-concavity in an otherwise standard production function. We can write the firm's problem as follows:

$$\min_{L^M, K} wL^M + \rho K + r \qquad \text{subject to} \tag{1}$$

$$X = \begin{cases} 0 & if A < \widetilde{A} \\ G\left(L^M, K, 1\,;\, A\right) & if A \geq \widetilde{A} \end{cases} \tag{2}$$

Optimal choices of inputs lead to cost minimization. This yields a cost function, denoted by, $C(w, r; A, \rho)$, which maps wages, rents, the price of capital, and the value of the amenity into the minimized value of production costs. Because of free entry, firms in locations where $A \geq \widetilde{A}$ will produce until the point where production costs equal output prices (which are normalized to 1). This gives us the following equilibrium condition:

$$C(w, r; A, \rho) = \begin{cases} \infty & A < \widetilde{A} \\ 1 & A \geq \widetilde{A} \end{cases} \tag{3}$$

Note that because the price of capital is equal over space, ρ is a parameter in this expression. By applying Sheppard's Lemma to the convex portion of the cost function, we obtain the following expression for labor demand:

$$L_d^{M*} = \begin{cases} 0 & A < \widetilde{A} \\ C_w(w, r; A, \rho) & A \geq \widetilde{A} \end{cases}$$

2.2 Households

For simplicity, we assume that agricultural households are unitary, abstracting from intra-household resource allocation considerations. In each village, households inelastically demand a single unit of land, which they rent at rate r and use for farming. Households choose quantities of consumption, C, and decide how to allocate their labor endowment, E^L, between working on the farm, L^F, and working for a firm, L^M. Conditional on the choice of a location, indexed by A, the household's problem can be stated as follows:

$$\max_{C, L^M, L^F} U(C) \quad \text{subject to} \tag{4}$$

$$C \leq F\left(L^F, 1\,;\, A\right) - r + wL^M \tag{5}$$

$$E^L = L^M + L^F \tag{6}$$

where the utility function, $U(\cdot)$, is assumed to be continuously differentiable, with $U'(\cdot) > 0$, and $U''(\cdot) < 0$, and $F\left(L^F, 1\,;\, A\right)$ is the farm production function. We assume that $F_A > 0$, as higher A leads directly to higher farm output per unit of input, or it may lower input prices.[2] Equation (5) is the full income constraint, stating that the value of the household's consumption cannot exceed the value of farm profits plus off-farm labor earnings.

Equation (6) is the labor resource constraint, stating that the household's total labor endowment is equal to the total amounts of labor supplied to the market and to the farm. Implicitly, we are assuming that the household does not value leisure and that there is no unemployment. If individuals cannot spend their entire endowment of labor in off-farm

[2] Note that the consumer-farmer-worker does not obtain utility for services provided by housing. Nor does the household's utility depend on A directly. The entire consumer amenity aspect of A comes through the dependence of household indirect utility on the farm production function, $F(\cdot)$.

employment, they must spend the rest of their time farming. Under this assumption changes in A have no effects on the extensive (probability of working) or intensive (hours of work) margins of labor supply.

Note that we also assume, for simplicity, that labor on the farm can only be directly supplied by the individual farmer household. This makes sense in a rural context with large families and contracting imperfections. We solve the household's problem in two cases, depending on the value of the amenity. If $A \geq \widetilde{A}$, then there is no job rationing, and the agricultural household model is separable. Following Benjamin (1992), we solve the model recursively. The household first chooses quantities of labor and land to maximize farm profits, which yields the farm's profit function:

$$\pi^* (w, r\,;\, A) = \max_{L^F} F\left(L^F, 1\,;\, A\right) - wL^F - r$$

This gives us the household's income, which we can write as:

$$M^* = \pi^* (w, r\,;\, A) + wE$$

Optimal choices of consumption lead to an indirect utility function, which maps income and prices into the maximum amount of household utility that can be attained:

$$\begin{aligned} V &= \widetilde{\psi} \left(\pi^* (w, r\,;\, A) + wE\right) \\ &\equiv \psi (w, r\,;\, A) \end{aligned} \tag{7}$$

On the other hand, if $A < \widetilde{A}$, then there are no off-farm employment opportunities for work, and $L^M = 0$. The household's supplies all of its labor to the farm, yielding the following value of income and consumption:

$$C = F\left(E^L, 1\,;\, A\right) - r$$

This implies that the farmer's indirect utility is:

$$\begin{aligned} V &= \widetilde{\psi} \left(F\left(E^L, 1\,;\, A\right) - r\right) \\ &= \psi (w, r\,;\, A) \end{aligned}$$

In locations where $A < \widetilde{A}$, at some wage levels, workers may want to supply labor to the market. This could be because at some level of farming labor, $L^{F*} \leq E^L$, the marginal product of labor is zero. However, because of the absence of non-farm employment opportunities, the household supplies its entire labor endowment to the farm. The combination of non-concavity in the manufacturing production function with respect to A and eventual zero marginal product of labor in farming creates dual labor markets (Sen, 1966; Dickens and Lang, 1985).

If migration is costless, workers will be perfectly mobile across locations, so that the following must hold:

$$\overline{V} \equiv \psi\left(w, r\,;\,A\right) \text{ for all } A \tag{8}$$

On the other hand, if migration is costly, there can be a wedge in utility across space.

2.3 Land Markets

Both farmers and firms inelastically demand a single unit of land, which they rent at rate r. Land is supplied on a spot market, according to marginal cost, so that:

$$r = h'\left(N_f + N\right) \tag{9}$$

where N_f is equal to the number of firms, N is equal to the number of agricultural households, and $h(\cdot)$ represents the total cost of clearing land for production and dividing into plots. We assume that $h(\cdot)$ is increasing and strictly convex in the total population of firms and households, $N_f + N$.

2.4 Spatial General Equilibrium

We consider spatial equilibrium conditions to hold in the long-run. Firms are created and labor reallocates, and as this occurs, wages and rents will adjust to equalize production costs and indirect utility over space. Mobility provides us with three equilibrium conditions: (3), (8), and (9). Using these, we can show a series of propositions from the model. We first show that because A is a productive amenity, increases in A will result in an entry of firms and an in-migration of households.

We can implicitly differentiate the equilibrium conditions, (3), (8), and (9), and solve for the comparative statics.

$$\begin{bmatrix} C_w & C_r & 0 \\ \psi_w & \psi_r & 0 \\ 0 & -1 & h''(\cdot) \end{bmatrix} \begin{bmatrix} \partial w / \partial A \\ \partial r / \partial A \\ \partial\left(N_f + N\right) / \partial A \end{bmatrix} = \begin{bmatrix} -C_A \\ -\psi_A \\ 0 \end{bmatrix}$$

Solving the second equation gives:

$$\frac{\partial r}{\partial A} = \frac{1}{\Delta}\left(\underset{+}{C_w}\underset{+}{\psi_A} - \underset{-}{C_A}\underset{+}{\psi_w}\right) > 0$$

where the denominator, Δ, is positive and given by:

$$\Delta = \underset{+}{\psi_w}\underset{+}{C_r} - \underset{+}{C_w}\underset{-}{\psi_r} > 0$$

So that increasing A leads to increases in land values (rents).

Solving for the first equation gives:

$$\frac{\partial w}{\partial A} = \frac{1}{\Delta}\left(\underset{-}{C_A}\underset{-}{\psi_r} - \underset{+}{C_r}\underset{+}{\psi_A}\right) \gtrless 0$$

So that increasing A leads to ambiguous effects on wages. Note that in low amenity villages, the marginal product of agricultural labor is very low because households in these villages allocate their entire labor endowment to the farm. Road improvements in these villages can attract some labor labor away from farming without raising wages.

3 Road Development in Indonesia

Indonesia is a lower middle income country with a population of nearly 250 million people, and a 2012 current GDP of $878 billion. The country's transition from a low-income to a lower middle-income country happened mainly because of its strong economic performance over the last three decades. Indonesia's economy has been one of the fastest growing 'tiger' economies in Asia, especially between the mid 1970s and the late 1990s, enjoying annual growth rates of more than six percent and making major progress in tackling poverty, which decreased from 60 percent to 10 percent during that time Hill (2000). This period overlaps with the Suharto regime and his government's investments in major public works programs, including those that targeted road improvements.

When Suharto assumed power in 1967, rehabilitating roads that had been built by the Dutch colonists and left to deteriorate became a top priority of his first five-year development plans. However, due to oil price shocks and the consequent collapse of state oil revenues in the late 1970s, spending on road infrastructure stagnated until the early 1990s, when the budget share allocated to road improvements increased substantially. Between Repelita IV (1984 - 1989) and Repelita V (1989 - 1994), the total budget for road improvements almost doubled from $2.1 to around $3.9 billion.[3] During Repelita V, transportation constituted the single largest item of the development budget and formed almost 20 percent of total planned development expenditures.[4]

Road improvements followed long-term national spatial plans. These plans mandated that particular regions, especially major islands and sparsely populated areas, were those that would receive infrastructure improvements. Importantly, these plans were revised very infrequently (approximately once a decade), suggesting that the road authorities did not regularly respond to changes in outcomes, and that location fixed effects can remove much of the targeting bias.[5]

Over this period, Indonesia experienced more than a 20 percent annual growth rate in manufacturing sector. This was accompanied by significant structural changes in labor market conditions, with a considerable shift of the labor force out of the informal sector and into the formal sector and from agriculture to manufacturing (Matsumoto and Verick, 2011). However, Indonesia's success in the early to mid 1990s was interrupted by the financial crisis of 1997-1998 that not only resulted in GDP dropping by more than 14 percent in one year, but also significantly slowed spending on transportation. After the crisis, Indonesia's experiences with decentralization gave local government more power and responsibility over the maintenance of national and provincial roads. As many local governments were not equipped to face these challenges, road surfaces started to deteriorate. During this time, the labor market appeared to be rather stable at the aggregate level.

3.1 Data on Road Roughness

In order to evaluate the impact of changes in road quality on welfare and labor markets, we use a novel longitudinal dataset on road roughness in Indonesia. Beginning in 1990, the Department of Public Works (*Departemen Pekerjaan Umum*, or DPU) conducted annual, high resolution surveys to track road quality and monitor pavement deterioration. These data are part of Indonesia's Integrated Road Management System (IRMS) and were used to determine which roads were servicable and which roads needed to be improved. With differing qualities, this type of data gathering systems are quite common in the developing world today. From the Indonesia road quality data collection system, we use data that record

[3]These figures are expressed in 2000 U.S. dollars.

[4]These figures are taken from various planning documents describing Indonesias five year development plans (Rencana Pembangunan Lima Tahun, abbreviated as *Repelita*).

[5]Unfortunately, we do not have access to the exact national spatial plans used, however this idea was confirmed by conversations with highway authorities at Indonesia's Department of Public Works (DPU). Also, in every national planning and budgeting document we do have access to, no information is provided at levels below the province.

the surface type, width, and the international roughness index (IRI) of every kilometer-post interval section of all major inter-urban roads in Indonesia, annually from 1990 through 2007. We merge these data with a digital map of the road network, using road-link identifiers, which yields a panel of road quality measures used for our analysis.[6]

The international roughness index (IRI) is a widely accepted measure of road quality in civil engineering and was developed in the 1980s by the World Bank. It is defined as the ratio of a vehicle's accumulated suspension motion (in meters), divided by the distance traveled by the vehicle during measurement (in kilometers).[7] All else equal, when driving on gravel roads or when faced with potholes and difficult pavement, drivers decrease speed and prolong their travel time. Consequently, road roughness directly increases transport costs. Road roughness also affects transport costs via fuel consumption, vehicle maintenance, and labor costs, all of which account for more than 54 percent of vehicle operating costs in Indonesia (Asia Foundation, 2008).[8]

Our primary measure of road roughness is a distance-weighted average of roughness for all roads in district d. Let $r = 1, ..., R(d)$ index road segments in district d, and let d_r denote the length of the road segment r. Our average roughness measure is defined as:

$$\text{IRI}_{dt} = \frac{\sum_{r=1}^{R} d_r \text{IRI}_{rdt}}{\sum_{r=1}^{R} d_r}$$

where IRI_{rdt} denotes the road roughness of road section r in district d at time t. Although we focus on this measure of road quality for the main analysis, our results are robust to using different measures, including roughness-induced travel times to provincial capitals or measures of market access.[9]

Figure 1 shows a significant leftward shift in the distribution of IRI_{dt} across districts between 1990 and 2007. After decentralization, however, Indonesia's roads started to deteriorate, which is evident in Figure 2.

3.2 Road Deterioration and Usage

We use the segment-level road quality data to describe the patterns of road deterioration and upgrading over the 1990 to 2007 period in more detail. We first assume that for segment r,

[6] Appendix Section B.1 provides details about the road quality data, the process of merging the interval data to network shapefiles, and information on the creation of variables.

[7] See Appendix Section B.1.2 for more details on IRI.

[8] Other significant cost factors included lubricants and tires (13%) which should increase as cars are driven on rougher roads.

[9] See Section 4 for more details.

road quality evolves according to the following dynamic equation:

$$\text{IRI}_{rt} = \text{IRI}_{r,t-1} \cdot (1 + \delta) + \alpha I_{rt} \tag{10}$$

where δ denotes road deterioration and I is an indicator for whether or not segment r was upgraded at time t. This indicator is constructed by inspecting the road segment panel data and looking for points in time in which roads are improved (i.e. $I_{rt} = 1$ whenever $\Delta\text{IRI}_{rt} < 0$). Table 1 shows road segment level regressions of equation (10) in columns 1 and 4, without and with segment fixed effects respectively. The estimation shows that the average road segment degrades by 17% per year. An event of road upgrading improves road quality by 0.48 log points, which is equvivalent to road upgrading undoing 3 years worth of degradation.

