

Access to Markets and the Benefits of Rural Roads: A Nonparametric Approach

Hanan G. Jacoby*

June 1998

Abstract

Transport infrastructure plays a central role in rural development, yet little is known about the size and especially the distribution of benefits from road investments. This paper develops and implements a method for nonparametrically estimating the benefits from road projects at the household level using the relationship between the value of farmland and its distance to agricultural markets. The empirical analysis, using data from Nepal, suggests that providing extensive road access to markets would confer substantial benefits on average, much of these going to poor households. However, the benefits would not be large enough or targeted efficiently enough to appreciably reduce income inequality in the population.

Keywords: Rural Roads, Income Distribution, Nonparametric Regression

JEL Classification: O12, D31, C14

*Development Research Group, The World Bank, 1818 H Street N.W., Washington DC 20433. The views expressed in this paper are those of the author and should not be attributed to the World Bank or affiliated organizations.

I. Introduction

Rural infrastructure is a major development priority (World Bank, 1994), yet little is known about the size and especially the distribution of benefits from such investments in LDCs. The distribution issue is salient, not only in the formulation of policy, but also in understanding the political constraints on the allocation of infrastructure investment.

Rural roads are an important form of public infrastructure, providing cheap access to both markets for agricultural output and for modern inputs. Given limited policy instruments for reaching the remote rural poor, road-building would seem desirable on distributional grounds. On the other hand, the benefits of infrastructure projects accrue mainly to landowners, who are generally not among the very poor. Thus, the extent to which rural road construction ameliorates income inequality is ultimately an empirical question.¹

In this paper, I examine the distributional consequences of rural roads using data from Nepal, a country with a largely agrarian economy, a sparse highway network, and extremely difficult terrain. To this end, I develop an empirical methodology for nonparametrically estimating the household-specific benefits from alternative road projects using information on the value of farmland and distance to agricultural markets. If land behaves like a standard asset, which is a testable assumption, then its value equals the discounted stream of maximal profits from cultivation. Hence, the income gains from lower transport costs should be capitalized in land values. With an estimate of the land value-distance relationship in hand, it is possible to describe the joint distribution of hypothetical road project benefits and household income.

¹ Howe and Richards (1984) discuss some distributional aspects of rural roads and present case studies. Also, van de Walle (1996) uses micro-data and a profit function approach to examine the distribution of benefits from irrigation in Viet Nam.

In principle, road benefits could also be estimated from the relationship between farm profits and distance to markets. However, there are several difficulties with this approach, the most nettlesome of which is that survey data rarely, if ever, provide accurate information on an essential component of profit, the cost of transporting goods and agricultural inputs to and from markets. Another difficulty is that profit (or production) functions assume a fixed technology and thus cannot easily account for potential adaptations of farmers to greater remoteness from markets, such as substitution of traditional for modern inputs or away from transport-intensive crops. The relationship between land value and distance to market is immune from such difficulties.

To be sure, the idea of using land values to estimate the *average* benefits of infrastructure investments in a population is hardly new, though it has not to my knowledge been applied to rural transport. In any case, such estimates do not address the primary question of this paper, which is a distributional one. The innovation here is to link a *household*-level benefit estimate with a measure of household income. In doing so, I take a nonparametric approach. While it is true that economic theory is largely silent on the parametric form of hedonic price functions (see Stock, 1991), in practice, relaxing parametric assumptions in hedonic models is much more likely to matter for distributional questions than for questions about average benefits.

Theoretically, my analysis is based on the Ellet-Walters model of rural transport (see Walters, 1968; Gersovitz, 1989), in which land rents decline with distance to markets through the influence of distance on effective prices. The model, laid out in the next section, provides a simple characterization of the potentially conflicting distributional consequences of road projects.

Section III discusses the nonparametric or, more precisely, semi-nonparametric estimation of the land value equation. Section IV describes the data, analyzes how farmer behavior is influenced by distance to market, and tests the appropriateness of a standard asset-pricing model for land. Section V presents the main empirical results, the analysis of land values and of the distribution of benefits from hypothetical road projects. Section VI concludes the paper.

II . Theoretical Framework

Basic Model

Farmers are assumed to cultivate a single crop using x kg per hectare of a modern input, say chemical fertilizer, and l hours per hectare of labor. Crop yield y (kg per hectare) is produced with a fixed technology represented by the neoclassical production function $y = f(x, l)$. Let w be wage rate and \tilde{v} and \tilde{p} be the “effective” or farm gate prices of output and fertilizer, respectively, discussed below. Per hectare land rent, r , is defined as the maximal profit that can be earned on a hectare of land,

$$r(w, \tilde{v}, \tilde{p}) \equiv \max_{l, x} \{ \tilde{p}y - wl - \tilde{v}x \} \quad (1)$$

and can be thought of as a long-run average.

Effective prices are determined by the economic geography, which is illustrated in Figure 1. A highway of arbitrary length through the countryside runs through a large city where all fertilizer is produced and output is purchased. The highway transects a series of otherwise isolated mountain valleys along which all farms are located. This is a convenient fiction, but not unlike the geography of Nepal. The total cost of transporting

goods between farms and the city has two components: a relatively large cost of headloading goods (i.e., using human porters) between the farm and the road and a relatively small cost of trucking goods along the road. All farmers trade agricultural output and fertilizer in a competitive market center located at the road juncture with their valley (markets at intermediate points up the valley are an inessential complication since goods must still be headloaded from the main highway). From a given farm, it takes h hours to walk to the market center and the portage cost of goods is b Rupees/(kg hours), where b is proportional to the wage. If the money prices of fertilizer and output at a particular market center are v and p , respectively, then the effective purchase price of fertilizer is $\tilde{v} = v + bh$ Rupees/kg and the effective selling price of output is $\tilde{p} = p - bh$ Rupees/kg. All labor can be obtained locally with zero transport costs.

Unprofitable land will not be cultivated, so the limit of cultivation in terms of walking time to the market center, h^* , is implicitly defined by $r(h^*; w, p, v) = 0$. As figure 1 illustrates, h^* declines across valleys as one moves away from the city, because p declines and v increases; ultimately, $h^* = 0$ and all cultivation ends.

As to the relationship between land rent and travel time, by the envelope theorem

$$\frac{\mathcal{J}_h r(w, \tilde{v}, \tilde{p})}{\mathcal{J}_h} = -b(y + x) \quad \text{for } h < h^* \quad (2)$$

So, the negative rent gradient is just equal in magnitude to the total transport costs per hour per hectare. Furthermore, by the convexity of the profit function in prices,

$$\frac{\mathcal{J}_h^2 r(w, \tilde{v}, \tilde{p})}{\mathcal{J}_h^2} > 0 \quad \text{for } h < h^* \quad (3)$$

Thus, along any given valley, the rent function is convex.

