Infrastructure, services, and smallholder income growth: evidence from Kenyan panel data

By:

Jordan Chamberlin

Invited paper presented at the 4th International Conference of the African Association of Agricultural Economists, September 22-25, 2013, Hammamet, Tunisia

Infrastructure, services, and smallholder income growth: evidence from Kenyan panel data

Corresponding author: Jordan Chamberlin

Affiliation: Michigan State University (currently: PhD student; by the time of the conference I will be an Assistant Professor with the same institution).

Length of text (excluding abstract, tables and references): 6765 words.

Abstract

This essay investigates the relationship between rural household income growth and access to electricity, extension services and roads. Following Dercon et al (2009), I outline a household income growth model that includes access to these public goods as growth conditioners. I estimate the model using panel data on Kenyan smallholders covering the period 2000-2010. I find that the expansion of electricity, feeder roads and agricultural extension services are all important conditioners. Access to agricultural extension services has a particularly strong impact on growth.

As an extension of the growth model, I investigate whether spatial spillovers in public goods exist and to what extent these spillovers contribute to household growth. I find strong evidence of spatial dependence in the growth process. Specification tests support a spatial Durbin model which allows for both endogenous spatial dependence in outcomes and exogenous spatial dependence through access conditions in neighboring areas. Direct and indirect spatial spillover effects have the greatest impact through feeder roads, suggesting that conventional (non-spatial) estimates of partial effects may be particularly prone to downward bias. Results from this study also indicate the importance of rural services and electricity to household growth outcomes, suggesting that studies of rural infrastructure which focus exclusively on road networks will miss important dimensions of rural accessibility and economic remoteness.

1. Introduction

Enabling smallholder growth is a key development policy objective in sub-Saharan Africa, where a majority of the rural population is characterized by low levels of farm productivity, household income and asset wealth. Furthermore, different expressions of growth are related to one another: without growth in farm productivity, household asset accumulation may be constrained; low levels of productive assets may also constrain investments which are necessary for achieving productivity growth.

A key public policy response to this challenge is the provision of growth-enabling public goods. The role of public goods in enabling smallholder farm growth is generally characterized as follows: public goods enhance the productive potential of privately-held productive assets by increasing the marginal productivity of such assets. Infrastructure that facilitates transportation and communication, such as roads and telephone services, lower the costs of accessing input and output markets. Reduced access costs imply relaxed constraints on the decision-making problems faced by utility maximizing farm managers. Theoretically, this should enable farmers to make more efficient production and marketing decisions. As a consequence, improved infrastructure and other public assets should enable farm productivity growth and household asset accumulation.

This paper empirically evaluates the role of public assets as a conditioner of asset growth experienced by small farm households in Kenya over the last 10 years. I examine a key household-level component of the rural development process: household asset accumulation. I offer a conceptual model in which asset accumulation is a function of farm productivity and household income, which are strongly conditioned by the provision of public assets. The public assets I include in my analysis include: road, electricity and telecommunications infrastructure, and agricultural extension services.

A distinguishing feature of this work is the incorporation of spatial dependence in a panel modeling framework. Micro-econometric models typically assume independence of observations in the cross-section. However, recent growing amount of empirical studies on smallholder decision making show important local interdependence of decision making outcomes (much of this comes from the literature concerned with technology adoption). Because the growth processes of interest here are directly related to technology and farm management decisions, I test and control for spatial dependence in the growth outcomes of interest, as well as in the error terms of our models. I note that without such controls, estimation results may be severely compromised: if the data generating process is characterized by spatial dependence in the dependent variable, standard linear approaches to estimation (e.g. using fixed or random effects estimators) will give inconsistent estimates; if the model is characterized by a spatial error component, standard estimators will be inefficient, at best, and may be inconsistent if the error structure derives from an omitted covariate with a spatial expression.

The remainder of this paper is structured as follows: the next section offers a conceptual framework. Thereafter, I describe the empirical model. I then describe estimation results and discuss their implications for our understanding of the role of public goods in achieving targeted smallholder growth trajectories.

2. Rural household income and public goods in Kenya 2000-2010

Rural households in Kenya have experience only moderate nominal income growth, on average, over the past decade (Table 1). In real terms, incomes have actually declined in many places. To be certain, some households have experienced growth, but the overall trends have been disappointing. Raising rural incomes is a strategic priority for the government (GOK 2007, 2002).