Instead of specifying a fixed per-period deterioration rate, we also allow for greater usage to cause road quality to deteriorate more rapidly:

$$\delta_{rt} = f\left(\text{Use}_{rt}\right)$$

Using a linear approximation to $f\left(\cdot\right)$, we write $\delta_{rt} = \phi_1 + \phi_2 \, Use_{rt}$, to estimate:

$$\text{IRI}_{rt} = \text{IRI}_{r,t-1} \cdot (1 + \phi_1 + \phi_2 \, \text{Use}_{rt}) + \alpha I_{rt} \tag{11}$$

Since vehicle counts along each roads segment are unavailable, we use local economic conditions, such as district GDP and the number of large manufacturing firms in the district, as proxies for use of roads. Column 2 provides estimates of equation (11), and a version of equation (11) with main effects is presented in column 3; the same estimates with segment fixed effects are provided in columns 5 and 6. Accelerated degradation with more intensive road use should result in positive estimates of GDP and firms, yet the data do not show this even when using segment level fixed effects. Overall the table shows that local economic conditions do not affect road depreciation rates. We conjecture that this is due to the selection of road materials for a given road use, but do not pursue this here. For our purposes it is enough to know that deterioration rates do not vary with local economic conditions.

3.3 Two Stage Budgeting for Road Improvements

Roads in Indonesia are maintained by three different levels of government: National, Provincial, and District. In each case, maintenance and capital expenditure funds are obtained from general taxation funds. Toll roads and dedicated taxes represent an insignificant share of road financing (World Bank, 2012). Pooled funds are distributed to the different road maintenance agencies and funding decisions are then made locally. For more details, see Figure 3.

We first investigate road upgrading decisions I_t – the second part of equation (10) – using reduced form models of investment. Decisions for road upgrading can be made in two ways. First, a threshold type road management system suggests improvements should be allocated to roads that exceed a threshold level of roughness. If this is the case, β_2 and β_3 of the following regression (12) are not significant:

$$I_{jt} = \alpha + \beta_1 \text{IRI}_{t-1} + \beta_2 \text{ Years Since Last Upgrade}_{rt} + \beta_3 \left(\text{Years Since Last Upgrade}_{rt}\right)^2 + u_{jt} \tag{12}$$

Another possible road management model is one in which roads are paved on a rotating basis regardless of road roughness. In that case β_2 and β_3 should be large and significant, the opposite of β_1 that should not be significantly different from zero. Table 2, column 1 presents regression estimates of equation 12. Results show that Indonesian road investment decisions are in fact a hybrid of these two polar cases. Road quality conditional on time since last upgrade is statistically different from zero and suggests more degraded roads are improved sooner. Years since last upgrade and its square term are both significantly different from zero and suggest a profile in which the probability of investment is decreasing until 6.5 years have elapsed and increases thereafter (conditional on road roughness).

Next, we construct a variable to approximate the total budget for national, provincial, and district road improvements. Let $R(a)$ denote the set of road segments under administrative authority a (national, provincial, or district). Our budget proxy for adminstrative authority a in year t is the following:

$$B_{at} = \sum_{r \in R(a)} d_r I_{rt}$$

where d_r is the distance of segment r in kilometers. This budget proxy is just equal to the total kilometers of roads upgraded in year t that are administered by authority a. Note that the administrative authorities vary over space; in our data (Java, Sumatra, and Sulawesi), there is a single national authority, 17 provinces, and 218 districts.

Columns 2 and 3 in Table 3 show that budgets for all types of roads in the province (district) are strongly correlated to the probability any given segment is upgraded in the province (district), consistent with two stage budgeting. In the second stage of the budgeting road authorities decide *which* roads to upgrade, and optimal allocation suggests that local economic conditions should matter for that decision. Columns 4 and 5 add these local economic conditions as controls. Column 5 shows that indeed local number of firms are significantly correlated to whether a given road segment is improved, which is the expected kind of endogenous variation in road upgrading.

3.4 Road Budgets as Instruments for Road Quality

We showed that Indonesian road improvement decisions use a two stage budgeting process (See Deaton and Muellbauer (1980) for an overview of two stage budgeting). Once agencies observe their funding allocation, the selection of which roads to upgrade takes place. We argue that the two-step budgeting for roads in Indonesia allows us to construct an instrument for road quality, which equals total road funding at the provincial or district level. In this section, we show that road *budgets* are relevant instruments for IRI and are uncorrelated with local economic conditions, which is necessary for the exclusion restriction.

In Table 4 columns 1 and 2 the road budget of provincial road authorities is regressed against lagged province GDP and lagged number of firms respectively. In both cases we cannot reject that the coefficients on economic conditions are uncorrelated to the road budget. Columns 3 and 4 show an analogous exercise for district road budgets against district GDP and district number of firms. Again, we cannot reject that the correlation between allocated budgets to roads and local economic conditions is zero and is consistent with two stage budgeting. The national road budgets do respond to local economic conditions, hence we do not use them as an instrument for road quality.

As shown in the previous subsection, available road funding directly impacts the probability that a road segment is improved. Columns 5 and 6 in the table show that in a road segment fixed effects model budgets have a large and significant effect on road quality. The negative sign of the coefficient confirms that whenever the province or district has a more generous road fund any given road segment in the province (district) is more likely to be of better quality (less rough).

4 Results

In this section, we first provide evidence on the relationship between changes in road roughness and household welfare. Next, we empirically evaluate the many different mechanisms through which these welfare effects might occur, following the discussion in Section 2. To do so, we make use of panel data regression techniques. Panel data allow us to control for unobserved individual-specific (or household-specific) heterogeneity which may be correlated with the provision of road improvements. If road authorities targeted locations that were historically poor (or wealthy) with upgrades, then fixed effects would be sufficient to remove all of the targeting bias from parameter estimates. However, to the extent that road authorities targeted faster growing areas, or areas experiencing declines in activity, we need an instrument that shifts road quality but is orthogonal to local demand for roads. Our main specifications combine panel data and fixed effects with instrumental variables estimates that use the budget instrument described in the previous section.

Our primary data source for individual and household-level panel data is the Indonesian Family Life Survey (IFLS). The IFLS is a nationally representative longitudinal survey that was collected in 1993, 1997, 2000 and 2007. The IFLS is representative of 83 percent of Indonesia's total population and follows more than 30,000 individuals over a 14 year period. These individuals are observed in more than 300 villages (*desa*), which are located in 13 of Indonesia's 27 provinces. The IFLS is notable for its low attrition rate, as more than 87 percent of the original households are tracked through all four waves of the survey.

The IFLS contains detailed modules of labor market activity, including questions about employment status, occupation, hours worked in different sectors, whether or not a job is in the formal sector, and self-reported earnings by sector. The data also contain many different demographic characteristics, including age, gender, and educational attainment. For these results, we focus our analysis on the islands of Java, Sumatera, and Sulawesi, as the road quality data for Kalimantan and parts of eastern Indonesia is limited. Figure 4 shows the locations of IFLS villages used throughout our analyses.

4.1 Welfare Effects

In this section, we provide evidence on the effect of road roughness on welfare, measured with individual earnings and household level consumption. We estimate the following household and individual-level panel regression equations:

$$y_{it} = \alpha_i + \alpha_t + \beta \log \mathrm{IRI}_{d(i)t} + \mathbf{x}'_{it}\theta + \varepsilon_{it} \tag{13}$$

where i indexes individuals (or households), y_{it} is the dependent variable observed for i at time t, α_i is a fixed effect for i, and α_t is a time effect. The variable $\log \mathrm{IRI}_{d(i)t}$ measures the log of the average roughness for all roads in i's district, denoted $d(i)$, at time t. The vector $\mathbf{x}_{it}$ represents a set of time-varying controls, including district-level GDP, household size, and month of survey indicators. For individual-level regressions, controls also include age and education. Unless stated otherwise, all parameters are estimated with GMM, where we instrument $\log \mathrm{IRI}_{d(i)t}$ with the budget shifting instruments described in the previous section. Whenever using GMM together with these instrument, we refer to it as IV-GMM.[10] Standard

[10] Our first stage relationships between road roughness and budget shifters, across districts-years, households-years, and individual years, can be found in Appendix Table A.1.

errors are clustered at the village level.

Using the log of household consumption per capita as the dependent variable, we report the results of (13) in Table 5, Panel A.[11] We find that a one percent increase in road roughness significantly decreases consumption per capita by 0.26 percent. We report the Kleibergen and Paap (2006) Wald rank F-statistic, robust to clustering and heteroskedasticity, which at the value of 35.45 rejects the null hypothesis of weak instruments.

In Panel B, we similarly estimate the relationship between individual log total monthly earnings and road roughness. Total monthly earnings consist of wage earnings and net profits from either primary and/or secondary job, depending on whether you are formally (working for the government or the private sector) or self-employed, respectively.[12] Controlling for hours worked allows us to interpret the estimate of β as changes in effective wages. We find that a one percent increase in road roughness significantly decreases total earnings by 0.89 percent.

In the next subsections, we examine many different mechanisms behind these effects, guided by our theoretical framework.

4.2 Inmigration of Workers and Entry of Firms

To the extent that road quality is a productive amenity, valued by farm producers and formal sector manufacturing firms, our model predicts that road improvements would encourage entry of firms and an in-migration of workers. This could result in additional employment opportunities and allow individuals to move out of agriculture into the formal manufacturing sector, enjoying relatively higher returns than before. Depending on the relative shifts in labor demand and labor supply, welfare could improve directly through increased wages as well.

To examine the hypothesis on in-migration of workers, in Table 6, Panel A, we report IV-GMM estimates of the following cross-sectional regression equation:

$$y_d = \alpha + \beta \Delta \log \mathrm{IRI}_d + \mathbf{x}_d' \theta + \varepsilon_d \tag{14}$$

where y_d is the log of district d's share of recent migrants, calculated as the number of people from Indonesia's Household Census of 2000 in each district who reported living in another

[11]Household consumption per capita is defined as the sum of total monthly expenditures on food (meat, fish, dairy, spices, sugar, oil, beverages, alcohol, tobacco, snacks, eating out, staples, vegetables, dried food) and non-food items (utilities, personal goods, household goods, recreation expenses, transport expenses, lottery, clothers, furniture, medical expenses, ceremonies, taxes, other, rent or other house costs, education).

[12]Family workers are excluded from earnings regressions since their income is reported as missing or zero in most cases.

district five years prior, and $\Delta \log \text{IRI}_d$ is the change in log roughness between 1995 and 2000. Controls in this regression, denoted by $\mathbf{x}_d$, include long lags of non-oil GDP and population, both measured in 1990. In column 2, we control for province fixed effects, too.

Results suggest that a one percent increase in the road deterioriation rate significantly decreased share of migrants by 0.7 percent. For the average district, with about 6% of recent migrant population share, a one percent increase in the rate of road deterioration would result in a reduction of the migrant share to approximately 5.6%. Calculated at the average population of Indonesia's districts ($\approx$ 643,000), this represents a reduction of total recent migrants from 38,594 to 35,892, a decrease of only 2,700 individuals. This is a small effect, which is not surprising given barriers to mobility in Indonesia (Bazzi, 2013).

In Panel B, we use data from Indonesia's Annual Survey of Manufacturing Establishments (*Survei Tahunan Perusahaan Industri Pengolahan*, or SI) to test whether changes in road roughness affected the entry of large manufacturing firms. The SI is an annual census of manufacturing plants with more than 20 employees and contains detailed information on plant's cost variables, their industry of operation, employment size and measures of value added. The data also contains information on the firm's starting date and its location at the district level. We use these data to measure firm entry into the district, where we identify new firms as those that appear in the dataset for the first time.[13]

In column 1, we regress the log of the number of new firms in district d at time t on the log of road roughness in district d at time t, using district-level panel data from 1990 to 2007. We control for district and year fixed effects, controls used in (14). In column 2, the dependent variable is the number of manufacturing jobs, equal to the total number of workers hired by manufactures in the SI.

Our IV-GMM estimates suggest that a one percent increase in road roughness is associated with a 1.2 percent decrease in the number of new manufacturing firms, and a 1.3 percent decrease in the number of manufacturing jobs. This is equvivalent to a reduction of slightly more than 1 firm, and the loss of approximately 215 jobs over only one percent increase in road rougness. From 1990 to 2000, the average district experienced an average reduction in roughness of approximately 46%. From the Panel B estimates, this should have led to a 55.2% increase in the number of firms (42.3 more firms, on average), and a 59.8% increase in the number of workers (7,700 more workers, on average).

In summary, we find that road improvements trigger the entry of manufacturing firms and increase local labor demand. We also find that the worker's mobility due to better roads was much smaller than the one from firms. Next, we use data on firms directly to examine

[13]We dropped all firms coded as state-owned enterprises (less than 3 percent of all firm-year observations). Also, throughout the discussion, we use plants and firms interchangeably since less than 5% of plants in the dataset are operated by multi-plant firms (Blalock and Gertler, 2008).

the within-firm effect of road improvements to further explore the mechanism behind the significant labor demand increase.

4.3 Firm-Level Outcomes

An advantage of using the SI data is that it contains firm-level identifiers, allowing researchers to track changes in firm-level outcomes over time. In Table 7, we estimate the relationship between road roughness and log firm-level employment (Panel A), log value added (Panel B), and log total factor productivity (TFP) (Panel C). Employment and value added measures are recorded directly in the SI data. We use a control-function approach to calculate TFP through productivity residuals estimation (e.g. Olley and Pakes, 1996; Levinsohn and Petrin, 2003).[14] Our regressions take the same form as (13), with firm-year observations ranging from 1990 to 2007.

In Column 1 of Table 7, we estimate regressions using only year, industry, and district fixed effects. We find that a one percent increase in road roughness translates to a 0.06 percent decrease in the number of workers, a 0.14 percent decrease in log value added, and a 0.10 percent decrease in total factor productivity. However, in Column 2, when we condition on firm fixed effects as well, the significant effects of road roughness on total workers disappear. This suggests that changes in labor demand come from the entry of new firms, rather than changes in labor demand for existing firms. This may be consistent with a perfectly competitive environment, or it may also reflect non-trivial adjustment costs; as road quality deteriorates, firms may want to be using less labor, but that would not be possible without reductions in capital, which may be costly. However, even with firm fixed effects, road deterioration reduces firms' value added and total factor productivity.