Notice that convexity of the rent function does not require that farmers both purchase fertilizer and sell output at the same time (though nonparticipation in the latter market means that rents depend upon the endogenous shadow price of output). However, if one moves far enough away from a market center, farmers may stop selling output and buying fertilizer altogether, and the rent gradient would be zero beyond this point (and thus the rent function not *strictly* convex). Convexity is also robust to the assumption of a single crop or production technology. Figure 2 shows how the rent function along one of the valleys in Figure 1 reflects the profit maximizing choice of available crops or technologies; as travel time rises, farmers may switch away from bulkier crops or from agricultural practices that are intensive in modern inputs.

Roads and Welfare

Using the above framework, consider the welfare implications of building a road of given length off of the main highway into a particular valley. The local scale of the project ensures that it has no general equilibrium effects on wages or prices. To avoid specifying the source of public finance, assume that the project is funded by earmarked foreign aid. Let I denote the length of the road in foot-travel (hours walking time) equivalents. By enabling truck transport, the road effectively cuts portage costs by some fraction m . The new rent function, suppressing its dependence on prices and on m^2 is

² The parameter m reflects road quality. I do not consider the welfare effects of variation in m because, as a practical matter, the cost of upgrading road surface (from earth to gravel or asphalt) far outweighs the small reduction in vehicle operating cost, once truck transport is feasible (Beenhakker and Lago, 1983). Although improvements in road conditions, given surface type, can substantially reduce vehicle operating cost, these cost-savings are likely to be small compared to those of a new road. Of course, in the extreme case where an existing road is impassible to trucks, a road improvement is tantamount to a new road.

$$\begin{aligned} \mathbf{s}(I, h) &= \mathbf{r}(mh) && \text{for } h < I \\ &= \mathbf{r}(h - I(1 - m)) && \text{for } h \geq I \end{aligned} \quad (4)$$

Figure 3 illustrates the rise in land rents as a result of the project, along with the expansion in the limit of cultivation h^* .

An interesting policy question concerns the length of the road, specifically whether it makes sense to build a lot of short roads or fewer long ones, given a fixed construction budget. A key input into this decision is the marginal social benefit of road length, where “society” in this case refers to the farmers in the typical valley. Suppose that household income z is determined by farm profit (land rent) and labor earnings, the latter which is assumed fixed across households for expository purposes; i.e., $z = \mathbf{s}(I, h)A + e$, where A is total landholdings and e is earnings. Let $G(h, A)$ denote the joint cumulative distribution function for distance from market center and landholdings. Finally, let $\mathbf{y}(z)$ be the increasing, strictly concave, indirect utility function. Assuming it is additive, the social welfare function is

$$W(I) = \int_A \int_0^{h^*(I)} \mathbf{y}(\mathbf{s}(I, h)A + e) dG(h, A) \quad (5)$$

Notice that households already on a road do not benefit from its extension further up the valley (i.e., $\mathbf{s}_I = 0$ for $h < I$), so that differentiating (5) with respect to I yields

$$W'(I) = \int_A \int_I^{h^*} \mathbf{y}' \mathbf{s}_I A dG(h, A) \quad (6)$$

To illuminate the distributional issues, it is instructive to fix land area for the moment and to decompose marginal social benefit as follows

$$W'(I, A) = [1 - G(I|A)] \{ E[\mathbf{y}'|A, h > I] E[\mathbf{s}_I|A, h > I] + \text{Cov}[\mathbf{y}', \mathbf{s}_I|A, h > I] \} A \quad (7)$$

The first term in (7) is simply the fraction of households living off the road. The first term in the curly brackets is the average marginal value of the road extension for these beneficiaries, assuming their marginal utilities of income and appreciations in plot value are uncorrelated. The second term in curly brackets accounts for this correlation, which must be negative because $y''s_h A > 0$ and $s_{Ih} < 0$. In other words, farther up the valley, where land rents are low and thus households are poorer, rents rise by less due to the road extension. The size of this “targeting inefficiency” depends crucially on the shape of the rent function; indeed the covariance is zero if the rent function is linear in travel time. In sum, the marginal social benefit of a road extension is higher when: (i) the fraction of the population living off-road is higher; (ii) the average off-road household is poorer (i.e., has a higher y'); (iii) the off-road rent gradient is steeper; and (iv) the targeting inefficiency (the covariance term) is smaller.

With land area variable the distributional issue becomes cloudier. First, for any given h , both the benefit from the road extension and household income are increasing in landholdings.³ While this effect exacerbates inequality, there is also an effect working in the opposite direction. It is often the case in developing countries, and Nepal is no exception, that poorer households are found in more remote areas, perhaps because poor migrants settle on the fringes of cultivation. If A and h are negatively correlated, then the beneficiaries of a road extension (those off the original road) tend to be poorer, resulting in a greater marginal social benefit than if A and h are uncorrelated. Allowing earnings, e , to vary in the population complicates matters further, but the basic point remains, namely

³ Note that in this model renters do not benefit from road construction; their higher rent payments just offset the greater profitability of the land.

that road construction has ambiguous distributional effects. The goal of the empirical work is to resolve this ambiguity in the case of Nepal.

III. Econometric Specification

The theoretical analysis is framed in terms of rents, but my data are on land *values* at the plot level. According to the standard asset-pricing model

$$\log(V) = \log(\mathbf{r}(h; w, p, v)) - \log(r) \quad (8)$$

where V is the present market value of a plot of land and r is the constant discount rate.⁴

Assume that this formula is valid, at least until it is tested in Section IV.

Besides negativity and convexity in h , economic theory imposes no restrictions on the form of \mathbf{r} . It is therefore desirable to estimate the rent function nonparametrically, but this is not feasible when it includes many other variables besides h . For example, any plot characteristic, such as soil quality, that shifts the production function f should also shift the rent function \mathbf{r} . Additionally, in the absence of accurate data on input and output prices, geographic price variation can be swept out of the rent function by including market center (regional) dummy variables.⁵ A within-market analysis also ameliorates the problem of endogenous placement of roads and/or markets; in particular,

⁴ If farmers are risk neutral, then $V = \sum_{t=0}^{\infty} \frac{E_0[\mathbf{r}_{t+1}]}{(1+r)^{t+1}}$, where E_0 is the expectations operator conditional on today's information set. If it is further assumed that profits per hectare follow a random walk so that $E_0[\mathbf{r}_{t+1}] = \mathbf{r}_t$, then the formula simplifies to $V = \mathbf{r}/r$, where $\mathbf{r} = \mathbf{r}_0$ (see Clark et. al, 1993).

⁵ Strictly speaking, this procedure is only an approximation. Since h enters effective prices linearly, $\log(p)$ cannot be additively separable in money prices. Even with accurate price data, it would be difficult to impose the full structure of the theoretical model in the estimation of the rent function. On the other hand, given the relatively low cost of trucking, it is unlikely that price variation along the main highway is sufficiently great to render this approximation inaccurate.

market centers may be located closer to more fertile land, where rents are higher. Given these considerations, I take a semi-nonparametric approach by assuming that $\log(\mathbf{r}) = \log \mathbf{q}(h) + \boldsymbol{\xi}' X + \mathbf{c}' M$, where \mathbf{q} is a nonparametric function, X is a vector of plot characteristics, and M is a vector of regional dummy variables.