One of the root causes of rural poverty is remoteness (Stifel et al 2003, Stifel and Minten 2008, Barret 2008). Key mechanisms underlying this relationship include higher cost of consumption goods and agricultural inputs, lower prices for agricultural outputs, greater price volatility and greater costs of accessing public services. In the case of Kenya, there is considerable evidence that remote households are less likely to be engaged in markets, less likely to have income from non-farm sources, and are more likely to be poor (e.g. Barret et al 2001, Renkew et al 2004). A negative relationship between income and distance from towns is fairly pronounced beyond thresholds of about 10km or so (Figure 1).

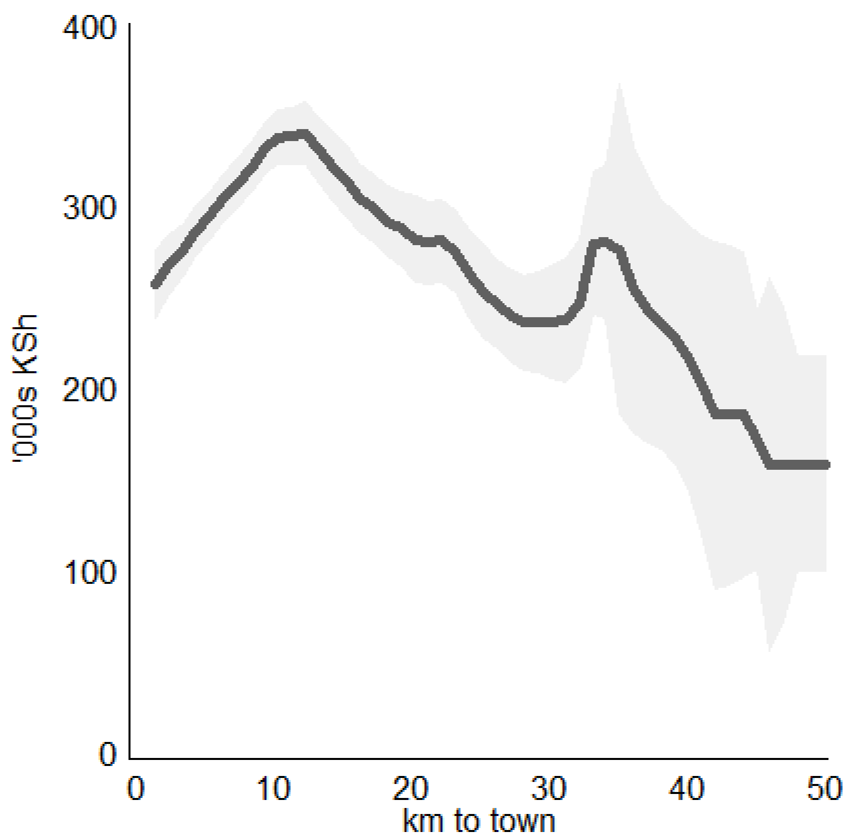
One of the strategic mechanisms for increasing agricultural incomes and reducing poverty in Kenya has been investments in rural infrastructure and other public goods. Key investment targets include expansion or improvements to infrastructure, such as electricity, telecommunications and transportation networks, and better access to education, health care, agricultural extension and other public services. Over the past decade such investments have taken place throughout the country and access improvements have been documented for a range of indicators. Chamberlin and Jayne (2009) show that the provision of physical infrastructure and key services has been gradually expanding for most areas of the country. However, with few exceptions, the pace of expansion has been incremental. Figure 2 shows boxplots representing the reported distances between rural households and a variety of public goods: the nearest source of electricity, agricultural extension office, all-weather (tarmac) and feeder road. In the long run, improvements are clearly taking place, although the inter-period improvements are often minor.