4.4 Labor Supply: Extensive and Intensive Margins

Another way better roads could increase total earnings and consumption is through increased labor supply. Individuals might work more hours, transition out of unemployment as a result of increased labor demand and employment opportunities. In Table 8, Panel A, we report results of estimating (13) using when y_{it} is defined as an indicator for whether or not individual i reported working at time t. Individuals were classified as working if in the last week, they reported actually working, trying to work, helping to earn income, worked at least one hour for pay, had a business and were usually working but just not last week, or worked on a family owned business (farm or nonfarm), and reported their occupation

[14]Precise details on the implementation of this approach for the SI data can be found in Poczter et al. (2014).

in either primary or/and secondary job. The indicator is only defined for individuals older than 15 years. Approximately two thirds of the individual-year observations in the IFLS data reported working last week. IV-GMM estimates show that road improvements have no significant effect on probability of working at conventional significance levels.[15]

In Panel B, we report estimates of the effect of roughness on working hours. In this regression, both hours worked and roughness are measured in logs. We again statistically insignificant effects of road roughness on hours worked.[16] Taken together, the evidence from Table 8 suggests that changes in road roughness do not improve total earnings or consumption through their effects on the extensive or intensive margins of labor supply.

4.5 Sector Switching

Next, we investigate the extent to which changes in road roughness result in sector switching, movement out of agriculture or the informal sector and into formal sector or manufacturing employment. In Table 9, we report estimates of the relationship between log average road roughness and measures of employment and hours worked in different sectors, using our individual-level panel regression specification (13). Across the columns of this table, we vary the dependent variables in the regression. In column 1, we use an indicator variable for whether the individual reported any employment in that sector. In column 2, we modify that indicator to be equal to one only if the individual reported that sector as his or her primary occupation, and in column 3, we use the log of total hours worked by sector as the dependent variable. Along the rows, we vary the sector (in Panel A) or use a measure to distinguish between formal and informal employment (Panel B).[17] As before, we report GMM estimates which use the budget-share instruments, and cluster standard errors at the village level.

Despite not finding any effects on total hours worked or on the probability of being employed, we find strong evidence that changes in road roughness affect how individuals sort into sectors of employment. From the first row, we find that increases in road roughness reduce employment and hours worked in manufacturing. A one percent increase in road roughness is associated with a 0.1 percent reduction in reporting working in manufacturing and a 6.2 percent reduction in hours worked in manufacturing.

[15]The confidence interval around the point estimate reported in Panel A, Column 2 implies that if roughness increased by one standard deviation (1.887, or 0.635 log points), we could reject the hypothesis that employment would decrease by more than 3 percent, a moderate effect size.

[16]Note that this insignificance is not due to a lack of statistical power. From column 2, the confidence interval around the point estimate implies that we could reject the hypothesis that a one percent increase in road roughness causes more than a 0.15 percent reduction in hours worked.

[17]Formal and informal employment defined in Section 4.1

We find limited evidence that individuals sort out of agriculture and into manufacturing when roads get better (row 2). Although the coefficients are positive, they are imprecisely estimated. Instead, on average, individuals seem to be moving out of sales and services, and into the manufacturing sector, when roads get better. A one percent increase in road roughness is associated with a 0.08 percent increase in the probability of reporting sales and services as a primary occupation, and a 4.5 percent increase in hours worked in the sales and services sector. The sales and services sector, as we define it, tends to include very low-skill jobs, such as becak drivers, street vendors, and hairdressers.

To the extent that this sort of informal, marginal employment is responsible for suplus labor in poor road quality villages, these results confirm the predictions of the model, particularly the dual economy interpretation. There were no significant substitution effects reported for the other sector. Given that we have already shown that road improvements were accompanied by an entry of manufacturing firms, this suggests that with increase in local labor demand, individuals respond by increasing their employment in manufacturing, and transitioning out of sales and services, and possibly agriculture.

In Panel B, we capture these same patterns by measuring switching between the formal and informal sector. A one percent increase in road roughness is associated with a 0.08 percent increase in the probability of working in the informal sector, and a 6 percent increase in hours worked in the informal sector.

In Table 10, we investigate heterogeneity in the relationship between working in manufacturing and agriculture and different individual and household characteristics. Column 1 repeats the estimates reported in Table 9, Column 2, using all individual-year observations. In columns 2 and 11, we split the sample by five different characteristics: age (above and below median, columns 2 and 3), gender (columns 4 and 5), education (above and below median, columns 6 and 7), initial household per-capita consumption expenditures (above and below median, column 8 and 9), and whether the household owned any farmland in the first year of the data (columns 11 and 12).

Despite finding that increases in road roughness had no average effects on the probability of working in agriculture, we find from Panel A significant positive effects for older individuals, less educated workers, males, and poorer workers (below-median PCE). For these individuals, increases in road roughess result in increased employment in agriculture. This suggests that the most marginal individuals are those who are most likely to be impacted by changes in road roughness.

In Panel B, we find that increases in road roughness reduce employment in manufacturing more for older workers, less educated workers, males, and poorer workers (below-median PCE). This suggests that when road improvements encourage the entry of formal sector firms, they can absorb the labor of lower skilled, marginal workers, and these individuals could benefit substantially from road improvements.

4.6 Total Earnings by Sector

As discussed in Section 2, improvements in road quality may have ambiguous effects on formal sector wages. In the short run, if firms are more mobile than workers, increases in local labor demand should offset increases in local labor supply, generating upward pressure on wages. We provide some evidence for the relative immobility of labor in Section **??**. However, our model also suggests that dual labor market effects may suppress wages, to the extent that in areas with very poor market access, opportunities for outside employment are limited, and there is a surplus of labor in agriculture.

In Table 11, we provide estimates of the reduced form relationship between wages and road roughness, differentiating by sector, using a regression specification similiar to (13). As in Section **??**, we use as dependent variables log total earnings by sector and control for total hours worked so that we can interpret the effects of changes in roughness as changes in wages.

From column 1, we see, as already mentioned in Section **??**, that a one percent increase in road roughness is associated with a significant 0.89 percent decrease in total earnings. However, when we separately estimate the relationship between road roughness and earnings by sector, we find that most of this effect comes from agricultural earnings (column 3). Although the coefficients on road roughness are negative for all other sectors, including manufacturing earnings (column 2), sales and services earnings (column 4), and other earnings (column 5), these coefficients are not significantly different from zero. However, for agriculture, the effect is large and statistically signficant. A one percent increase in road roughness is associated with a 3.4 percent decrease in agricultural earnings, conditional on hours worked. As described in the model, there are many ways in which agricultural earnings and farm profits could increase because of better roads, including reductions in the costs of inputs and higher output prices. Another plausible explanation is that as surplus labor farms reduce agricultural labor because of better outside options, farm profits and total earnings per hours worked increases.

4.7 Land Prices and Land Rents

As firms and workers move into areas affected by road quality improvements, they increase the local demand for land and housing, and we would expect to see increases in the value of land in those areas. To provide evidence for this, we use household-level panel data from the IFLS on land values and land rents and estimate hedonic rent regressions (e.g. Blomquist et al., 1988; Roback, 1982). Our regression specifications are similiar to (13), but when using household-level data, we add a vector of housing controls, including variables for the number of rooms, indicators for access to electricity, piped water, internal toilets, and types of floor, walls, and roof material, in addition to our standard controls (district-level non-oil GDP and month or survey indicators). Standard errors are clustered at the village level.

Table 12 reports the results. In column 1, the dependent variable is the log of the land rent per-room. This variable is measured as the actual rent paid or, if rents were not paid, the monthly rent that the household estimates they would have been paid if they had been renting. Note that much of the data on land rents is imputed. Of the 33,049 household-years in the IFLS with rent data, 29,508 observations (89.3%) are based on imputed rents, with the remaining 3,541 observations (10.7%) representing rents actually paid. Because of this, we also control for an indicator for whether observations are based on actual rents. The coefficient is negative and statistically significant, and suggests that a one percent increase in average road roughness leads to a 0.24 percent decrease in rents.

In column 2, we use a different measure of land prices, based on self-reported land values. In the IFLS, households are asked to report the total market value for a series of different types of land they own, including land used for a farm business, land used for a non-farm business, and other owned farm or non-farm land not being used for business. This measure has limitations because it is based on self-reported data, instead of market transactions. Another issue with this measure is that because individuals were not asked about the size of their land holdings, we cannot control for total hectares owned. This leaves open the possibility that differences in total land values across households could be explained by differences in land sizes. Nevertheless, similar to the results in column 1, the coefficient on average road roughness is negative and statistically significant, suggesting that a one percent increase in average road roughness leads to a 0.66 percent decrease in land values. Taken together, both measures of land prices, despite their imperfections, provide substantial evidence for the notion that road improvements lead to increases in the value of land.

4.8 Robustness

In this section, we have only focused on presenting our main results, in the interest of simplifying the exposition. However, in Appendix Section A, we subject our results to a number of robustness checks. We first present fixed-effects least squares results and find that our GMM estimates are quite similar. We also estimate effects separately for rural areas, and although our sample sizes fall and standard errors become less precise, the point estimates are always quite similiar. We also estimate effects separately for non-moving households; in our main specifications, changes in outcomes could entirely be a product of the fact that individuals were moving between locations, but we find no strong evidence of this. We also use different measures of road quality, including a roughness-induced travel time measure between village v and that village's nearest provincial capital, and measures of market access.[18] Finally, we provide GMM estimates using a Hausman IV, where we instrument district d's road quality with the average road quality of all other districts in the same province. In general, our results are very robust to these different robustness checks.

4.9 Decomposition of Total Earnings Effects

TO BE COMPLETED.

Let Y denote a consumer's total earnings. This is just the total hours worked in each sector, multiplied by the wage assigned to each sector:

$$Y = \sum_{s=1}^{S} w_s H_s$$

If we consider a marginal change in road quality, we can totally differentiate this expression with respect to A and obtain:

$$\frac{\partial Y}{\partial A} = \sum_{s=1}^{S} w_s \left(\frac{\partial H_s}{\partial A} \right) + H_s \left(\frac{\partial w_s}{\partial A} \right)$$

Expressing this in terms of elasticities, we have:

$$E_{Y,A} = \sum_{s=1}^{S} \left(\frac{w_s H_s}{Y} \right) E_{H_s,A} + \left(\frac{w_s H_s}{Y} \right) E_{w_s,A}$$

where $E_{y,x} = \partial \log y / \partial \log x$ is the elasticity of y with respect to x. So, we can decompose the percent change in income with respect to road improvements to be an income share weighted sum of elasticities of hours worked and wages with respect to road improvements.

5 Conclusion

Even though transportation infrastructure investments typically account for a significant proportion of countries' budgets, little is known about their effects in developing countries, where spatial disparities are particularly pronounced. This paper aims to understand the role

[18]For more on calculating the roughness-induced travel time measure and measures of market access, see Rothenberg (2013).

road improvement (or deterioration) can play in such countries, not only through looking at possible welfare effects, but also by investigating the different possible mechanisms through which these effects materialize. While much of the previous literature on this topic has focused on the construction of new roads, we add to the literature by evaluating the effects of substantial changes in road quality due to maintenance and upgrading of already existing roads in Indonesia.

Using a novel dataset that documents substantial variation in road quality in Indonesia, and combining this with high quality household panel data that spans years 1990 through 2007, we provide reduced form evidence that road improvements significantly increase welfare, measured either with consumption or income. Additionally, using an annual census of manufacturing firms, we show that these positive welfare effects partially materialize though increased labor market demand, generated by the entry of new firms rather than extended hiring by existing firms. However, we do not see substantial changes in the extensive or intensive margin of labor supply, but instead observe occupational shifts from agriculture into higher paying, newly available manufacturing jobs. In addition, while manufacturing wages typically don't exhibit an upward push, we do observe significant improvements in agricultural profits. This not only implies the wage gap between these two sectors is narrowed, but also confirms the predictions of our stylized model of dual labor markets. The latter shows under what conditions productive amenities, such as transport infrastructure, may translate into positive welfare effects.

The methodological contribution of this paper is in addressing the common concerns of targeting bias and spillover effects by suggesting a new instrument, replicable in many instances. We take advantage of Indonesia's institutional two-step budgeting setup for road funding, where different authorities, such as provinces or districts, are in charge of road quality and funding of different parts of the road network. This allows us to construct a time varying instrument for road quality, which equals total road funding at the provincial or district level. Thus, we identify the effects from the set of roads that get maintained when road budgets allow for it, but which get less maintenance when road budgets are tight or scaled back.

The evidence presented in this paper shows that road improvements alone can present an important stepping stone in economic development through opening up labor market opportunities and decreasing the income gap at the same time. On the flip side, deterioration of roads may have adverse affects in the opposite direction and may bring about important and unanticipated welfare effects that governments should be aware of when cutting transportation budgets.

References

Anderson, J. E. and E. Van Wincoop (2004): "Trade Costs," *Journal of Economic Literature*, 42, 691–751.

Asia Foundation (2008): "The Cost of Moving Goods: Road Transportation, Regulations and Charges in Indonesia," Survey report.

Bank, W. (2014): "World Bank: Indonesia @ONLINE," .

Bardhan, P. and C. Udry (1999): *Development Microeconomics*, Oxford: Oxford University Press.

Bazzi, S. (2013): "Wealth Heterogeneity, Income Shocks, and International Migration: Theory and Evidence from Indonesia," Working Paper.

Benjamin, D. (1992): "Household Composition, Labor markets, and Labor demand: Testing for Separation in Agricultural Household Models," *Econometrica*, 60, 287–322.

Bennett, C. R., A. Chamorro, C. Chen, H. de Solminihac, and G. W. Flintsch (2007): "Data Collection Technologies for Road Management," Technical report, East Asia Pacific Transport Unit, World Bank.

Blalock, G. and P. J. Gertler (2008): "Welfare Gains from Foreign Direct Investment Through Technology Transfer to Local Suppliers," *Journal of International Economics*, 74, 402–421.