As to the dependent variable in (8), Colwell and Munneke (1997) warn against using land value per-hectare because plot values may be nonlinear in area. If so, and if parcel size is correlated with distance to market, then such a specification will lead to a spurious correlation between land value and distance. Including the log of plot area, A , to account for this nonlinearity in the empirical specification of (8) gives

$$\log(V) = \log \mathbf{q}(h) + \mathbf{b} \log(A) + \boldsymbol{\xi}' X + \mathbf{c}' M + u \quad (9)$$

which nests the value per-hectare specification when $\mathbf{b} = 1$. Note, $\log \mathbf{q}(h)$ absorbs $\log(r)$, and the error term, u , reflects unobserved attributes of the plot.

Following Robinson (1988) and Stock (1991), equation (9) can be estimated by first using bivariate kernel regressions to “partial out” h from both sides. Since the number of kernel regressions required equals one plus the dimensionality of (A, X, M) , this method is computationally very expensive, especially given the large sample. A much cheaper approach uses the fact that h is a discrete valued regressor in the data with k distinct values. In the first stage, I include a $k - 1$ vector of dummy variables D in (9) for each value of h ; i.e., I replace $\log \mathbf{q}$ with $\mathbf{w} = \sum_{j=1}^{k-1} \mathbf{x}_j D_j$. Applying ordinary least squares (OLS) to this regression yields a consistent estimate of $(\mathbf{b}, \mathbf{g}, \mathbf{d}, \mathbf{x})$. In the second stage, I calculate $\bar{\mathbf{w}}$ for each observation and run a nonparametric regression of this variable against h to get a smoothed estimate of \mathbf{q} .

IV. Data and Preliminary Analysis

Nepal Living Standards Survey

The data for this study come from the 1995-96 Nepal Living Standards Survey (NLSS), a nationwide multi-topic survey collected by the Central Bureau of Statistics assisted by the World Bank (Central Bureau of Statistics, 1995). A stratified random sample of around 3400 households was drawn from four zones: Mountains, urban Hills, rural Hills, and Terai. In addition, a special sample of 1200 households was surveyed in the Arun valley (rural Hills), which I include in my analysis.

The NLSS contains a detailed agricultural module including information on the use of modern inputs, of which chemical fertilizer predominates, and on crop production and sales. In addition, the survey provides a listing of all plots owned or leased in by the household along with information on plot area, land quality, irrigation by season, net rent received by season (if leased out), and, of course, value of owned plots. The question on plot value reads as follows: “If you wanted to buy a plot exactly like this, how much would it cost you?” One indication of farmers’ awareness of land values, besides the fact that it is by far their most productive asset, is the frequency of land transactions, which is surprisingly high in the sample.⁶ Among the 3,621 landowning households, 5 percent bought land the previous year and 9 percent either bought or sold land. Moreover, although 85 percent of all plots are inherited, 28 percent of the landowning households purchased at least one of their current plots. Note also that any *nonsystematic*

⁶ Land transactions are sparse in most contexts. For this reason many hedonic studies use self-assessed land values (see, in particular, Mendelsohn, et al., 1994 and the citations in Colwell and Munneke, 1997) or housing values (see Bartik and Smith, 1987).

measurement error in plot values will not affect the coefficients of a regression in which plot value is the dependent variable.

A unique feature of the NLSS questionnaire is that it collects information at the household level on access to 14 different facilities. For each facility, the survey asks about travel time (in minutes, hours, and days) and mode of transport; i.e., by foot (without load), bicycle, motorcycle, car/bus, and mixed (foot+vehicle). Keep in mind that collecting data on actual distance, even using satellite telemetry, would be of little value in mountainous Nepal (except perhaps in the Terai). However, I do not use the household-level travel time information directly. Instead, I take the median of travel times by “wards” (in rural areas these are villages and environs) based on households that report travel times by foot, which the great majority do. The advantage of this procedure is that it standardizes travel times for mode of transport, which is potentially endogenous, and it mitigates the measurement error that is likely to be present in household level travel times.

I focus on market centers and agricultural cooperatives, since these facilities are the most likely to offer the opportunity to sell output and purchase modern inputs. In fact, over two-thirds of the households who report using chemical fertilizers obtained them from agricultural cooperatives, and most of the rest from private traders. My measure of h is the *minimum* of ward- median travel times to the market center and agricultural cooperative. Ideally, travel times to all relevant facilities should be considered separately, but this would lead to multicollinearity problems, not to mention making nonparametric estimation practically impossible. Median travel time to a market center or a cooperative in the sample of 3,724 cultivating households is 2 hours (mean=2.8 hrs.). Median travel

times are shorter on the plains of the Terai (1.25 hrs.) than in the Hills (2 hrs.) or Mountains (3 hrs.).

Finally, it is necessary to define a market area within which money prices (i.e., p , v , and w) do not vary. I take each district to be a distinct market; 73 out of the 75 districts in the country are represented in the NLSS sample and there are on average about 4 wards in a district (except in the oversampled Arun valley where this number is much higher). Though somewhat arbitrary, identifying a market by a district is consistent with evidence from the village questionnaires attached to the NLSS. These data show that very few villages have either market centers or agricultural cooperatives located in the same ward; households living in several different wards share these facilities.

Analysis of Fertilizer Use and Crop Sales

If proximity to markets influences land values through the effective prices of agricultural inputs and outputs, then purchases of modern inputs and sales of output should decline with distance from the market center. According to the model, observed fertilizer use per hectare, allowing for corner solutions, is $x^* = \max\{x(w, \tilde{v}, \tilde{p}), 0\}$. A similar equation holds for observed total crop sales per hectare, s^* , except that sales also depend on household consumption decisions. Analyzing marketed surplus (s^* minus consumption) is problematic because transport costs drive a wedge between selling and purchase prices and net sellers respond differently to variation in these transport costs than do net buyers (see, e.g., Omamo, 1998). Examining crop sales alone focuses on the selling decision, which is my primary interest. It is also probably reasonable to assume

that for net sellers consumption is relatively unresponsive to transport costs, given that income and substitution effects work in opposite directions.

Figures 4 and 5 plot nonparametric regression estimates of fertilizer use per hectare and crop sales per hectare, respectively, against travel time. The econometric specification is similar to equation (9) and the sample consists of 3,712 cultivating households.⁷ Since both fertilizer use and crop sales are heavily censored at zero, I estimate the first-stage models by tobit, in addition to OLS.⁸ For the tobits, I drop observations with perfect classification; i.e., where all cases of a given value of h or of a given district are censored (167 observations for fertilizer, 6 for crop sales).⁹ The second-stage nonparametric regressions are estimated using the LOWESS smoother with bandwidth=0.8. The choice of smoother is dictated by its robustness to outliers and by the fact that the 60 successive values of h are not equally spaced, which can lead to biases in kernel regressions (Fan, 1992).

The figures show that fertilizer purchases and crop sales per hectare decline steadily beyond travel times of about one hour. It is unclear why the curves are increasing for travel times of less than an hour, but one reason might be that cultivation is typically less intensive near urban areas. Also, comparing the curves based on OLS and tobit first-stage regressions indicates that accounting for censoring makes little substantive

⁷ Twelve households are dropped because they are uniquely identified by their district and their value of h .