Table 1: Trends in household income among rural smallholders in Kenya 2000-2010

	2000	2004	2007	2010
nominal KSh				
income	149,196	161,757	171,375	234,515
per capita income	22,093	38,738	32,334	45,211
real (2010) KSh				
income	391,714	321,354	245,683	234,515
per capita income	58,006	76,959	46,354	45,211

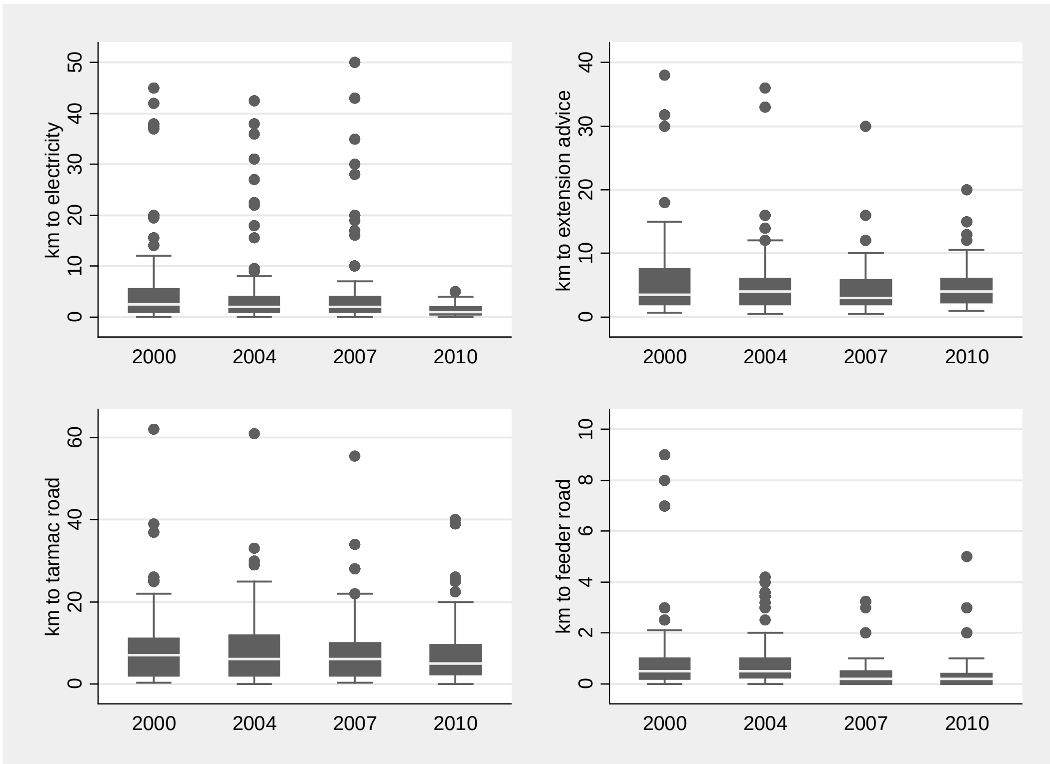
Note: per capita calculation based on the number of adult equivalents in the household in each year.

Figure 1: Household income and distance from the nearest district capital



Note: Data are in real 2010 Kenyan Shillings, pooled across the 2000, 2004, 2007 and 2010 rounds of the MSU/Tegemeo household survey. The dark line in the figure is a local polynomial estimate of household income over the distance gradient. The grey shaded area is the 95% confidence bounds.

Figure 2: Trends in distance to public services and infrastructure



Note: Box plots are based on the village-level averages of household reported distances. The dark shaded area in each plot represents the interquartile range (IQR: 75th – 25th percentile) of values reported for each year. The upper and lower bars are the most extreme values occurring within 1.5*IQR of the IQR.

Given these changes in Kenya's rural access landscape, is there any evidence of a positive impact on rural household income growth? On the face of it, the overall flatness of rural income growth, and the relatively modest improvements over time in access conditions would appear to suggest that investments in public goods have not had a major impact. However, aggregate statistics may obscure important patterns of change at the community and household level.

To investigate this relationship, I use household panel data covering 107 villages in all major production regions of the country.[1] I use 4 panel waves, collected in 2000, 2004, 2007 and 2010. Included in the dataset are a number of indicators of the distance to the nearest instance of a variety of infrastructure and services. Summary statistics of household income and other characteristics are presented in the table below. The majority of households in the sample derive their livelihoods primarily from agricultural production and marketing. Most households operate very small farms, with only moderate levels of input use and participation in output markets.

The average household is located about 4-5km from the nearest provision of electricity or extension services. Most households are several kilometers from an all-weather road, but are rarely located more than a kilometer from a feeder road. About half of the sample is located within 10km of a district capital, which is often the most important local market. Larger urban markets are usually much further away: the average estimated travel time from the farm gate is more than 12 hours.