Blomquist, G. C., M. C. Berger, and J. P. Hoehn (1988): "New Estimates of Quality of Life in Urban Areas," *American Economic Review*, 78.

Deaton, A. and J. Muellbauer (1980): "An Almost Ideal Demand System," *American Economic Review*, 70, 312–326.

Dickens, W. T. and K. Lang (1985): "A Test of Dual Labor Market Theory," *American Economic Review*, 75, 792–805.

Donaldson, D. (2010): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," Working Paper.

Fallon, P. R. and R. E. Lucas (2002): "The impact of financial crises on labor markets, household incomes, and poverty: A review of evidence," *The World Bank Research Observer*, 17, 21–45.

Hill, H. (2000): *The Indonesian Economy*, Cambridge: Cambridge University Press.

Islam, R. (2004): *The nexus of economic growth, employment and poverty reduction: An empirical analysis*, Recovery and Reconstruction Department, International Labour Office.

Jacoby, H. G. (2000): "Access to Markets and the Benefits of Rural Roads," *The Economic Journal*, 110, 713–737.

Kleibergen, F. and R. Paap (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133, 97–126.

Levinsohn, J. and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–342.

Manning, C. (2000): "Labour market adjustment to Indonesia's economic crisis: context, trends and implications," *Bulletin of Indonesian Economic Studies*, 36, 105–136.

Matsumoto, M. and S. Verick (2011): "Employment trends in Indonesia over 1996-2009: casualization of the labour market during an era of crises, reforms and recovery," Tech. rep., International Labour Organization.

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Poczter, S., P. Gertler, and A. D. Rothenberg (2014): "Financial Crisis and Productivity Evolution: Evidence from Indonesia," *The World Economy*, 37, 705–731.

Roback, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90, 1257–1278.

Rosen, S. (1979): "Wage-Based Indexes of Urban Quality of Life," in *Current Issues in Urban Economics*, ed. by P. Mieszkowski and M. Straszheim, Johns Hopkins University Press.

Rothenberg, A. D. (2013): "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia's Highways," Working Paper.

Sen, A. K. (1966): "Peasants and Dualism With and Without Surplus Labor," *Journal of Political Economy*, 74, 425–450.

Singh, I., L. Squire, and J. Strauss (1986): *Agricultural Household Models: Extensions, Applications, and Policy*, Baltimore: Johns Hopkins University Press.

World Bank (2012): *Investing in Indonesia's roads : improving efficiency and closing the financing gap.*

Yu, J., E. Chou, and J. Yau (2006): "Development of Speed-Related Ride Quality Thresholds using International Roughness Index," *Transportation Research Record*, 1974, 47–53.

Table 1: THE DETERMINANTS OF ROAD QUALITY

Dep. Var.: Road Segment $\log \mathrm{IRI}_{\mathbf{t}}$	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\log \mathrm{IRI}_{t-1}$	0.831	0.852	0.857	0.616	0.698	0.708
	(0.001)***	(0.004)***	(0.006)***	(0.001)***	(0.005)***	(0.006)***
Segment $\text{Upgraded}_{t-1} = 1$	-0.484	-0.492	-0.492	-0.450	-0.449	-0.449
	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
$\log \mathrm{GDP}_{t-1}$			0.005			0.008
			(0.001)***			(0.002)***
$\log \mathrm{GDP}_{t-1} \times \mathrm{IRI}_{t-1}$		-0.002	-0.003		-0.006	-0.006
		(0.000)***	(0.000)***		(0.000)***	(0.000)***
$\log \#\mathrm{Firms}_{t-1}$			-0.007			0.010
			(0.001)***			(0.002)***
$\log \#\mathrm{Firms}_{t-1} \times \mathrm{IRI}_{t-1}$		-0.001	0.002		-0.002	-0.005
		(0.000)***	(0.001)***		(0.001)***	(0.001)***
N	972714	602872	602872	972714	602872	602872
Adjusted R^2	0.811	0.797	0.797	0.513	0.499	0.499
F Statistic	256469.40	148866.70	136271.68	31242.55	19506.45	17776.15
Road FE	No	No	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

GDP and # firms are measured at the district level. Robust standard errors in parentheses, clustered at the road segment level. */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_roadQuality.do`.

Table 2: THE DETERMINANTS OF ROAD INVESTMENT

	All Roads	National Roads		Provincial Roads		District Roads	
Dep. Var.: Segment Upgrade$_t$ (0 1)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
$\log \text{IRI}_{t-2}$	0.017	-0.014	-0.017	-0.005	-0.009	0.023	0.022
	(0.001)***	(0.002)***	(0.002)***	(0.001)***	(0.001)***	(0.002)***	(0.003)***
Years Since Last Upgrade	-0.335	-0.436	-0.478	-0.382	-0.427	-0.258	-0.291
	(0.001)***	(0.006)***	(0.006)***	(0.004)***	(0.006)***	(0.004)***	(0.005)***
Years Since Last Upgrade^2	0.026	0.039	0.046	0.033	0.040	0.018	0.022
	(0.000)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.000)***	(0.001)***
$\log \text{GDP}_{t-1}$		-0.005		-0.002		0.001	
		(0.001)***		(0.001)**		(0.001)	
$\log \#\text{Firms}_{t-1}$			-0.002		-0.003		0.001
			(0.000)***		(0.000)***		(0.001)
N	774719	237220	216508	265443	238341	27981	25165
Adjusted R^2	0.627	0.686	0.708	0.673	0.693	0.635	0.657
F Statistic	45361.69	20743.20	21132.55	28403.58	32178.52	5284.67	5374.47
Road FE	No	No	No	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses, clustered at the road segment level. */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_roadQuality.do`.

Table 3: Road Quality and Budgets

Dep. Var.: Road segment log IRI(t)	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Budget (National Roads)	-0.044			-0.047		
	(0.001)***			(0.001)***		
Budget (Provincial Roads)		-0.037			-0.026	
		(0.000)***			(0.000)***	
Budget (District Roads)			-0.015			0.006
			(0.001)***			(0.001)***
$\log \text{IRI}_{t-1}$				0.415	0.411	0.294
				(0.002)***	(0.002)***	(0.005)***
Years Since Last Upgrade				0.164	0.165	0.137
				(0.001)***	(0.001)***	(0.001)***
Years Since Last Upgrade^2				-0.012	-0.012	-0.009
				(0.000)***	(0.000)***	(0.000)***
N	955214	960837	306578	766335	769723	227814
Adjusted R^2	0.078	0.079	0.039	0.388	0.386	0.259
F Statistic	2520.469	2734.166	484.572	11979.469	12074.867	1777.209
Road FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

GDP and # firms at the district level. Robust standard errors in parentheses, clustered at the road segment level. */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_roadQuality.do`.

Table 4: Road Budgets and Local Economic Conditions

Dep. Var.: $\log \text{Budget}_t$	**Province Panel**				**District Panel**	
	Nat. roads	Nat. roads	Prov. roads	Prov. roads	District roads	District roads
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\log \text{GDP}_{t-2}$	-0.080		-0.584		-0.087	
	(0.185)		(0.506)		(0.118)	
$\log \#\text{Firms}_{t-2}$		-0.080		-0.291		-0.086
		(0.124)		(0.284)		(0.065)
N	241	256	249	265	545	568
Adjusted R^2	0.319	0.310	0.389	0.363	0.091	0.097
Road FE	No	No	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses, clustered at the road segment level. */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_roadQuality.do`.

Table 5: Reduced Form Effects of Road Roughness: Consumption and Total Earnings

Panel A: Dep Var: Log Consumption (Per Capita)	**(1)**
Log Avg IRI	-0.261
	(0.061)***
Adj. R^2	0.594
Regression F-Stat	191.771
Kleibergen-Paap F-Stat	35.450
DV Mean	11.046
N	22541
N Households	6928
N Desa / Kelurahan	852

Panel B: Dep Var: Log Total Earnings	**(1)**
Log Avg IRI	-0.894
	(0.328)***
Adj. R^2	0.265
Regression F-Stat	10.118
Kleibergen-Paap F-Stat	33.684
DV Mean	10.633
N	16935
N Individuals	6460
N Desa / Kelurahan	298
Household FE	Yes
Year FE	Yes
Controls	Yes

All regressions include a constant. Controls include: district GDP, month of survey indicators. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_hh.do`.

Table 6: Road Roughness and District-Level In-Migration of Workers and Entry of Firms

Panel A: Dep Var: Log Share of Recent Migrants (2000 Census)	(1)	(2)
Log Δ IRI (2000-1995)	-1.257	-0.690
	(0.153)***	(0.225)***
N	201	201
Adjusted R^2	0.366	0.474
F Statistic	38.992	8.436
Kleibergen-Paap F-Stat	14.779	11.570
DV Mean	-3.156	-3.156
Province FE	.	Yes

Panel B: Firm Outcomes	DV: Log # of Firms	DV: Log # of Workers
	(1)	(2)
Log Avg IRI	-1.194	-1.294
	(0.182)***	(0.286)***
N	3121	3121
N Kabupaten / Kota	204	204
Adj. R^2	0.910	0.914
Regression F-Stat	110.553	241.978
Kleibergen-Paap F-Stat	47.239	47.239
DV Mean	2.890	6.660
Kabupaten / Kota FE	Yes	Yes
Year FE	Yes	Yes

All regressions include a constant. Panel A reports the results of cross-sectional regressions of the share of recent migrants (last five years) in 2000 on changes in road roughness. In Panel A, the unit of analysis is the kabupaten (district), and controls include logs of 1990 population and 1990 non-oil GDRP. Robust standard errors are reported in parentheses. In Panel B, we report the results of panel regressions of road roughness on the log of one plus the number of firms and workers, with data from the SI. Controls include logs of population and non-oil GDRP. Robust standard errors, clustered at the kabupaten level, are reported in parentheses. For the IV specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_kabuYear.do`.

Table 7: Road Roughness and Firm-Level Outcomes

Panel A: Dep Var: Log Number of Workers	**(1)**	**(2)**
Log Avg IRI	-0.057	0.003
	(0.017)***	(0.012)
N	234272	230224
N Firms	29366	29366
Adjusted R^2	0.266	0.924
Regression F Statistic	11.563	0.069
Kleibergen-Paap F-Stat	2034.234	1821.897
Panel B: Dep Var: Log Value Added	**(1)**	**(2)**
Log Avg IRI	-0.138	-0.074
	(0.036)***	(0.031)**
N	234134	230096
N Firms	29362	29362
Adjusted R^2	0.467	0.859
Regression F Statistic	15.048	5.624
Kleibergen-Paap F-Stat	2032.975	1821.119
Panel C: Dep Var: Log TFP	**(1)**	**(2)**
Log Avg IRI	-0.099	-0.134
	(0.040)**	(0.037)***
N	126629	122859
N Firms	19073	19073
Adjusted R^2	0.480	0.725
Regression F Statistic	6.255	13.041
Kleibergen-Paap F-Stat	1190.051	1010.144
Year FE	Yes	Yes
Industry FE	Yes	Yes
Kabupaten / Kota FE	Yes	Yes
Firm FE	.	Yes

All regressions include a constant. Robust standard errors in parentheses, clustered at the firm level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_SI_indiv.do`.

Table 8: Road Roughness and Labor Supply: Extensive and Intensive Margins

Panel A: Dep Var: Working? (0 1)	**(1)**
Log Avg IRI	-0.037
	(0.034)
Adj. R^2	0.427
Regression F-Stat	34.111
Kleibergen-Paap Wald rk F-Stat	33.594
DV Mean	0.704
N	35193
N Individuals	12459
N Desa / Kelurahan	377

Panel B: Dep Var: Log Hours Worked (Conditional on working)	**(1)**
Log Avg IRI	-0.011
	(0.060)
Adj. R^2	0.237
Regression F-Stat	3.461
Kleibergen-Paap Wald rk F-Stat	32.698
DV Mean	5.148
N	22255
N Individuals	8168
N Desa / Kelurahan	321
Individual FE	Yes
Year FE	Yes
Controls	Yes

All regressions include a constant. Controls include: district GDP, individual age, education, household size, and month of survey indicators. In Panel B, the dependent variable is only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_indiv.do`.

Table 9: ROAD ROUGHNESS AND SECTOR SWITCHING

Dependent Variable:	**Any Employment?**	**Primary Sector Employment?**	**Log Hours Worked**
Panel A: By Industry	**(1)**	**(2)**	**(3)**
Manufacturing	-0.114	-0.129	-6.240
	(0.036)***	(0.033)***	(1.802)***
Adj. R^2	0.462	0.493	0.395
Regression F-Stat	8.807	10.079	8.151
Kleibergen-Paap F-Stat	32.711	32.711	32.910
DV Mean	0.285	0.253	12.219
Agriculture	0.055	0.055	2.889
	(0.033)*	(0.034)	(2.494)
Adj. R^2	0.648	0.653	0.399
Regression F-Stat	6.559	4.923	2.887
Kleibergen-Paap F-Stat	32.711	32.711	32.695
DV Mean	0.419	0.367	13.013
Sales and Services	0.054	0.082	4.523
	(0.035)	(0.035)**	(1.963)**
Adj. R^2	0.498	0.533	0.443
Regression F-Stat	8.511	6.062	7.049
Kleibergen-Paap F-Stat	32.711	32.711	32.930
DV Mean	0.316	0.279	13.688
Other	-0.026	-0.015	0.026
	(0.018)	(0.011)	(0.802)
Adj. R^2	0.608	0.521	0.474
Regression F-Stat	1.903	3.954	2.160
Kleibergen-Paap F-Stat	32.711	33.594	32.771
DV Mean	0.105	0.067	3.646
Panel B: Formal vs. Informal	**(1)**	**(2)**	**(3)**
Informal	0.079	0.091	6.074
	(0.033)**	(0.032)***	(2.592)**
Adj. R^2	0.560	0.598	0.366
Regression F-Stat	11.468	6.857	7.955
Kleibergen-Paap F-Stat	32.698	32.707	33.050
DV Mean	0.636	0.576	23.298
Individual FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_indiv.do`.