⁸ The first-stage parameter estimates are suppressed for brevity. To summarize, education of the head and the set of district dummies are jointly significant in all regressions, and the demographic variables (number of adult males and females and male and female children), included only in the crop sales equation to capture consumption variation, are also jointly significant.

⁹ The high rate of censoring within certain districts precludes the use of semiparametric methods that are robust to deviations from normality. In particular, the censored LAD estimator fails to converge when

difference. In sum, this analysis supports the notion that transport costs influence farm profits through input use and crop marketing decisions. The fertilizer result, in particular, confirms the importance of the intensive margin of cultivation, implying that the farm profit function, and hence the rent function, should be convex.

Analysis of Plot Values and Rents

Underlying the asset-pricing formula given by equation (8) are several strong assumptions about land and credit markets,¹⁰ which may not hold in Nepal. To test equation (8), along with the validity of self-reported plot values, I compare values with rents received on plots that are leased out (mainly sharecropped). My analysis is based on a sample of 381 plots that were either rented out during both agricultural seasons, or were rented out in the wet (the main growing season) and left fallow in the dry. It should be noted, however, that about twice this many plots (around six percent of all those owned) were rented out during at least one season. Net rents are summed across both seasons, and include the value of in-kind payments while netting out the cost of inputs provided to tenants. The median rent to value ratio (r/V) is 0.055, which can be viewed as an estimate of the discount rate r . Formally, I run the regression

$$\log(V) = \mathbf{h}_0 + \mathbf{h}_1 \log(\mathbf{r}) + \mathbf{z} \quad (9)$$

and test whether $\mathbf{h}_1 = 1$.

district dummies are included in the specifications. The overall rates of censoring in the samples are 43 percent for fertilizer and 49 percent for sales.

¹⁰ For example, where land is the sole form of collateral for loans, its price may reflect its collateral value in addition to the capitalized value of the stream of rents (see Chalamwong and Feder, 1988).

Table 1 reports a series of estimators of h_1 that make progressively less restrictive assumptions about the form of correlation between $\log(\mathbf{r})$ and the error term \mathbf{z} . The OLS estimate in column (1) assumes that $\log(\mathbf{r})$ and \mathbf{z} are uncorrelated, and it falls well short of unity. However, one reason for this low estimate could be attenuation bias due to random measurement error in rents. Specification (2) thus instruments rents with plot area, which does indeed raise the estimate of h_1 , though it remains marginally below unity. Specification (3) corrects for the possibility that credit market conditions (i.e., $\log(r)$) and rents might covary across markets by including district fixed effects (using only districts that contribute more than one plot to the sample); h_1 is still precisely estimated, but it is no longer statistically different from unity. Finally, specification (4) includes household fixed effects, using the 90 households that contribute more than one plot to the sample. This estimator corrects for the possibility that *household*-specific interest rates and rents are correlated, as well as for any selection bias induced by restricting the sample to those households that rent out land. This last estimate of h_1 is again indistinguishable from unity, so that the maintained assumptions of the asset- pricing model and of no systematic reporting bias in plot values cannot be rejected.

V. Main Results

Plot Values and Distance to Market

Information is available on a total of 13,672 plots owned by 3,621 households. Plot characteristics include suitability for rice cultivation (khet land), whether irrigated, and if so whether seasonally or year round, mode of irrigation (tubewell, canal, other),

system of irrigation (self-managed, agency managed, community managed), and the “quality” of the plot based on a four grade classification used by the land revenue department for tax assessments. Land value is missing for 13 plots and quality for 98 plots, so the plot value regressions are based on 13,651 observations. However, when predicting plot values based on the regression results, these missing observations can be recovered by imputing plot quality with its modal value.

Table 2 reports two specifications of the plot value regression. The first assumes that $q(h) = h$ and the second performs the first-stage in the semi-nonparametric estimation described in Section III. In both specifications, the plot characteristics have significant and sensible coefficients; for example, plots that are suitable for rice, plots with year round irrigation,¹¹ high quality plots, and, of course, larger plots are more valuable. However, the value per-hectare specification ($b = 1$ in equation (9)) is resoundingly rejected, which is important because plot size turns out to be negatively correlated with distance to markets. Thus, restricting b to one would have led to an underestimate of the rate of decline of land values with travel time (Colwell and Munneke, 1997). As it is, travel time is strongly negatively associated with land values.

Figure 6(a) plots the $\bar{q}(h)$ derived from the two specifications in Table 2 (normalized by taking deviations from means). In the first case, $\bar{q}(h) = h^{-0.222}$, which is obviously convex. The nonparametric estimate of $q(h)$ (LOWESS; bandwidth=0.8) is also roughly convex, except that the function is increasing above travel times of 12 hours,

¹¹ Since irrigation is an investment, which may be responsive to distance to market, I also ran the plot value regressions unconditional on the irrigation variables. In this case, the coefficient on log travel time is -0.225 (0.0217), which is almost identical to its value of -0.222 in specification (1) of Table 2.

though only 5 percent of plots are located this far from markets.¹² To formally compare the parametric and nonparametric models, I use the bootstrap, drawing 100 percent random samples in two stages, first by sampling plots within households and then by sampling households. On each replication, the two models are run and the difference in the slope of $\bar{q}(h)$ is calculated at each value of h . As a test of whether the actual slope differences in Figure 6(a) are significantly different from zero, Figure 6(b) plots, at each h , the ratio of the actual slope difference to the standard deviation over the 100 bootstrap replications. When the absolute value of this ratio exceeds 1.96, the equality of slopes can be rejected at the 5 percent level. That this hypothesis is rejected near the endpoints is perhaps not too surprising. However, the nonparametric estimate is also significantly steeper (more negatively sloped) than the parametric estimate in much of the one to three hour travel time range, where the data are most dense. Although other parametric models might fit the data better, this test shows that there is sufficient power to reject a reasonable parametric alternative, thus supporting a nonparametric approach.

Benefits and Distributional Consequences of Road Projects

The expected appreciation in value of a given plot as travel time falls from an initial value of h_0 to a value of h_1 is

$$\left[\bar{q}(h_1) - \bar{q}(h_0) \right] \exp \left[\bar{b} \log(A) + \bar{g}X + \bar{d}M \right] E[\exp(\bar{u})] \quad (10)$$

The last term in this expression takes into account unobserved heterogeneity in plot values, and can be estimated nonparametrically as the average exponentiated residual from

¹² Bandwidth choice is subjective, but the main features of the nonparametric fit are robust to bandwidth.