Public goods such as infrastructure and extension services have important spatial expressions: they are located in particular places and their location influences how accessible they are to rural populations. Consider, for example, the relationship between distance from the nearest agricultural extension office and the probability of receiving extension advice, shown graphically in Figure 3. Physical expansion of service provision, therefor, corresponds to a reduction in distance for would-be rural beneficiaries.

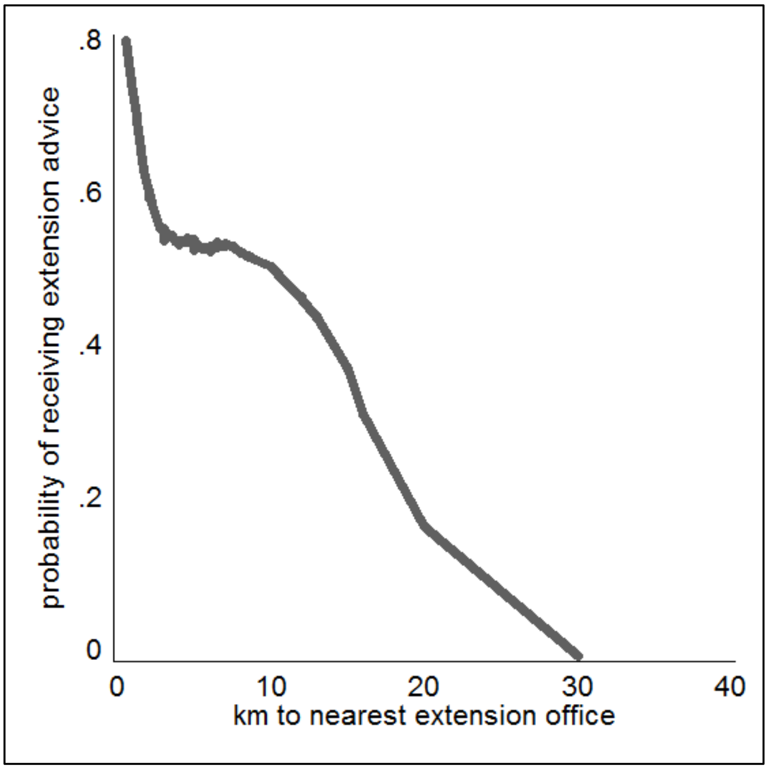
A second point to note is the close proximity of villages in the sample. If the impact of public goods provision extends out over space to non-trivial extents (i.e. beyond a few kilometers), then it is reasonable to assume that neighboring villages share some degree of access to the same conditions. This opens an additional avenue for potential impact of public good investments: the indirect impacts of improvements in public good provision in neighboring areas.

[1] This dataset has been described extensively by Chamberlin and Jayne (2009). See details of the sampling methodology in Argwings-Kodhek et al (1999).

Table 2: Household and community characteristics

		average values by year				*percentiles in 2010*				
	unit	**2000**	**2004**	**2007**	**2010**	**5th**	**25th**	**50th**	**75th**	**95th**
household characteristics										
income	KSh	149,196	161,757	171,375	234,515	28,765	76,250	151,963	285,676	705,664
per capita income	KSh/adult	22,093	38,738	32,334	45,211	5,305	14,211	29,113	55,648	138,429
off-farm income share	%	36%	37%	41%	40%	0%	12%	37%	67%	92%
marketed share of production	%	44%	42%	45%	41%	0%	13%	41%	67%	88%
high-value % of marketed output	%	35%	34%	30%	34%	0%	1%	13%	71%	100%
used inorganic fertilizer	1=yes	70%	72%	76%	75%	0	1	1	1	1
inorganic fertilizer use rate	kg/ha	49.92	46.42	48.02	44.65	0	0.03	35.35	70.81	133.53
land holding size	ha	5.66	5.76	5.41	4.78	0.75	1.75	3	5	15
adult equivalents	adult	7.28	4.76	6.17	6	1.66	3.96	5.7	7.68	10.96
female-headed household	1=yes	16%	20%	23%	27%	0	0	0	1	1
value of productive assets	KSh	25,529	8,364	41,892	49,827	0	2,200	7,050	23,300	214,500
age of household head	years	45.85	56.53	58.7	60.56	39	51	61	70	82
education of household head	years	5.07	6.2	6.31	6.48	0	3	7	10	15
community characteristics										
distance to electricity	km	4.82	4.27	4.07	1.32	0	0.5	1	2	4
distance to extension office	km	5.21	4.93	4.13	4.77	1.5	2.25	4	6	10.5
distance to tarmac road	km	7.85	7.77	7.5	6.9	0.3	2.25	5	9.5	20
distance to feeder road	km	1.05	0.85	0.35	0.3	0	0	0.2	0.4	1
distance to district capital	km	12.41	12.41	12.41	12.41	3	6	9.9	16.75	30
time to city of 50,000 or more	hours	3.97	3.97	3.97	3.97	0.31	1.72	3.42	5.58	11.09
rural population density	p/km2	308	353	382	408	109	244	379	502	744
average annual rainfall	mm	531	531	531	531	248	372	554	700	751