Table 10: Road Roughness and Sector Switching: Heterogeneity

		Age		Gender		Education		Initial PCE		Own Farmland?	
Panel A: Dep Var: Working in Agriculture (Primary) (0 1)	**All Obs (1)**	**$<$ Med (2)**	**$\geq$ Med (3)**	**F (4)**	**M (5)**	**$<$ Med (6)**	**$\geq$ Med (7)**	**$<$ Med (8)**	**$\geq$ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	0.055	0.034	0.074	0.034	0.082	0.116	-0.002	0.100	-0.008	0.118	0.079
	(0.034)	(0.082)	(0.034)**	(0.042)	(0.043)*	(0.047)**	(0.038)	(0.046)**	(0.038)	(0.036)***	(0.054)
Adj. R^2	0.664	0.579	0.678	0.708	0.640	0.646	0.627	0.650	0.630	0.633	0.622
Regression F-Stat	4.923	2.786	4.627	2.276	4.244	4.952	1.668	4.443	3.226	3.042	4.107
Kleibergen-Paap Wald rk F-Stat	32.711	19.528	33.321	31.895	31.717	26.469	30.785	27.690	29.354	26.561	26.547
DV Mean	0.367	0.314	0.382	0.328	0.392	0.491	0.212	0.494	0.226	0.197	0.529

		Age		Gender		Education		Initial PCE		Own Farmland?	
Panel B: Dep Var: Working in Manufacturing (Primary) (0 1)	**All Obs (1)**	**$<$ Med (2)**	**$\geq$ Med (3)**	**F (4)**	**M (5)**	**$<$ Med (6)**	**$\geq$ Med (7)**	**$<$ Med (8)**	**$\geq$ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	-0.129	-0.129	-0.134	-0.117	-0.138	-0.207	-0.073	-0.197	-0.071	-0.145	-0.157
	(0.033)***	(0.080)	(0.035)***	(0.044)***	(0.041)***	(0.044)***	(0.041)*	(0.046)***	(0.042)*	(0.054)***	(0.040)***
Adj. R^2	0.490	0.392	0.493	0.543	0.461	0.505	0.468	0.528	0.471	0.484	0.479
Regression F-Stat	10.079	2.862	10.382	3.439	8.778	10.285	4.021	7.270	4.635	9.754	5.366
Kleibergen-Paap Wald rk F-Stat	32.711	19.528	33.321	31.895	31.717	26.469	30.785	27.690	29.354	26.561	26.547
DV Mean	0.253	0.308	0.237	0.185	0.298	0.241	0.268	0.248	0.258	0.324	0.185
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_heterogeneity.do`.

Table 11: Road Roughness and Total Earnings by Sector

	Dep Var: Earnings by Sector			
	Manu-facturing (1)	**Agri-culture (2)**	**Sales and Services (3)**	**Other (4)**
Log Avg IRI	-0.533	-3.474	-0.160	-0.132
	(0.363)	(1.224)***	(0.404)	(0.201)
Adj. R^2	0.282	0.125	0.197	0.066
Regression F-Stat	1.818	2.181	2.182	1.940
Kleibergen-Paap F-Stat	23.238	19.200	24.000	32.073
DV Mean	11.177	7.806	11.024	0.135
N	4160	5464	3692	10763
N Individuals	1722	2130	1540	4235
N Desa / Kelurahan	227	189	231	260
Individual FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

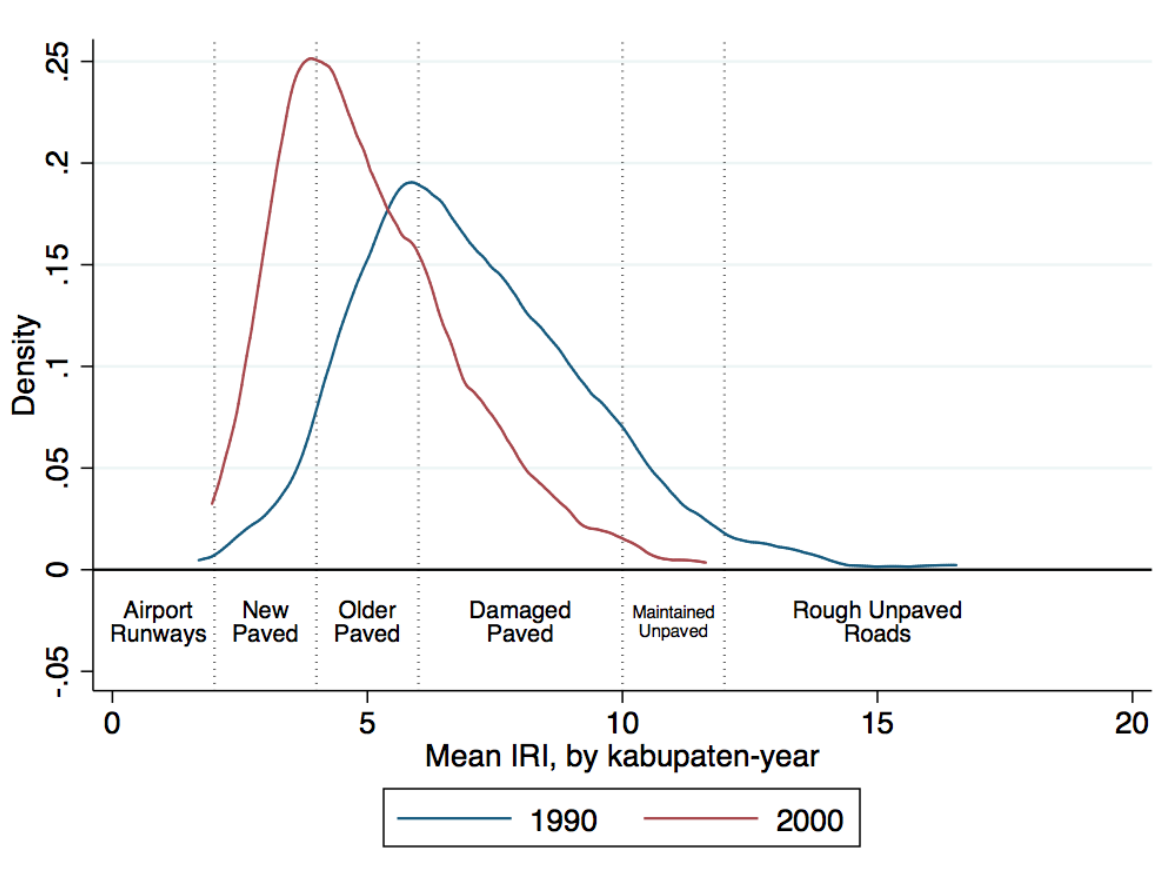
All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators, and hours worked. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_indiv.do`.

Table 12: Road Roughness and Land Prices

	DV: Log Rent per Room (1)	**DV: Log Land Value (2)**
Log Avg IRI	-0.244	-0.658
	(0.116)**	(0.217)***
Adj. R^2	0.360	0.488
Regression F-Stat	25.589	12.117
Kleibergen-Paap F-Stat	36.627	19.970
DV Mean	8.276	14.819
N	19115	7441
N Households	6475	2498
N Desa / Kelurahan	793	321
Household FE	Yes	Yes
Year FE	Yes	Yes
Controls	Yes	Yes

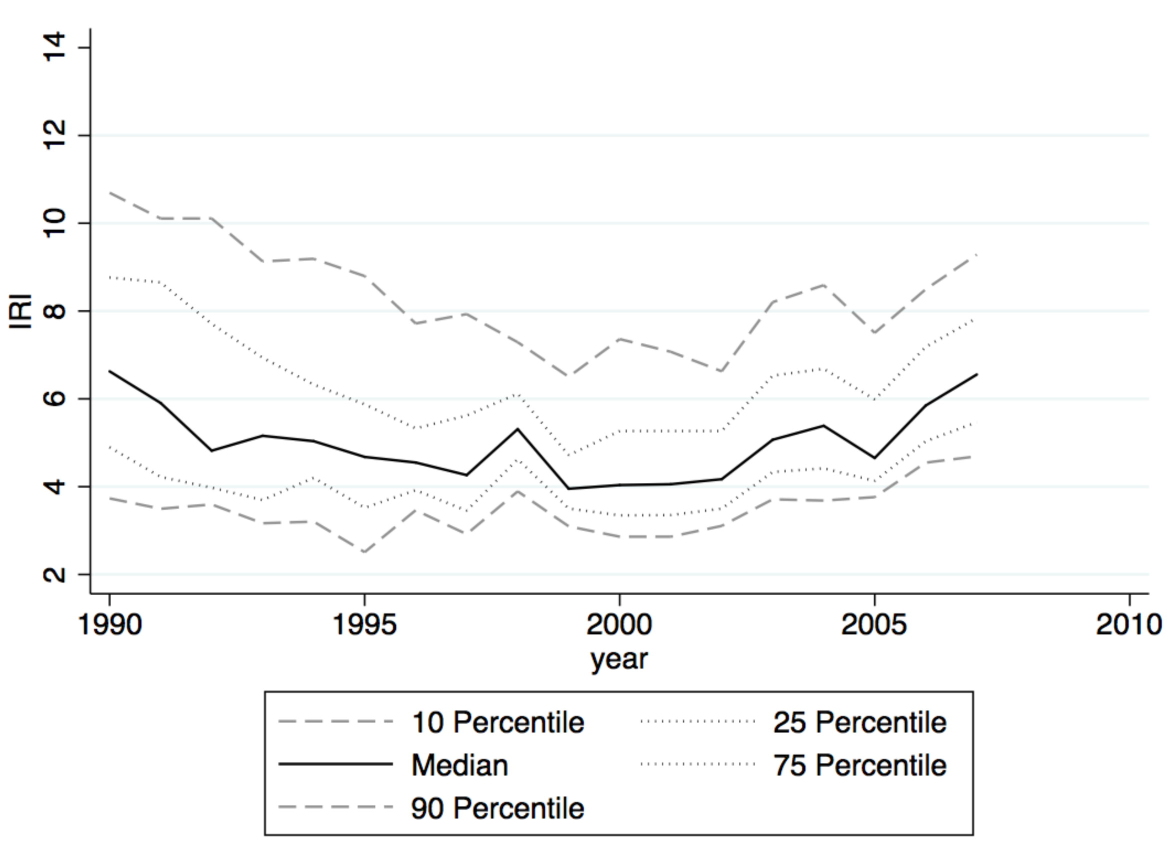
All regressions include a constant. Controls include: district GDP, month of survey indicators, number of rooms, indicators for electricity, piped water, internal toilets, types of floor (tiled, cement, dirt), mason walls, and tiled roof. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_hh.do`.

Figure 1: Changes in the Distribution of Road Roughness



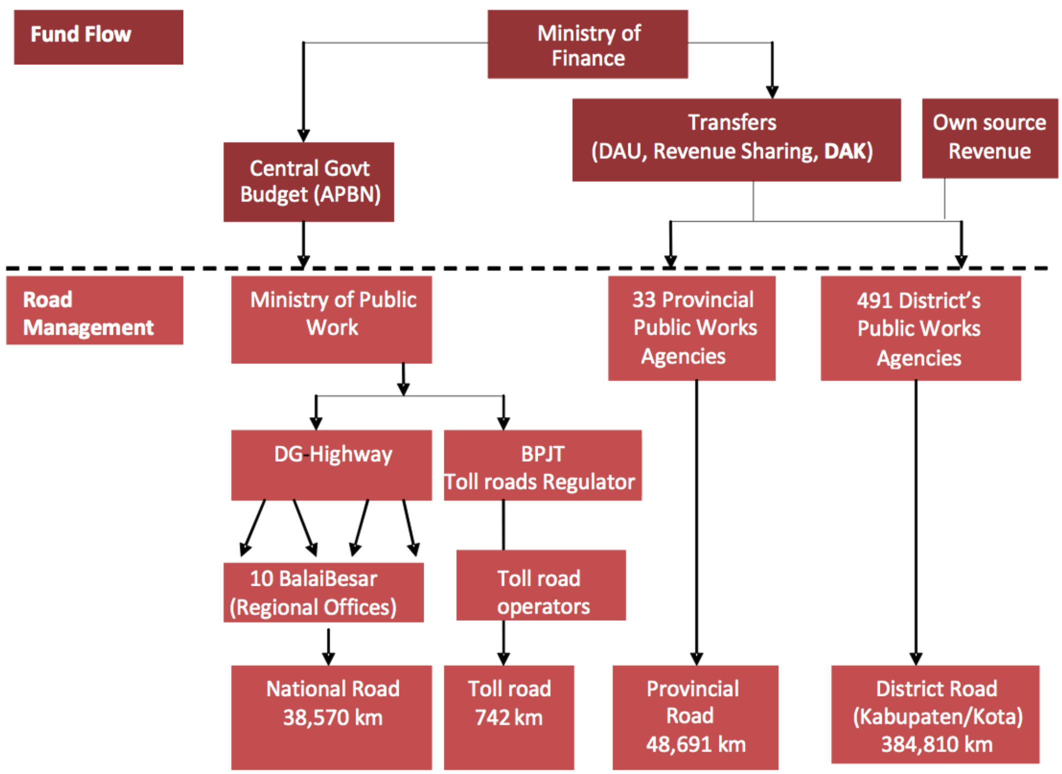
Note: Authors' calculations.