(9). Although nonparametric regression only provides an estimate of q at actual values of h , $\bar{q}(h_1)$ can be estimated by linear interpolation between known values of \bar{q} . To simulate the benefits from the road project discussed in Section II, let $h_1 = mh_0$ for $h_0 < I$ and $h_1 = h_0 - I(1 - m)$ for $h_0 \geq I$. I set the value of m at 0.1 to reflect the fact that the cost per ton-km of headloading in Nepal is roughly ten times the cost of trucking.¹³ Total household (capitalized) benefit from the project is simply the sum of the appreciations in value of each of the household's plots.¹⁴

Following standard practice, I use total household consumption expenditures as a measure of income and per-capita expenditures, adjusted for regional price differences, as a measure of welfare. Let us say that a road project is progressively (regressively) targeted if the ratio of benefit to total household expenditures, the benefit ratio, is decreasing (increasing) in per-capita expenditures. Figure 9 plots nonparametric regression curves (LOWESS; bandwidth=0.8) of benefit ratios against log per-capita expenditures for three road projects, corresponding to three values of I . The estimates are based on the full sample of 4,573 households, which includes both households who do not cultivate and those who do not own land.¹⁵

¹³ This number is based on information in Walters (1968), but it is still approximately valid according to World Bank transport economists familiar with Nepal.

¹⁴ In the few cases where the reduction in travel time occurs on the increasing portion of \bar{q} , benefit is set to zero. Note that it is not possible to calculate benefits from the additional land brought under cultivation as h^* increases. However, as long as one considers a marginal road extension, these benefits are an envelope phenomenon and hence are zero.

¹⁵ All statistics reported for this sample (e.g., the expenditure deciles in Figure 7) are calculated using population weights to insure that they are nationally representative. Ninety-three percent of the households (on weighted basis) live in rural areas.

Figure 7 shows that as the road is extended farther up the “representative” valley, benefits are targeted more progressively. Indeed, the distribution of benefits from a short road, one that extends a mere 1.5 hours walking time up the valley, is slightly regressive. Sixty percent of the sample would lie along such a road, whereas 85 percent would lie along the 3-hour road and 98 percent would lie along the 8-hour road. The key factor behind the increasing progressivity is the strong tendency for poorer households to live farther away from markets, as confirmed by the nonparametric regression curve in Figure 8. Evidently, this factor dominates the opposing tendency of poorer households to have less valuable landholdings, also illustrated in Figure 8.

While suggestive, Figure 7 does not address the question of whether the benefits of a hypothetical road project are sufficiently large and distributed sufficiently progressively to reduce overall income inequality. To tackle this question first requires converting the capitalized benefits into a permanent income flow, which in turn requires an assumption about the discount rate, r . Suppose that current household expenditures equal permanent income. For a pure farm household, with no off-farm employment and which leases in no land, permanent income equals rV_{TOTAL} , where V_{TOTAL} is the sum of the value of all plots owned; more generally, income comes from other sources as well. Thus, in a regression of total household expenditures on V_{TOTAL} , the coefficient on V_{TOTAL} should equal the discount rate r , with other sources of income consigned to the residual. A slight refinement of this regression, which includes ward (village) fixed effects and which instruments V_{TOTAL} for measurement error using total land area owned, gives $\hat{\beta} = 0.058$

with a standard error of 0.010.¹⁶ This estimate of the discount rate is remarkably close to that derived earlier from the rent to value ratio (i.e., 0.055). In the calculations that follow, I set $r = 0.06$.

Denote household benefits per capita from a road of length I by $B(I)$. Per capita expenditures (permanent income) as a function of road length is thus $z(I) = z_0 + rB(I)$, where z_0 is baseline per capita expenditures. Assuming that $y(z) = z^{1-e}$, $e > 0$, the social welfare function (see (5)) can be written as $W(I) = [\bar{z}(I)(1-I(I))]^{1-e}$, where \bar{z} is mean per capita expenditures and $I \in [0,1]$ is Atkinson's measure of inequality.¹⁷

Differentiating with respect to I gives

$$\frac{W'(I)}{W(I)} \propto \frac{\bar{z}'(I)}{\bar{z}(I)} - \frac{I'(I)}{1-I(I)} \quad (10)$$

In other words, the rate of increase in social welfare as the road is extended can be decomposed into the rate of increase in mean income and the rate of increase in income equality, $1-I$. An assessment of whether road building has important distributional consequences can be made by comparing the magnitudes of these two components at reasonable values of the inequality aversion parameter e . Specifically, define

$$\Omega(I; e) = -\frac{I(I) - I_0}{1 - I_0} \bigg/ \left(\frac{\bar{z}(I) - \bar{z}_0}{\bar{z}_0} - \frac{I(I) - I_0}{1 - I_0} \right) \quad (11)$$

¹⁶ The OLS estimate of r , which does not account for measurement error, is substantially smaller at 0.0076 (0.0007). These regressions are based on the sample of 3,621 households that own land.

¹⁷ Recall that $I = 1 - \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{z_i}{\bar{z}} \right)^{1-e} \right]^{1/(1-e)}$ for $e \neq 1$ where N is the number of individuals. When

$e = 1$, $W(I) = \log[\bar{z}(I)(1-I(I))]$, where $I = 1 - \prod_{i=1}^N \left(\frac{z_i}{\bar{z}} \right)^{1/N}$.

where the zero subscript denotes values prior to the road project. $\Omega(I; \mathbf{e})$ represents the contribution of reduced income inequality to the increase in social welfare from the building of a road of length I .

Figure 9 plots $\Omega(I; \mathbf{e})$ using the Nepal data for I 's ranging from 0.5 to 10 hours, beyond which value practically every household is on the road so that $W'(I) \approx 0$. Note that putting all households in the sample on a road would raise \bar{z} by ten percent, quite a substantial gain in permanent income.¹⁸ Interestingly, at high values of inequality aversion ($\mathbf{e} = 4$), building a short road actually increases income inequality ($\Omega < 0$). For any value of \mathbf{e} , the contribution of inequality reduction to the increase in social welfare rises with the length of the road, again because of the strong tendency for the poor to live in more remote areas. However, unless one chooses a value of \mathbf{e} above 2 or so, which is usually considered rather large, the increase in social welfare is due overwhelmingly to higher mean income. Rural road construction thus appears to be like a tide that lifts all boats rather than a highly effective means of reducing income inequality.

VI. Conclusion

Transport infrastructure plays a central role in rural development, but the distributional consequences of rural roads have not received adequate theoretical or empirical attention. This paper develops and implements a method for nonparametrically estimating the benefits from road projects at the household level and for examining the distribution of these benefits across income classes. The findings for Nepal suggest that

¹⁸ Keep in mind that because I have weighted the sample to represent the population of Nepal, all of these calculations take proper account of the fact that population is less dense in more remote areas.

providing extensive road access to markets would confer substantial benefits on average, much of these going to poor households. However, the benefits would not be large enough or targeted efficiently enough to appreciably reduce income inequality in the population (unless there is an exceptionally high degree of inequality aversion). Thus, while my analysis may paint a more optimistic picture than the World Bank's (1994, p. 80) general assessment that infrastructure is "a blunt instrument for intervening directly on behalf of the poor," rural road construction is certainly not the magic bullet for poverty alleviation.