Figure 3: Relationship between location and access to agricultural extension services



Note: Data for 2010 only.

3. Conceptual framework

Smallholder household incomes are derived from a variety of sources, the most important of which are the value of farm production, earnings from participation in agricultural output markets, and from sales of household labor in agricultural and non-agricultural labor markets. Income may be used to satisfy consumption needs, as well as reinvested in production inputs, productive capital or human capital development, thereby improving future prospects for additional income growth and asset accumulation. The major channels of agricultural income growth are through increases in farm productivity, production specialization and/or market engagement. Non-farm income growth may result from returns to experience or specialized skills. The opportunities to engage in such pathways are conditioned by household endowments and, critically, the provision of public goods. For example, improved roads lower the cost of accessing input and output markets; access to agricultural extension services can contribute to productivity gains; and the provision of electricity enables a wider variety of rural non-farm industrial activities. The structural linkages in this system are complex. Here, I adopt a reduced form approach that relates growth outcomes with enabling conditions.

Let us denote household income by A, with A_{it} referring to the outcome for household i at time period t. Following Dercon et al. (2009), I model a household growth process $(\ln A_{it} - \ln A_{it-1})$ as a function of changes in capital stock of public goods (as proxied by distance to tarmac and feeder roads, electricity and agricultural extension services). I define an empirical growth model that allows for transitional dynamics, as follows. Consider changes in y over a time period of length p:

(1) $$\left(\ln A_{it} - \ln A_{it-p}\right)/p = \xi + \alpha \ln A_{it-p} + \beta \ln K_{it-p} + \gamma \left(\ln R_{it} - \ln R_{it-p}\right)/p + \lambda H + \\ + \mu_i + u_{it}$$

where ξ represents sources of growth common to all households. H reflects fixed characteristics of the household, such as location, that also affect growth. K represents exogenous levels of capital stocks (roads, electricity, extension), the term $(\ln R_{it} - \ln R_{it-p})$ represents transitory shocks such as changes in rainfall and prices, and μ_i denotes unobserved individual heterogeneity (i.e. a vector of time-invariant individual-specific effects). To deal with unobserved heterogeneity, I use FE to time-demean equation (1). In doing so, all time invariant explanatory variables drop out of H and $\mu_i$drops out completely.

As an extension of this model, I also allow for spatial dependence in the cross-section. Following Manski (1993), I consider three types of dependence: (a) endogenous interaction effects, whereby the income growth experienced by a household is dependent upon the income growth in neighboring households; (b) exogenous interaction effects, whereby the growth outcomes experienced by a household depend upon the provision of public goods in neighboring locations (as well as in their own locations); and (c) correlated effects related to unobserved factors affecting neighboring households in similar ways.

The first channel may be thought of as a kind of local growth spillover process, akin to the growth spillovers which have now been widely documented for contiguous regions (e.g. Ertur et al 2006, Le Gallo et al 2003, Richaud et al 1999). At the local level, empirical studies of growth spillovers have not yet been implemented, to my knowledge. However, such spillovers are consistent with growth multiplier effects operating through consumption and production linkages (Johnston and Mellor 1961, Mellor and Lele 1973, Haggblade et al 2007). Consumption linkages include spending by wealthier rural households on local consumption goods and services. As local incomes rise, the resulting increases in demand benefits local producers. Additionally, with increases in farm productivity (which both enables and is enabled by higher incomes) may drive growth in local non-farm activities, which may enhance off-farm income earning opportunities.