Figure 2: Changes in Roughness-Based Travel Time



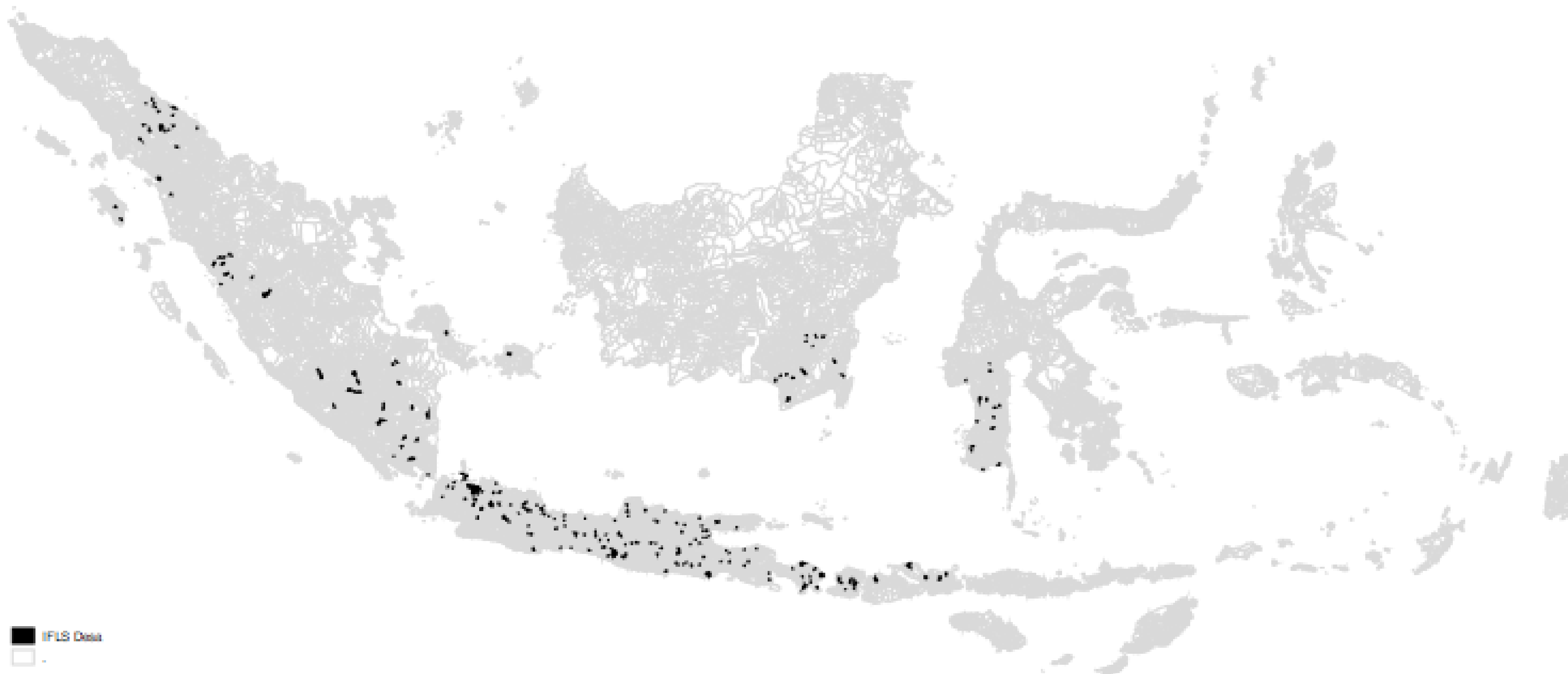
Note: Authors' calculations.

Figure 3: Institutional Arrangements for the Road Sector in Indonesia



Source: (World Bank, 2012).

Figure 4: IFLS Villages



Note: Authors' calculations.

A Robustess Tables

Table A.1: Budget IV First Stage

	Kabupatens (1)	**Households (2)**	**Individuals (3)**
Log Budget Proxy (Nat'l Roads)	0.122 (0.010)***	0.123 (0.024)***	0.169 (0.012)***
Log Budget Proxy (Prov Roads)		0.046 (0.025)*	
Log Budget Proxy (Kabu Roads)	-0.023 (0.004)***		
Log Budget Proxy (Nat'l Roads, t-1)	0.091 (0.012)***	-0.036 (0.014)***	-0.054 (0.015)***
Log Budget Proxy (Prov Roads, t-1)	-0.034 (0.005)***	-0.093 (0.015)***	-0.064 (0.012)***
Log Budget Proxy (Kabu Roads, t-1)		-0.052 (0.014)***	-0.013 (0.006)**
Log Budget Proxy (Nat'l Roads, t-2)	0.099 (0.014)***		
Log Budget Proxy (Prov Roads, t-2)	-0.034 (0.006)***	-0.078 (0.018)***	-0.065 (0.020)***
Log Budget Proxy (Kabu Roads, t-2)		-0.045 (0.013)***	-0.036 (0.013)***
Log B-Proxy (Nat'l Roads)^2	-0.014 (0.001)***	-0.012 (0.003)***	-0.018 (0.002)***
Log B-Proxy (Prov Roads)^2	-0.005 (0.000)***	-0.008 (0.003)**	-0.002 (0.001)*
Log B-Proxy (Kabu Roads)^2			-0.004 (0.002)**
Log B-Proxy (Nat'l Roads)^2 (t-1)	-0.010 (0.001)***	0.005 (0.002)**	0.007 (0.002)***
Log B-Proxy (Prov Roads)^2 (t-1)		0.013 (0.003)***	0.009 (0.002)***
Log B-Proxy (Kabu Roads)^2 (t-1)	-0.002 (0.001)**	0.011 (0.004)**	
Log B-Proxy (Nat'l Roads)^2 (t-2)	-0.011 (0.001)***	-0.004 (0.001)***	-0.003 (0.001)**
Log B-Proxy (Prov Roads)^2 (t-2)		0.012 (0.003)***	0.009 (0.003)***
Log B-Proxy (Kabu Roads)^2 (t-2)	-0.004 (0.001)***	0.015 (0.004)***	0.011 (0.004)***
N	3121	22617	66254
Adj. R^2	0.413	0.672	0.644
Regression F-Stat	77.708	149.049	97.222
DV Mean	1.000	1.831	1.828
Kabupaten-Years	Yes	.	.
Household-Years	.	Yes	.
Individual-Years	.	.	Yes
Year FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

All regressions include a constant. Kabupaten-year controls include logs of population and non-oil GDRP. Household controls include: district GDP and month of survey indicators. Individual controls include: district GDP, age, education, household size, month of survey indicators. Robust standard errors in parentheses, clustered at the kabupaten level in column 1, and the village level in columns 2 and 3. */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_firstStage.do`.

Table A.2: Variable Names for Robustness Tables

Variable	**Description**	**Original Table**
Y_1	Working? (0 1)	Table 8
Y_2	Log Hours Worked (0 1)	Table 8
Y_3	Log Total Earnings	Table 11
Y_4	Log Manufacturing Earnings	Table 11
Y_5	Log Agricultural Earnings	Table 11
Y_6	Log Sales and Services Earnings	Table 11
Y_7	Log Other Sector Earnings	Table 11
Y_8	Working in Agriculture (0 1)	Table 9
Y_9	Working in Agriculture, Primary (0 1)	Table 9
Y_{10}	Log Hours Worked, Agriculture	Table 9
Y_{11}	Working in Manufacturing (0 1)	Table 9
Y_{12}	Working in Manufacturing, Primary (0 1)	Table 9
Y_{13}	Log Hours Worked, Manufacturing	Table 9
Y_{14}	Working in Sales and Services (0 1)	Table 9
Y_{15}	Working in Sales and Services, Primary (0 1)	Table 9
Y_{16}	Log Hours Worked, Sales and Services	Table 9
Y_{17}	Working in Other (0 1)	Table 9
Y_{18}	Working in Other, Primary (0 1)	Table 9
Y_{19}	Log Hours Worked, Other	Table 9

This table contains the names and descriptions of variables, for reference to individual and household-level robustness tables, Appendix Tables A.3-XXX.

Table A.3: Robustness: Individual-Level Results (Part 1)

Panel A: Preferred Spec. (GMM)	Y_1 (1)	Y_2 (2)	Y_3 (3)	Y_4 (4)	Y_5 (5)	Y_6 (6)	Y_7 (7)
Log Avg IRI (Budget IV)	-0.037	-0.011	-0.894	0.058	-1.303	-3.474	0.080
	(0.034)	(0.060)	(0.328)***	(0.105)	(0.664)**	(1.224)***	(0.215)
Panel B: OLS Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Log Avg IRI	0.016	0.010	-0.465	0.030	-1.149	-1.284	0.072
	(0.020)	(0.035)	(0.190)**	(0.059)	(0.389)***	(0.770)*	(0.132)
Log Avg IRI, Non-Movers	0.034	-0.011	-0.382	0.012	-1.128	-1.401	0.141
	(0.022)	(0.038)	(0.215)*	(0.064)	(0.431)***	(0.839)*	(0.161)
Log Avg IRI, Rural Areas	0.081	-0.018	-0.563	0.032	-0.709	-0.638	0.351
	(0.027)***	(0.052)	(0.354)	(0.113)	(0.601)	(0.851)	(0.268)
Log Travel Time to Prov Capital	0.016	0.027	-0.147	-0.021	-0.580	-1.674	-0.025
	(0.009)*	(0.031)	(0.114)	(0.032)	(0.268)**	(0.533)***	(0.064)
Log Local Market Access (GDRP weights)	0.017	0.028	-0.181	-0.015	-0.665	-1.699	-0.030
	(0.010)	(0.032)	(0.127)	(0.035)	(0.294)**	(0.570)***	(0.070)
Log Local Market Access (Pop weights)	0.018	0.027	-0.179	-0.013	-0.650	-1.642	-0.035
	(0.010)*	(0.032)	(0.127)	(0.034)	(0.290)**	(0.567)***	(0.069)
Panel C: GMM Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Log Avg IRI, Non-Movers	-0.014	-0.013	-0.805	-0.002	-1.224	-3.782	0.051
	(0.036)	(0.069)	(0.371)**	(0.124)	(0.748)	(1.367)***	(0.238)
Log Avg IRI, Rural Areas	0.027	-0.015	-0.819	-0.100	-0.607	-1.867	-0.135
	(0.037)	(0.083)	(0.482)*	(0.152)	(0.880)	(1.283)	(0.318)
Log Avg IRI, Hausman IV	-0.022	-0.018	-0.842	-0.101	-1.807	-2.599	-0.276
	(0.031)	(0.058)	(0.306)***	(0.102)	(0.644)***	(1.237)**	(0.220)
Log Travel Time to Prov Capital, Budget IV	0.014	-0.052	-0.267	-0.080	-0.571	-2.184	-0.052
	(0.019)	(0.039)	(0.268)	(0.077)	(0.566)	(1.243)*	(0.141)
Log Local Market Access (GDRP weights), Budget IV	-0.004	-0.053	-0.312	-0.087	-0.595	-1.488	-0.075
	(0.025)	(0.047)	(0.304)	(0.083)	(0.624)	(1.293)	(0.163)
Log Local Market Access (Pop weights), Budget IV	-0.003	-0.052	-0.291	-0.084	-0.567	-1.226	-0.082
	(0.024)	(0.046)	(0.297)	(0.081)	(0.612)	(1.266)	(0.163)

Refer to Appendix Table A.2 for variable names and definitions. All regressions include a constant, individual and year fixed effects. Individual controls include: district GDP, age, education, household size, month of survey indicators. Robust standard errors in parentheses, clustered at the village level. */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_robustness_indiv.do`.

Table A.4: Robustness: Individual-Level Results (Part 2)

Panel A: Preferred Spec. (GMM)	Y_8 (1)	Y_9 (2)	Y_{10} (3)	Y_{11} (4)	Y_{12} (5)	Y_{13} (6)
Log Avg IRI (Budget IV)	0.055	0.055	2.889	-0.114	-0.129	-6.240
	(0.033)*	(0.034)	(2.494)	(0.036)***	(0.033)***	(1.802)***
Panel B: OLS Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log Avg IRI	0.062	0.055	1.458	-0.085	-0.089	-4.559
	(0.018)***	(0.018)***	(1.349)	(0.022)***	(0.020)***	(1.130)***
Log Avg IRI, Non-Movers	0.066	0.055	1.230	-0.087	-0.100	-5.001
	(0.019)***	(0.020)***	(1.450)	(0.025)***	(0.023)***	(1.231)***
Log Avg IRI, Rural Areas	0.069	0.071	1.020	-0.089	-0.106	-4.781
	(0.029)**	(0.030)**	(2.273)	(0.032)***	(0.028)***	(1.469)***
Log Travel Time to Prov Capital	0.026	0.021	0.840	-0.050	-0.055	-1.477
	(0.010)**	(0.011)*	(0.603)	(0.012)***	(0.012)***	(0.569)***
Log Local Market Access (GDRP weights)	0.029	0.024	1.161	-0.048	-0.055	-1.699
	(0.011)**	(0.012)**	(0.719)	(0.013)***	(0.012)***	(0.613)***
Log Local Market Access (Pop weights)	0.028	0.022	1.123	-0.046	-0.053	-1.610
	(0.011)**	(0.011)*	(0.713)	(0.013)***	(0.012)***	(0.602)***
Panel C: GMM Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log Avg IRI, Non-Movers	0.056	0.046	1.498	-0.121	-0.147	-6.515
	(0.036)	(0.037)	(2.611)	(0.038)***	(0.036)***	(2.010)***
Log Avg IRI, Rural Areas	0.046	0.040	1.926	-0.120	-0.128	-7.331
	(0.045)	(0.045)	(3.601)	(0.042)***	(0.035)***	(2.090)***
Log Avg IRI, Hausman IV	0.086	0.090	3.639	-0.166	-0.162	-9.519
	(0.032)***	(0.033)***	(2.066)*	(0.037)***	(0.033)***	(1.840)***
Log Travel Time to Prov Capital, Budget IV	0.017	0.010	-1.502	-0.099	-0.125	-5.232
	(0.020)	(0.018)	(1.198)	(0.029)***	(0.031)***	(1.632)***
Log Local Market Access (GDRP weights), Budget IV	0.028	0.006	-1.863	-0.119	-0.144	-5.891
	(0.027)	(0.021)	(1.500)	(0.035)***	(0.034)***	(1.859)***
Log Local Market Access (Pop weights), Budget IV	0.028	0.005	-1.831	-0.117	-0.141	-5.755
	(0.027)	(0.021)	(1.468)	(0.034)***	(0.034)***	(1.826)***

Refer to Appendix Table A.2 for variable names and definitions. All regressions include a constant, individual and year fixed effects. Individual controls include: district GDP, age, education, household size, month of survey indicators. Robust standard errors in parentheses, clustered at the village level. */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_robustness_indiv2.do`.