Another lesson of this research is that data on land values and characteristics, particularly at the plot level, can be extremely useful in measuring the benefits of infrastructure investments in LDCs, and not just of rural roads. To be sure, land may not always behave like a typical asset, so that benefit capitalization may be imperfect, but the asset-pricing model can be tested to determine whether the methodology developed in this paper is appropriate in a particular context.

Finally, it is important to mention the other benefits of rural roads besides cheaper transport to and from agricultural markets, such as better access to schools and health facilities and, more generally, to a greater variety of consumer goods. Insofar as farmers prefer to live near their farms, at least some of these gains are likely to be capitalized in farmland values. Separating out these distinct benefits of rural roads is left as a topic for future research.

References

- Bartik, Timothy J. and V. Kerry Smith. 1987. "Urban Amenities and Public Policy," Chapt. 31 in *Handbook of Regional and Urban Economics*, Vol. II, ed. by Edwin Mills. Amsterdam: North Holland.
- Beenhakker, Henri L. and Armando M. Lago. 1983. *Economic Appraisal of Rural Roads*. World Bank Staff Working Papers No. 610. Washington: The World Bank.
- Central Bureau of Statistics. 1995. *Nepal Living Standards Survey: Interviewers Manual*, June. Washington: The World Bank.
- Chalamwong, Yongyuth and Gershon Feder. 1988. "The Impact of Landownership Security: Theory and Evidence from Thailand," *World Bank Economic Review*, 2(2):187-204.
- Clark, J. Stephen, Murray Fulton, and John T. Scott, Jr. 1993. "The Inconsistency of Land Values, Land Rents, and Capitalization Formulas," *American Journal of Agricultural Economics*, February: 147-55.
- Colwell, Peter F. and Henry J. Munneke. 1997. "The Structure of Urban Land Prices," *Journal of Urban Economics*, 41:321-36.
- Fan, Jianqing. 1992. "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87:998-1004.
- Gersovitz, Mark. 1989. "Transportation, State Marketing, and the Taxation of the Agricultural Hinterland," *Journal of Political Economy*, 97(5):1113-37.
- Howe, John and Peter Richards. 1984. *Rural Roads and Poverty Alleviation*. London: Intermediate Technology Publications.
- Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Analysis," *American Economic Review*, 84(4):753-71.
- Omamo, Steven Were. 1998. "Transport Costs and Smallholder Cropping Choices: An Application to Siaya District, Kenya," *American Journal of Agricultural Economics*, forthcoming.
- Robinson, P. M. 1988. "Root- N -Consistent Semiparametric Regression," *Econometrica*, 56(4):931-54.

Stock, James H. 1991. "Nonparametric Policy Analysis: An Application to Estimating Hazardous Waste Cleanup Benefits," in William A. Barnett, James Powell, and George Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge: Cambridge University Press.

Walters, Alan A. 1968. *The Economics of Road User Charges*. World Bank Staff Occasional Papers No. 5. Washington: The World Bank.

World Bank. 1994. *World Development Report 1994: Infrastructure for Development*. New York: Oxford University Press.

van de Walle, Dominique. 1996. *Infrastructure and Poverty in Viet Nam*, LSMS Working Paper, No. 21. Washington: The World Bank.

Table 1
Plot Values and Rents

	(1) OLS	(2) IV ^b	(3) IV ^b	(4) IV ^b
η_1	0.658	0.825	1.047	1.034
Standard Error	0.064 ^a	0.108 ^a	0.091	0.115
H ₀ : $\eta_1 = 1$ (p-value)	0.000	0.105	0.606	0.767
Fixed Effects	none	none	district	household
N	381	381	373	278

^aRobust standard error accounts for district-level clustering.

^bLog of plot area is the excluded instrument.

Table 2
Plot Value Regressions

	Means	(1)	(2)
Log(hours travel time)	3.47 (4.01) ^a	-0.222 (10.3)	---
H ₀ : $\mathbf{x} = 0$ (p-value) ^b		---	0.0000
Log(area in hectares)	0.298 (1.16) ^a	0.558 (37.8)	0.547 (37.4)
Suitable for rice (khet)	0.434	0.149 (3.89)	0.144 (3.84)
<i><u>Irrigation:</u></i>			
seasonal	0.175	0.341 (4.55)	0.325 (4.47)
year-round	0.138	0.497 (6.29)	0.450 (5.88)
canal	0.239	0.060 (1.0)	0.052 (0.88)
tubewell	0.030	-0.034 (0.41)	-0.061 (0.74)
self-managed	0.189	-0.082 (1.664)	-0.051 (1.09)
agency managed	0.018	-0.219 (1.663)	-0.209 (1.67)
<i><u>Quality:</u></i>			
dwaim	0.240	-0.250 (5.27)	-0.226 (5.02)
sim	0.304	-0.499 (9.04)	-0.465 (8.84)
chahar	0.365	-0.833 (13.9)	-0.792 (13.9)
District dummies (p-value)		0.0000	0.0000
R ²		0.546	0.568

Notes: Robust t-values accounting for household-level clustering in parentheses. Omitted categories: pakho/bari (dry) for khet; non-irrigated for irrigation; other for irrigation mode; community managed for irrigation management; and awal (highest grade) for quality. Sample size is 13,561 plots belonging to 3,586 households.

^aStatistics are for levels. Standard deviations are in parentheses.

^bJoint significance of the dummy variables for each of the 60 values of h .

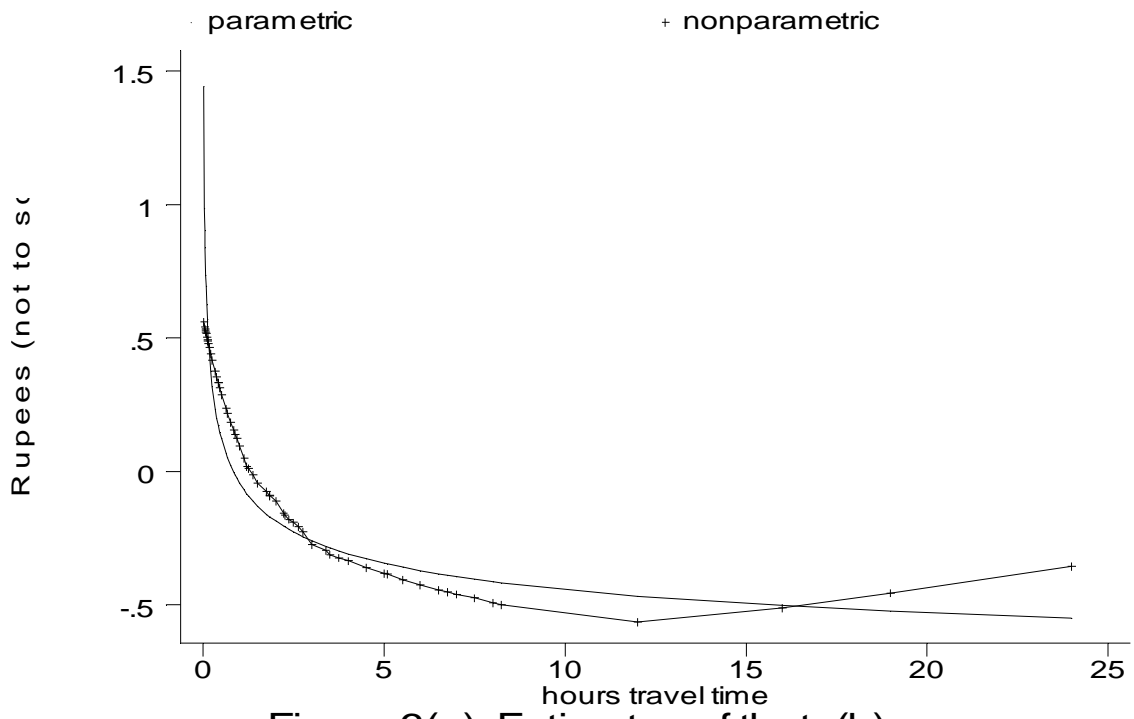


Figure 6(a): Estimates of theta(h)