To implement this idea, let us redefine our outcome of interest as $y_{it} = \left(\ln A_{it} - \ln A_{it-p}\right)/p$ and define the right hand side elements of equation (1) as the vector $\boldsymbol{x}_{it}$, with i indexing cross-sectional observations $i = 1, \dots, N$ and t indexing time periods $t = 1, \dots, T$. We then write our model for a given observation as:

(2) $$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + u_{it}$$

Now, to allow for endogenous spatial dependence, let us rewrite this model as

(3) $$y_{it} = \delta \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{x}_{it}\boldsymbol{\beta} + u_{it}$$

where $\sum_{j=1}^{N} w_{ij} y_{jt}$ is a spatial lag of the dependent variable, $\delta(|\delta| < 1)$ is the spatial lag parameter. This term represents the endogenous spatial dependence in growth outcomes for neighboring households.

The other two channels of spatial dependence (exogenous dependence and correlated effects) may be thought of as aspects of a spatial diffusion process through which changes in infrastructure act upon the growth enabling conditions throughout a local region. Given the close proximity of many of the villages in our data, shared exposure to similar access conditions is likely. If this shared exposure operates directly, i.e. through observed access conditions in neighboring locations, we may extend our model as follows:

(4) $$y_{it} = \delta \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \theta \sum_{j=1}^{N} w_{ij} \boldsymbol{x}_{jt} + u_{it}$$

where $\sum_{j=1}^{N} w_{ij} \boldsymbol{x}_{jt}$ is a spatial lag of the independent access variables, $\theta(|\theta| < 1)$ is the spatial lag parameter. This term represents the exogenous spatial dependence of growth outcomes on neighboring access (and possibly other) conditions.

If the spillover is not directly related to observable access conditions in the model, but to unobserved factors which influence growth outcomes for neighboring households in similar ways, then this is simply reflected in spatially autocorrelated errors:

(5) $$u_{it} = \rho \sum_{j=1}^{N} w_{ij} u_{jt} + \varepsilon_{it}$$

where $\sum_{j=1}^{N} w_{ij} u_{jt}$ is the spatial autocorrelation structure, and $\rho(|\rho| < 1)$ as the spatial autocorrelation parameter.

Note that these channels of dependence may coexist. In other words, the nested structure of a model consisting of (4) and (5) allows for all three channels of spatial interaction (a model sometimes called the Manski model; see Appendix A for a typology of spatial models). *A priori*, I do not discard any possible channel of spatial dependence, including its absence.

4. Empirical model

Public goods

I use a number of indicators of access to public goods, all of which are based on household reported distances to the nearest instance of each type. The types of public goods that I evaluate are presented in the table below.

Table 3: Public goods variables used in this study

Variable	Description	2000	2004	2007	2010
km advice	Km to nearest agricultural extension office	x	x	x	x
km elect	Km to nearest electricity supply	x	x	x	x
km tarmac	Km to nearest all-weather road	x	x	x	x
km feeder	Km to nearest motorable road	x	x	x	x

Endogenous initial state variables

The initial value of the variable whose growth we are measuring is likely to be endogenous. In equation (3) this variable is written as $\ln A_{it-p}$. In this paper, I deal with the endogeneity of $\ln A_{it-p}$ by using a control function approach (Wooldridge 2010b). I implement the control function as follows: let us model $\ln A_{it-p}$ as a function of a set of instruments which are not included in equation (3). We obtain the residuals from this first stage regression and add them to our estimating version of equation (3). The inclusion of the control function residuals, if they are based on valid instruments, acts to break the endogeneity of $\ln A_{it-p}$ in the primary model. Furthermore, the control function approach provides a simple test of whether or not the suspected regressor is truly endogenous: a significant coefficient estimate on the control function residual supports the endogeneity assumption (Wooldridge 2012).

The instruments I use in the control function are all time-varying household characteristics observed at time $(t - p)$. I instrument initial income levels with log per capita value of livestock and whether or not the household owned a television set.

Long-run and short-run changes in access

There is considerable enthusiasm for panel estimators that control for unobserved heterogeneity in the cross-section, as failure to account for such factors may result in severe bias (if the outcome of interest is in fact confounded by such heterogeneity). Fixed-effects (FE), first-difference (FD) and correlated-random effects (CRE) are examples of such estimators.