Table A.5: ROBUSTNESS: INDIVIDUAL-LEVEL RESULTS (PART 3)

Panel A: Preferred Spec. (GMM)	Y_{14} **(1)**	Y_{15} **(2)**	Y_{16} **(3)**	Y_{17} **(4)**	Y_{18} **(5)**	Y_{19} **(6)**
Log Avg IRI (Budget IV)	0.054 (0.035)	0.082 (0.035)**	4.523 (1.963)**	-0.026 (0.018)	-0.015 (0.011)	0.026 (0.802)
Panel B: OLS Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log Avg IRI	0.046 (0.020)**	0.053 (0.020)***	3.064 (1.319)**	-0.020 (0.012)*	-0.014 (0.009)	-0.304 (0.628)
Log Avg IRI, Non-Movers	0.054 (0.021)**	0.059 (0.020)***	2.928 (1.334)**	-0.018 (0.013)	-0.007 (0.010)	-0.320 (0.651)
Log Avg IRI, Rural Areas	0.033 (0.025)	0.048 (0.027)*	0.738 (1.381)	-0.023 (0.013)*	-0.005 (0.008)	0.381 (0.556)
Log Travel Time to Prov Capital	0.029 (0.020)	0.029 (0.015)*	1.492 (1.351)	0.005 (0.014)	0.003 (0.009)	0.024 (0.469)
Log Local Market Access (GDRP weights)	0.030 (0.020)	0.030 (0.015)**	1.702 (1.339)	-0.000 (0.015)	0.001 (0.010)	-0.170 (0.530)
Log Local Market Access (Pop weights)	0.030 (0.020)	0.030 (0.015)**	1.709 (1.347)	-0.001 (0.015)	0.001 (0.010)	-0.216 (0.538)
Panel C: GMM Results	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Log Avg IRI, Non-Movers	0.062 (0.037)*	0.104 (0.036)***	5.937 (1.919)***	-0.032 (0.018)*	-0.011 (0.013)	-0.038 (0.872)
Log Avg IRI, Rural Areas	0.079 (0.041)*	0.093 (0.038)**	3.143 (2.061)	-0.025 (0.016)	-0.006 (0.011)	-0.779 (0.550)
Log Avg IRI, Hausman IV	0.091 (0.034)***	0.096 (0.032)***	5.513 (2.037)***	-0.021 (0.020)	-0.011 (0.014)	-0.325 (0.948)
Log Travel Time to Prov Capital, Budget IV	0.100 (0.033)***	0.111 (0.032)***	7.353 (1.839)***	-0.014 (0.014)	-0.005 (0.009)	-0.465 (0.513)
Log Local Market Access (GDRP weights), Budget IV	0.099 (0.037)***	0.127 (0.036)***	7.742 (2.081)***	-0.016 (0.015)	-0.004 (0.010)	-0.533 (0.580)
Log Local Market Access (Pop weights), Budget IV	0.098 (0.037)***	0.125 (0.036)***	7.515 (2.062)***	-0.015 (0.015)	-0.004 (0.010)	-0.547 (0.573)

Refer to Appendix Table A.2 for variable names and definitions. All regressions include a constant, individual and year fixed effects. Individual controls include: district GDP, age, education, household size, month of survey indicators. Robust standard errors in parentheses, clustered at the village level. */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_robustness_indiv3.do`.

Table A.6: ROAD ROUGHNESS AND PROBABILITY OF WORKING: HETEROGENEITY

		Age		**Gender**		**Education**		**Initial PCE**		**Own Farmland?**	
Dep Var: Working? (0 1)	**All Obs (1)**	**$<$ Med (2)**	**$\geq$ Med (3)**	**F (4)**	**M (5)**	**$<$ Med (6)**	**$\geq$ Med (7)**	**$<$ Med (8)**	**$\geq$ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	-0.037	-0.123	0.002	-0.020	-0.066	-0.011	-0.095	-0.013	-0.071	-0.062	-0.023
	(0.034)	(0.070)*	(0.033)	(0.048)	(0.033)**	(0.044)	(0.040)**	(0.042)	(0.044)	(0.041)	(0.045)
Adj. R^2	0.454	0.350	0.459	0.374	0.448	0.443	0.462	0.445	0.474	0.471	0.436
Regression F-Stat	34.111	59.174	9.186	23.800	24.119	8.171	42.791	18.410	16.287	21.484	17.711
Kleibergen-Paap F-Stat	33.594	26.782	34.343	33.836	32.212	26.660	34.965	29.379	29.570	27.925	27.021
DV Mean	0.679	0.514	0.748	0.528	0.837	0.714	0.641	0.696	0.662	0.644	0.717
N	35193	7920	27273	18249	16932	19483	15710	17521	16106	17080	16910
N Individuals	12459	3383	9076	6354	6099	6547	5912	6129	5840	6208	5876
N Desa / Kelurahan	377	301	320	328	297	305	311	276	324	316	278
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_heterogeneity.do`.

Table A.7: Road Roughness and Hours Worked: Heterogeneity

		Age		Gender		Education		Initial PCE		Own Farmland?	
Dep Var: Log Hours Worked	**All Obs (1)**	**< Med (2)**	**≥ Med (3)**	**F (4)**	**M (5)**	**< Med (6)**	**≥ Med (7)**	**< Med (8)**	**≥ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	-0.011	-0.219	0.043	0.089	-0.086	0.018	-0.034	-0.036	0.042	0.012	0.048
	(0.060)	(0.147)	(0.062)	(0.100)	(0.066)	(0.087)	(0.070)	(0.083)	(0.073)	(0.065)	(0.083)
Adj. R^2	0.329	0.282	0.312	0.318	0.276	0.316	0.352	0.319	0.345	0.323	0.342
Regression F-Stat	3.461	5.349	2.778	1.579	4.117	1.314	3.512	1.737	2.113	1.843	3.102
Kleibergen-Paap F-Stat	32.698	19.528	33.307	31.877	31.713	26.456	30.790	27.691	29.295	26.550	26.547
DV Mean	5.119	5.046	5.140	4.965	5.220	5.117	5.120	5.104	5.135	5.170	5.070
N	22255	3050	19205	8321	13930	13114	9141	11547	9765	9987	11573
N Individuals	8168	1374	6794	3122	5044	4662	3506	4181	3654	3770	4152
N Desa / Kelurahan	321	250	292	262	291	271	275	245	285	272	259
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_heterogeneity.do`.

Table A.8: Road Roughness and Working in Sales and Services (Primary): Heterogeneity

DV: Working in Sales and Services (Primary) (0 1)	All Obs (1)	Age: < Med (2)	Age: ≥ Med (3)	Gender: F (4)	Gender: M (5)	Education: < Med (6)	Education: ≥ Med (7)	Initial PCE: < Med (8)	Initial PCE: ≥ Med (9)	Own Farmland?: No (10)	Own Farmland?: Yes (11)
Log Avg IRI	0.082	0.193	0.084	0.148	0.061	0.095	0.144	0.077	0.114	0.027	0.102
	(0.035)**	(0.087)**	(0.036)**	(0.053)***	(0.041)	(0.044)**	(0.051)***	(0.044)*	(0.054)**	(0.056)	(0.044)**
Adj. R^2	0.534	0.418	0.547	0.618	0.432	0.574	0.479	0.541	0.518	0.519	0.522
Regression F-Stat	6.062	1.353	5.356	4.751	3.993	2.495	8.776	3.612	4.178	7.006	1.773
Kleibergen-Paap Wald rk F-Stat	32.711	19.528	33.321	31.895	31.717	26.469	30.785	27.690	29.354	26.561	26.547
DV Mean	0.279	0.296	0.274	0.377	0.213	0.248	0.316	0.213	0.352	0.360	0.201
N	22259	3050	19209	8323	13932	13117	9142	11548	9768	9991	11573
N Individuals	8169	1374	6795	3123	5044	4663	3506	4181	3655	3771	4152
N Desa / Kelurahan	321	250	292	262	291	271	275	245	285	272	259
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_heterogeneity.do`.

Table A.9: Road Roughness and Working in Other Sectors (Primary): Heterogeneity

		Age		Gender		Education		Initial PCE		Own Farmland?	
Dep Var: Working in Other Sectors	**All Obs (1)**	**< Med (2)**	**≥ Med (3)**	**F (4)**	**M (5)**	**< Med (6)**	**≥ Med (7)**	**< Med (8)**	**≥ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	-0.015	-0.001	-0.015	-0.015	-0.024	-0.002	-0.036	-0.009	-0.022	-0.009	-0.020
	(0.011)	(0.032)	(0.015)	(0.013)	(0.020)	(0.008)	(0.023)	(0.014)	(0.021)	(0.017)	(0.015)
Adj. R^2	0.515	0.175	0.590	0.616	0.428	0.169	0.491	0.433	0.515	0.498	0.545
Regression F-Stat	3.954	6.657	3.132	5.038	3.301	1.370	5.334	1.910	4.190	2.748	3.222
Kleibergen-Paap Wald rk F-Stat	33.594	26.782	34.343	33.836	32.212	26.660	34.965	29.379	29.570	27.925	27.021
DV Mean	0.067	0.041	0.078	0.057	0.078	0.011	0.129	0.029	0.107	0.075	0.059
N	35193	7920	27273	18249	16932	19483	15710	17521	16106	17080	16910
N Individuals	12459	3383	9076	6354	6099	6547	5912	6129	5840	6208	5876
N Desa / Kelurahan	377	301	320	328	297	305	311	276	324	316	278
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_heterogeneity.do`.

Table A.10: Road Roughness and Working in Informal Sector (Primary): Heterogeneity

		Age		Gender		Education		Initial PCE		Own Farmland?	
DV: Working in Informal Sector	**All Obs (1)**	**$<$ Med (2)**	**$\geq$ Med (3)**	**F (4)**	**M (5)**	**$<$ Med (6)**	**$\geq$ Med (7)**	**$<$ Med (8)**	**$\geq$ Med (9)**	**No (10)**	**Yes (11)**
Log Avg IRI	0.091	-0.114	0.109	0.078	0.097	0.139	0.040	0.119	0.056	0.043	0.101
	(0.032)***	(0.107)	(0.033)***	(0.041)*	(0.046)**	(0.043)***	(0.043)	(0.044)***	(0.042)	(0.045)	(0.040)**
Adj. R^2	0.601	0.439	0.615	0.689	0.548	0.542	0.623	0.571	0.641	0.570	0.588
Regression F-Stat	6.857	2.051	8.569	2.929	6.864	4.411	5.546	4.140	4.109	5.310	5.360
Kleibergen-Paap Wald rk F-Stat	32.707	19.474	33.295	31.991	31.671	26.425	30.925	27.860	29.087	26.569	26.578
DV Mean	0.576	0.455	0.611	0.643	0.531	0.669	0.460	0.621	0.525	0.442	0.703
N	22147	3033	19114	8300	13843	13042	9105	11487	9724	9943	11523
N Individuals	8133	1366	6767	3114	5017	4640	3493	4160	3643	3755	4138
N Desa / Kelurahan	320	249	292	262	290	270	275	244	285	272	258
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

All regressions include a constant. Controls include: district GDP, age, education, household size, month of survey indicators. Dependent variables are only defined if the individual reported working. Robust standard errors in parentheses, clustered at the (initial) village level. For the GMM specifications, we report the adjusted R^2 from the analagous reduced-form regression (regressing the dependent variable on the instruments). */**/*** denotes significant at the 10% / 5% / 1% levels. This table was constructed using the following .do file: `$do_files/analysis/analysis_IFLS_heterogeneity.do`.

B Data Appendix

B.1 Road Quality Data

Data on the quality of Indonesia's highway networks were produced by DPU as part of Indonesia's Integrated Road Management System (IRMS). This appendix section begins by providing some background on road management in Indonesia, describing the road classification system and discussing IRMS coverage. It then discusses the measures of road quality that are collected in IRMS and how they are measured. I then discuss how the road network data were created.

B.1.1 Background on Road Management

Indonesia's national road network is currently managed and maintained by the Department of Public Works (*Departemen Pekerjaan Umum*, DPU), specifically by the Directorate General of Highways (*Direktorat Jenderal Bina Marga*). According to Law No. 38, 2004, roads are classified into four different types of roads, primarily based on their function for users. Arterial roads (*jalan arteri*) serve as the major transportation linkages between urban areas, and are characterized by longer distances, higher speeds, and limited access. Speeds are meant to be a minimum of 60 km/h, and width should be at least 11 meters to accommodate larger traffic volumes. Collector roads (*jalan kolektor*) serve "collector or distributor transportation" and are characterized by medium distance travel with medium speeds. Collector roads are subdivided into primary collector roads (*jalan kolektor primer*), which should have a minimum speed of 20 km/h and width of 9 meters, and secondary collector roads, which should have a minimum speed of 20 km/h and width of 9 meters. Local roads (*jalan lokal*) and Neighborhood Roads (*jalan lingkungan*) serve local areas at lower speeds, and are characterized by unlimited access.

Roads can also be classified by their management authority, or "status" (*wewenang penyelenggaraan*). Generally, arterial and primary collector roads are managed by the national government (specifically by DPU). Secondary and tertiary collector roads are managed by provincial governments, while local and neighborhood roads are managed by the kabupaten, kecamatan, and desa governments. Table B.1 describes the road classification system, minimum speed and width guidelines, and management authorities.

Table B.2 depicts the coverage of the IRMS dataset by road function and managing authority, as measured by counts of the number of kilometer-post observations that appear in the entire dataset. Most of the observations, and indeed most of the road network, is made up by collector roads (K1-K3), though the category with the next largest coverage is the arterial roads. Local and neighborhood roads are not very well surveyed in this dataset. Although the network of village and kabupaten roads is doubtless extremely dense, I cannot use this dataset to say very much about it. But since the data do cover arterial and collector roads, the major roads connecting regions and cities in Indonesia, this dataset seems particularly well suited for evaluating models of economic geography and regional trade.