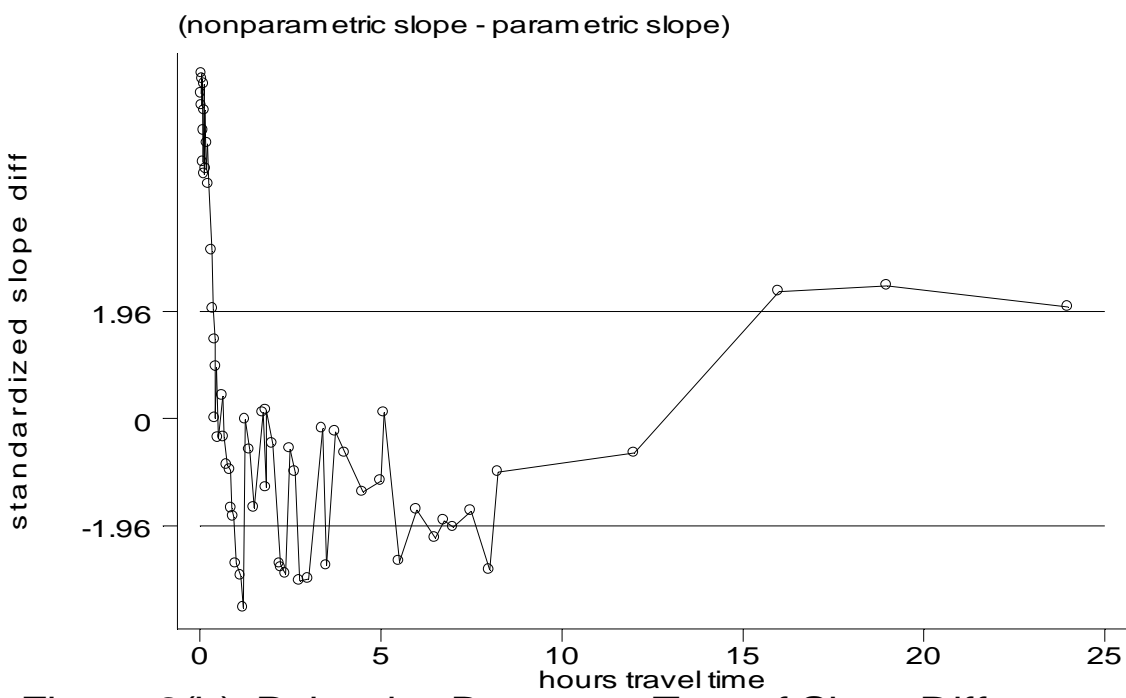


Figure 6(b): Pointwise Bootstrap Test of Slope Difference

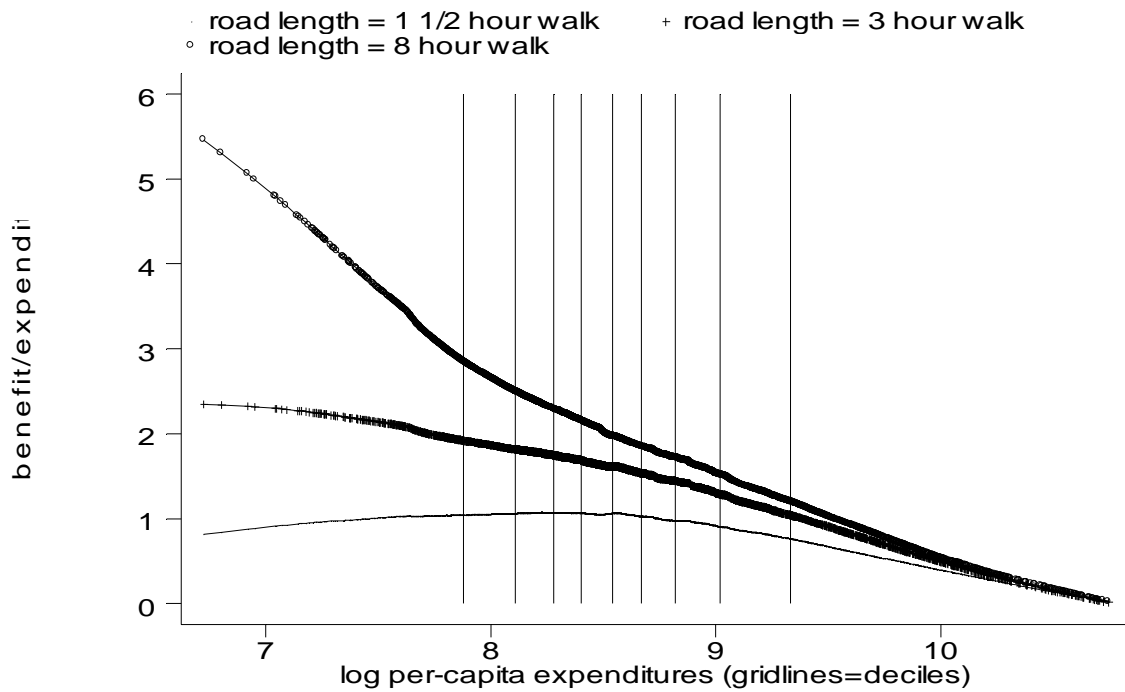


Figure 7: Distribution of Benefit from Road Projects

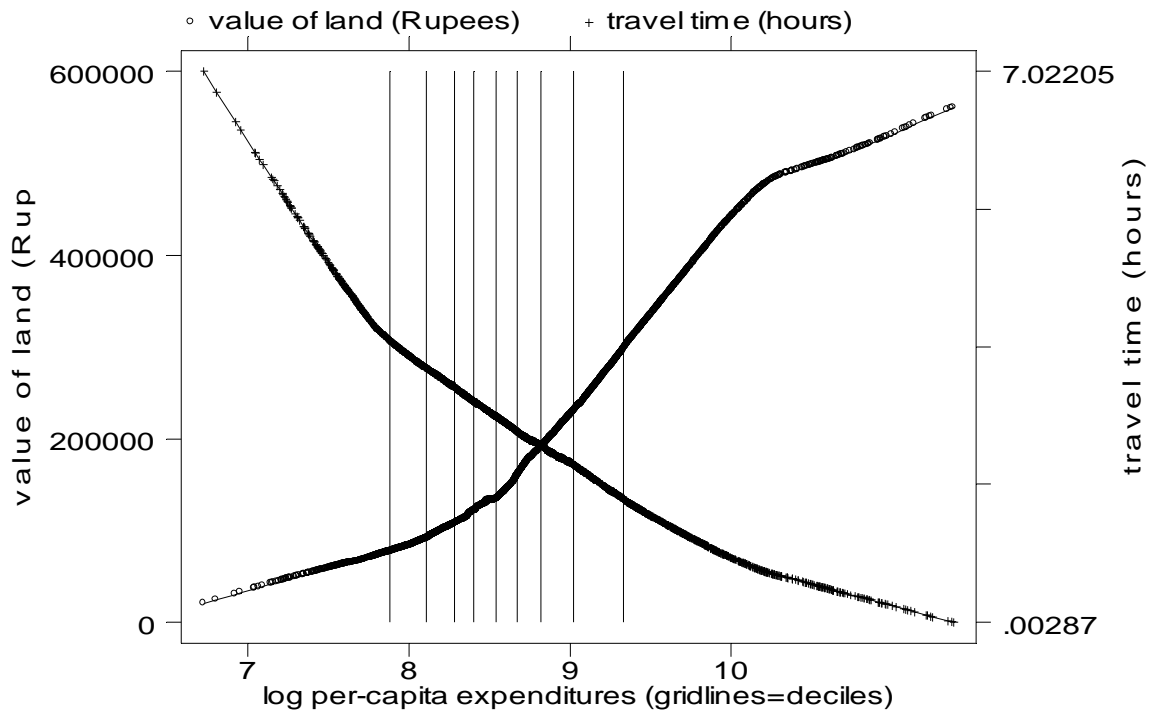