However, there are potential drawbacks to this strategy. When there is little variation in the covariate of interest and/or there is a high degree of measurement error – i.e. when the signal-to-noise ratio is low – the results from FD, FE or CRE estimation may be seriously compromised, with coefficient estimates severely attenuated towards zero. This is sometimes referred to as attenuation bias (Deaton 1997: p108, Baltagi 2008: p205-208, Wooldridge 2010: p365).2

This problem is usually framed as a tradeoff: differencing approaches exacerbate measurement error bias even as they eliminate heterogeneity bias. In other words, order to remove the inconsistency arising from unobserved heterogeneity, precision has been sacrificed.

Deaton (1997) notes that "a consistent but imprecise estimate can be further from the truth than an inconsistent estimator" (p108). Furthermore, "we must also be aware of misinterpreting a decrease in efficiency as a change in parameter estimates between the differenced and undifferenced equations. If the cross-section estimate shows that β is positive and significant, and if the differenced data yield an estimate that is insignificantly different from both zero and the cross-section estimate, it is not persuasive to claim that the cross-section result is an artifact of not "treating" the heterogeneity." (p108).

McKinnish (2008) shows that the measurement error problem can be extended to include cases where an indicator is measured with precision, but where this indicator is an imprecise measure of the true factor relevant to the outcome of interest. She notes that time-series variation in panels – i.e. the variation that remains after removing fixed effects – often reflects idiosyncratic changes in the independent variable that have little or no influence on the dependent variable.

[2] The case for attenuation bias is usually made for FE and FD methods, but since CRE estimates approach those of FE estimators, under the CRE assumptions, this argument applies to CRE also. See Wooldridge (2010) for comparison of CRE and FE estimator properties. Solon (1985) and Griliches and Hausman (1986) are the seminal studies of attenuation bias from measurement error in panel data.

"For example, we may know the exact value of state welfare benefits from administrative records, but not all of the variation in these benefit levels will necessarily influence behavior. In particular, we expect many outcomes to respond differently to short-term and long-term variation in conditions. This differential effect of long-term and short-term variation can generate the same bias as "true" measurement error." (p 336).

In this case, measurement error as conventionally defined is not an issue, although the resulting "measurement error problem" is the same. To paraphrase her argument in the context of this study, consider a simplified household growth model:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}$$

$$x_{it} = z_{it} + \nu_{it}$$

where is y_{it} is the growth outcome of interest, x_{it} is the access measurement for the corresponding period, and z_{it} is the sustained component of access. Even in the absence of measurement error on x_{it}, x may still not capture the underlying causal relationship of primary interest: if z is highly correlated over time and two observations of x from adjacent periods are differenced, most of the information about z will be eliminated, leaving variation which is mainly associated with the noise component ν.

We have already seen that most of the access indicators in the Kenya panel dataset vary incrementally between successive panel waves. Furthermore, there are indications that measurement error may be a problem: for example, intra-village variation in the reported distance to the nearest paved road often exceeds (by a large margin) the variation in household locations within the village. This suggests that farmers are either measuring distance to very different targets, or there is measurement error (or both). Furthermore, given the slow pace of changes, the substantive changes may best be measured over longer periods of time. In other words, even in the absence of measurement error, the sustained component of access changes is probably more important to income growth outcomes. To address this, I compare estimates from fixed effects (FE), first-differences (FD) and long-differences (LD), where the LD estimator is based on differencing levels in the first and last period of the panel. Given the four years of observations in the dataset and the single-period lag structure of the model, the LD estimator means that one year drops out. McKinnish (2008) suggests that a comparison of these methods will yield insights about the presence and magnitude of attenuation-type biases.

Spatial dependence

Spatial models pose special challenges for estimation. The cost of ignoring spatial dependence in the dependent variable (and/or a spatial lag in the independent variables) is high due to the simple fact that if one or more relevant explanatory variable are omitted from a regression equation, the estimator of the coefficients for the remaining variables is biased and inconsistent (i.e. the omitted variable problem; Wooldridge 2010). In contrast, ignoring spatial dependence in the disturbances, if present, will only cause a loss of efficiency (assuming, of course, that this non-spherical spatial error term is not an artifact resulting itself from an omitted variable). Furthermore, even when correctly specified, models with lagged dependent variables which are estimated via least squares will generally result in inconsistent parameter estimates inconsistent estimation of the spatial parameters, and inconsistent estimation of standard errors (Le Sage and Pace 2009).