B.1.2 Measures of Road Quality

There are a number of different devices that transport engineers have developed to collect measurements of road quality, and there are several different measures of road quality. The most widely used measure of road roughness, and the measure used in this study, is the international roughness index (IRI), developed by the World Bank in the 1980s. IRI is constructed as a filtered ratio of a standard vehicle's accumulated suspension motion (in meters), divided by the distance travelled by the vehicle during measurement (in kilometers). Expressed in units of slope (m/km), IRI is a characteristic of a vehicle's longitudinal profile. Importantly, since it is a measure of a physical quantity, IRI is standardized, as opposed to other subjective measures of ride quality. Figure **??** shows the relationship between different ranges of IRI and surface type; generally, larger roughness levels correspond to worse surfaces, but the mapping is not one-to-one.

Bennett et al. (2007) distinguish between several different types of devices for measuring road roughness and provide a good overview of their relative strengths and weaknesses. Over the course of its existence, Indonesia's IRMS has largely made use of two different types of measuring devices.[20] Before 1999, roads were surveyed using devices like the ROMDAS, which estimate IRI indirectly. The ROMDAS machine is a calibrated bump integrator, which must first be calibrated and estimates IRI from correlation equations. It is very useful for measuring roughness on bumpy roads and can record high levels of IRI, but the device must be calibrated manually, and measurement error can occur if the device is miscalibrated.

The ROMDAS device is also portable, meaning that it can be used inside different vehicles (each of which would require unique calibrations). The portability contrasts with devices like the high-speed laser profilometer, which is essentially a separate vehicle reserved entirely for the purposes of collecting road quality data. The device uses lasers and optical techniques to scan the road as it is traversed and create measures of surface profiles. These instruments are very accurate, but are much more expensive. Moreover, they might become mis-calibrated on extremely rough roads. Indonesia started using the high speed laser profilometer for collecting its road quality data in 1999, licensing vehicles from the Australian government.

Road width and surface type are more straightforward variables to measure, involving visual inspection and simple measurement. I categorize a kilometer-post interval as being unpaved if it is either an earth, gravel, or sand road, or if it was given a granular base (crushed stone) treatment, a first step in the process of paving.

B.1.3 Creation of Road Network Data

Using GIS shapefiles of the road network provided to me by DPU, I have georeferenced the kilometer post observations of road quality, in order to capture the evolution of Indonesia's transportation network over space and time. This proved to be a challenging exercise, because the identifiers for each road-link-interval observation were not consistent over time, and because the identifiers in the shapefile and in the linearly referenced dataset were often different, even though both did refer to exactly the same link.

Once the IRMS interval data was successfully merged to the regional network shapefiles, I converted the GIS database of road links into a weighted graph of arcs and nodes, as commonly used in the transportation literature. Nodes represent locations (such as ports, cities, or the centroids of kabupatens, my unit of analysis), arcs represent the possibility of traveling between two nodes, and weights represent the cost of moving goods along a given arc. Weights were constructed according to

[20] I am very grateful for the extensive discussions I've had with Glen Stringer about IRMS; this section of the appendix benefits highly from our conversations.

the IRMS data on road quality, and for simplicity, the cost of moving along each road was assumed to be the same, no matter which way you were traveling.[21]

For computational reasons, I have used a simplified representation of Indonesia's road network, where the number of nodes and links was small enough for network algorithms to operate on it using a desktop computer.[22] Table B.3 depicts the number of network arcs, the total distance of the network, and merge statistics for the kilometer-post observations. Merge statistics are pretty good for arterial and collector roads, but the quality of merges falls substantially for local and neighborhood roads, due most likely to poor shapefile coverage for that type of road network.

The interval observations were not matched directly to their exact locations in the network, because I had no knowledge of the exact location of the kilometer posts. To deal with this, I first aggregated the kilometer-post interval observations to the road-link level by constructing distance-weighted averages of the road quality variables. Each network arc-year observation was then assigned the value of this average road quality variable that corresponds to its road link.[23]

B.1.4 Roughness, Speed, and Ride Quality

One effect that rough roads have on vehicles is that they require the driver to travel at lower speeds. When faced with potholes, ragged pavement, or poor surfaces, drivers slow down, and this reduction in speed increases travel time and hence the cost of travel. Of course, there is not a one-to-one relationship between road roughness and speed, because drivers choose the speed at which they travel, and different preferences for smoothness of the ride or the desired arrival time might induce different choices of speed.

Yu et al. (2006) explore the relationship between *jolt*, or the "jerk" experienced by road users, and subjective measures of ride quality and road roughness at different speeds.[24] Using survey data in which users were asked to rate the quality of particular rides, the authors find that people experience greater discomfort while traveling at higher speeds on rough roads, but lowering speed on rough roads can reduce discomfort. The authors provide a mapping between subjective measures of ride quality and roughness at different speeds, and this mapping can be used to infer the maximum speed that one can travel in order to achieve a ride of a certain quality, given pavement roughness. Table B.4 reproduces this mapping. Because travel times were unreasonably long for high quality rides given Indonesia's rough roads, and because the subjective quality measures were chosen by Western drivers, I have focused on the poor ride quality speed thresholds in my empirical work.

[21] Another tedious issue involved the construction of junction points where the road links intersected. The shapefiles were originally stored as MapInfo files, an older shapefile format that required conversion for use with Arcview, and in this conversion, information on where the roads crossed was lost, requiring painstaking editing. The shapefiles were also not designed to be used in any network analysis, so much care had to be taken to make them usable.

[22] The road lines were straightened using the "Generalize" command from ET Geotools, which employs the RamerDouglasPeucker algorithm for reducing the number of points that represents a line.

[23] In some cases, when a network arc had no data for a particular year, I assigned the network arc the average value of road quality for arcs with the same function. This was done because constructing the transport cost variables involved a search over the entire network, and if certain network arcs were coded as missing, this could distort the search substantially. Overall, imputation amounted to no more than 5 percent of network arc observations in any given year.

[24] *Jolt* is officially defined as the vector that specifies the time-derivative of acceleration; in other words, the third derivative of the vertical displacement of vehicle to time t.

Given the maximum speed that one can travel on roads of different roughness levels, it is straightforward to calculate travel times for each network arc, the primary measure of transport costs used in this study. Note that the travel times on road sections were computed using the detailed kilometer-post interval roughness data. These were then aggregated to the network arcs using distance-weighted averages.

B.2 Administrative Boundaries

Administrative boundary shapefiles were constructed by BPS for use during the 2000 Household Census. These shapefiles contain the polygon boundaries of all provinces, kabupatens, kecamatans, and desas for the entire extent of the Indonesian archipelago. However, after the fall of Suharto and a massive decentralization program, many new kabupatens were created, splitting existing kabupatens into new ones. For instance, in 1990 there were 290 kabupatens and kotas, but by 2003, there were 416 kabupatens and kotas. The fact that administrative boundaries are not fixed over time create difficulties for the analysis.

Because of the need for a geographic unit of analysis that was consistently defined over time, I used kabupaten borders as they were defined in 1990. BPS provided the administrative boundary shapefile for 2000, as well as a correspondence table between kabupaten codes in 2000 and kabupaten codes from 1990 to the present. This information was processed using ArcView to create the 1990 shapefiles that form the basis of the analysis. Throughout the paper, all survey data were appropriately merged back to the 1990 kabupaten definitions.

Table B.1: Indonesia's Road Classification System

Function	Code	Minimum Speed	Minimum Width	Management Authority
Arterial	A	60 KM/H	11 M	NATIONAL
Collector-1	K1	40 KM/H	9 M	NATIONAL
Collector-2	K2	20 KM/H	9 M	PROVINCIAL
Collector-3	K3	20 KM/H	9 M	PROVINCIAL
Local	L	20 KM/H	7.5 M	KABUPATEN & DESA
Neighborhood	Z	15 KM/H	6.5 M	KABUPATEN & DESA

Source: Departemen Pekerjaan Umum, 2008

Table B.2: Road Function and Managing Authority, Kilometer-Post Observations, 1990-2007

	Road Function			**Managing Authority**		
	Code	**Number of Obs.**	**Share of Total**	**Code**	**Number of Obs.**	**Share of Total**
Java	**A**	52,917	0.17	**N**	93,808	0.30
	K1	40,889	0.13	**P**	132,649	0.42
	K2	121,386	0.39	**K**	15,862	0.05
	K3	10,714	0.03	**S**	72,068	0.23
	L	15,862	0.05			
	Z	72,619	0.23			
	Total	314,387	1.00	**Total**	314,387	1.00
Sumatra	**A**	103,160	0.20	**N**	202,915	0.39
	K1	99,782	0.19	**P**	263,409	0.50
	K2	235,750	0.45	**K**	11,391	0.02
	K3	27,632	0.05	**S**	45,680	0.09
	L	11,391	0.02			
	Z	45,680	0.09			
	Total	523,395	1.00	**Total**	523,395	1.00
Sulawesi	**A**	54,496	0.21	**N**	143,147	0.54
	K1	87,728	0.33	**P**	72,198	0.27
	K2	71,234	0.27	**K**	18,232	0.07
	K3	1,887	0.01	**S**	29,371	0.11
	L	18,232	0.07			
	Z	29,371	0.11			
	Total	262,948	1.00	**Total**	262,948	1.00

Source: IRMS and author's calculations. Data come from kilometer-post observations. Standard deviations in parentheses.

Table B.3: Number of Network Arcs, Distances, and Merge Statistics (by road function)

		Road Function						
		A	**K1**	**K2**	**K3**	**L**	**Z**	**Miss**
	# of Arcs	1168	889	2618	309	315	37	.
	# of Road IDs	220	129	354	43	72	6	.
	Total Distance	2944.91	1970.65	5832.59	750.39	663.44	92.16	.
	Link-Years Merged	16538	13685	38719	3876	4689	14572	3015
Java	**Link-Years Unmerged**	1838	735	1842	45	971	21772	157
	% Merged	0.90	0.95	0.95	0.99	0.83	0.40	0.95
	Arc-Years Merged	20,844	16002	46350	5562	5670	666	.
	Arc-Years Unmerged	180	0	774	0	0	0	.
	% Merged	0.99	1.00	0.98	1.00	1.00	1.00	.
	# of Arcs	1485	1205	2975	453	277	22	41
	# of Road IDs	207	165	412	87	66	6	13
	Total Distance	4964.69	4469.43	11551.28	1492.97	571.67	56.44	147.56
	Link-Years Merged	24755	20035	49171	6808	2603	8730	1406
Sumatra	**Link-Years Unmerged**	718	373	537	52	394	9722	12
	% Merged	0.97	0.98	0.99	0.99	0.87	0.47	0.99
	Arc-Years Merged	26730	21690	51876	7830	4986	396	0
	Arc-Years Unmerged	0	0	1674	324	0	0	738
	% Merged	1.00	1.00	0.97	0.96	1.00	1.00	0.00
	# of Arcs	1624	2319	2051	15	391	.	45
	# of Road IDs	113	116	150	4	44	.	1
	Total Distance	2836.96	3805.92	4369.33	28.35	732.96	.	70.34
	Link-Years Merged	24006	24006	34711	30911	551	5670	5674
Sulawesi	**Link-Years Unmerged**	25	356	410	339	9	118	4755
	% Merged	1.00	0.99	0.99	0.99	0.98	0.98	0.54
	Arc-Years Merged	25794	35694	33660	270	7038	.	0
	Arc-Years Unmerged	3438	6048	3258	0	0	.	810
	% Merged	0.88	0.86	0.91	1.00	1.00		0.00

Source: IRMS and author's calculations. Missing function information is attributable to poorly coded shapefiles. Arc-Years could be unmerged potentially because there were no surveys done on that particular link; statistics are computed assuming a balanced panel. Road IDs are defined in the shapefile, while Link IDs are defined from the IRMS data.

Table B.4: Roughness and Ride-Quality Speed Limits

Max Speed	**Good**	**Fair**	**Mediocre**	**Poor**
120 km/h	$IRI \in [0.00, 1.49]$	$IRI \in [0.00, 1.89]$	$IRI \in [0.00, 2.70]$	$IRI \in [0.00, 3.24]$
100 km/h	$IRI \in [1.49, 1.79]$	$IRI \in [1.89, 2.27]$	$IRI \in [2.70, 3.24]$	$IRI \in [3.24, 4.05]$
80 km/h	$IRI \in [1.79, 2.24]$	$IRI \in [2.27, 2.84]$	$IRI \in [3.24, 4.05]$	$IRI \in [4.05, 4.63]$
70 km/h	$IRI \in [2.24, 2.57]$	$IRI \in [2.84, 3.25]$	$IRI \in [4.05, 4.63]$	$IRI \in [4.63, 5.40]$
60 km/h	$IRI \in [2.57, 2.99]$	$IRI \in [3.25, 3.79]$	$IRI \in [4.63, 5.40]$	$IRI \in [5.40, 6.25]$
50 km/h	$IRI \in [2.99, 3.59]$	$IRI \in [3.79, 4.54]$	$IRI \in [5.40, 6.25]$	$IRI \in [6.25, 8.08]$
40 km/h	$IRI \in [3.59, 4.49]$	$IRI \in [4.54, 5.69]$	$IRI \in [6.25, 8.08]$	$IRI \in [8.08, 10.80]$
30 km/h	$IRI \in [4.49, 5.99]$	$IRI \in [5.69, 7.59]$	$IRI \in [8.08, 10.80]$	$IRI \in [10.80, 16.16]$
20 km/h	$IRI \in [5.99, 8.99]$	$IRI \in [7.59, 11.39]$	$IRI \in [10.80, 16.16]$	$IRI \in [16.16, 32.32]$
10 km/h	$IRI \in [8.99, \infty)$	$IRI \in [11.39, \infty)$	$IRI \in [16.16, \infty)$	$IRI \in [32.32, \infty)$

Source: Author's calculations and Yu et al. (2006), Table 2. IRI denotes the international roughness index, measured in m/km. Ride quality levels are subjective and measured on a 5-point scale ("Very Good", "Good", "Fair", "Mediocre", and "Poor").