Figure 8: Land Value and Travel Time vs. Income

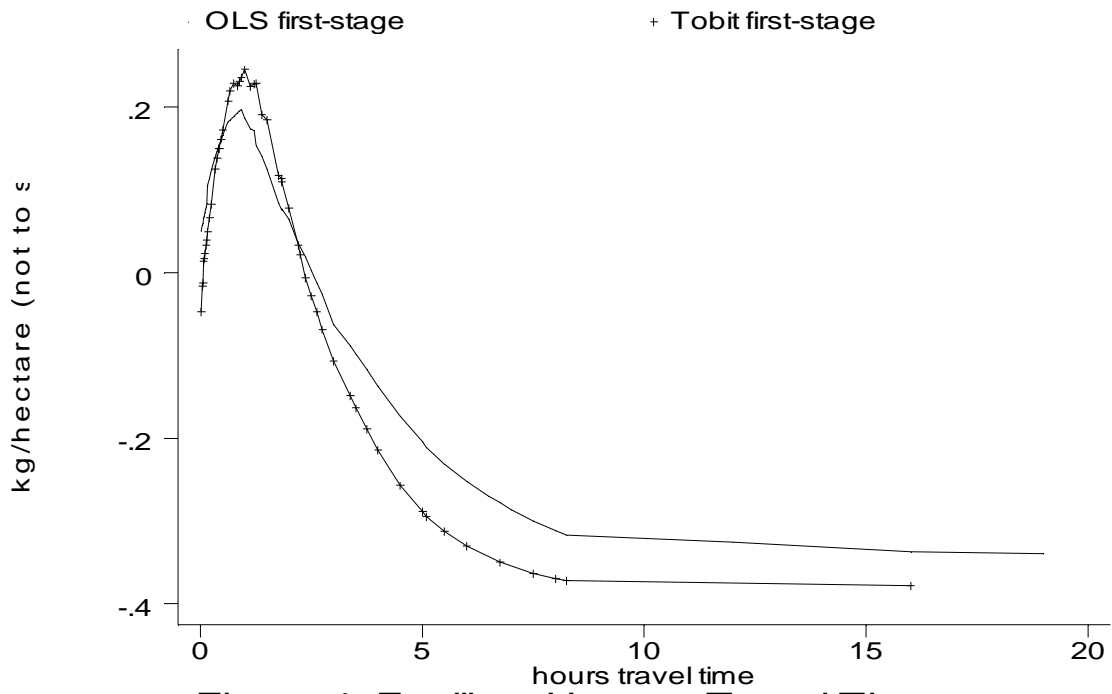


Figure 4: Fertilizer Use vs. Travel Time

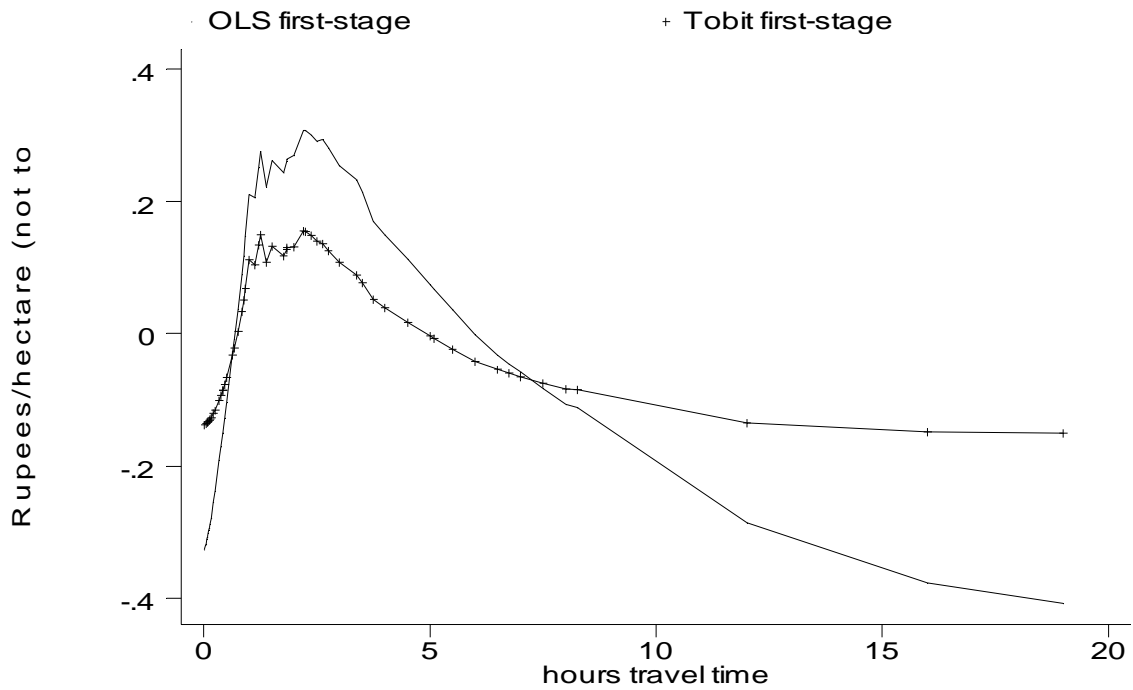


Figure 5: Crop Sales vs. Travel Time