There are three main approaches described in the literature for estimating models that include spatial interaction effects. The first is based on maximum likelihood (ML; see Le Sage and Pace 2009, chapter 3). The second is based on a generalized method of moments approach that uses instrumental variables to deal with the endogeneity of spatial lags (IV/GMM[3]; e.g. Kelejian and Prucha 2009). A third approach uses a Bayesian Markov Chain Monte Carlo (MCMC) approach (e.g. Le Sage and Pace 2009; chapter 5). ML estimators assume normality of errors; IV/GMM does not rely on this assumption. Both ML and IV/GMM approaches, however, assume that the $\varepsilon_{\mathrm{it}}$ are independently and identically distributed for all i and t with zero mean and variance σ^2.

Franzese and Hays (2007) compared the performance of the ML and IV/GMM estimators for panel data models with a spatially lagged dependent variable in terms of unbiasedness and efficiency. They find that the ML estimator weakly dominates the IV/GMM estimator in terms of efficiency, but that the IV/GMM estimator offers more robust estimates for some ranges of δ.

[3] The estimator used in this approach is frequently referred to as the Generalized Spatial 2-Stage Least Squares (GS2SLS) estimator.

However, Elhorst (2010) notes that they did not consider differences between spatial fixed or random effects. In this section, I describe a ML approach to estimating equation (3). I describe approaches for the spatial lag fixed effects model, but extensions to the error model are straightforward.

Fixed effects spatial lag model

To implement the FE approach, let us remove the time constant effects μ_i by demeaning the equation. This leaves us the following time-demeaned variables

(6) $y_{it}^* = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\boldsymbol{x}_{it}^* = \boldsymbol{x}_{it} - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}$

Now, stack the observations as successive cross-sections for $t = 1, \dots, T$ to obtain vectors of dimension$(NT, 1)$ for $\boldsymbol{Y}^*$ and $(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^*$, and an (NT, K) matrix for $\boldsymbol{X}^*$ of the demeaned variables.

As shown by Elhorst (2010; following Kelejian and Prucha [1998]), a consistent estimation procedure is as follows. Let $\mathbf{b}_0$ and $\mathbf{b}_1$ denote the OLS estimators of successively regressing $\boldsymbol{Y}^*$ and $(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^*$ on $\boldsymbol{X}^*$, and let $\mathbf{e}_0^*$ and $\mathbf{e}_1^*$ be the corresponding residuals. The ML estimate of δ is then obtained by maximizing the concentrated log-likelihood function

(7) $LogL = C - \frac{NT}{2} log[(\boldsymbol{e}_0^* - \delta \boldsymbol{e}_1^*)'(\boldsymbol{e}_0^* - \delta \boldsymbol{e}_1^*)] + T log|I_N - \delta \boldsymbol{W}|$

where C is a constant not depending on δ.

Third, estimators for $\boldsymbol{\beta}$ and σ^2 may be computed, using the numerical estimate of δ, as follows:

(8) $\boldsymbol{\beta} = \boldsymbol{b}_0 - \delta \boldsymbol{b}_1 = (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1} \boldsymbol{X}^{*\prime} [\boldsymbol{Y}^* - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^*]$

$\sigma^2 = \frac{1}{NT}(\boldsymbol{e}_0^* - \delta \boldsymbol{e}_1^*)'(\boldsymbol{e}_0^* - \delta \boldsymbol{e}_1^*)$

For reference, Elhorst and Freret (2007) derive the asymptotic variance matrix of these parameters.

Measuring direct and indirect impacts in spatial models

As indicated earlier, my primary analytical interest is in the partial effect of a change in access on the outcomes of interest. Unlike non-spatial linear models, where parameter estimates may be interpreted as partial effects, spatial dependence requires that we consider the dependence channels specified in the model. In particular, models containing spatial lags of the dependent variable require special interpretation of the parameters (Anselin and LeGallo, 2006; Kelejian, Tavlas and Hondronyiannis, 2006; Kim, Phipps, and Anselin, 2003; LeGallo, Ertur, and Baumont, 2003).

For a single period, we can represent the partial effect on outcome y for an individual i from a change in the rth explanatory variable x_r, in terms of an own derivative: [4]

$$\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii} \text{ where } S_r(W) = (I_n - \rho W)^{-1}(I_n \beta_r)$$

However, we might also consider the effect on observation i deriving from a change in x at observation j, i.e. $\frac{\partial y_i}{\partial x_{jr}} = S_r(W)_{ij}$. This expression, unlike in non-spatial models, is not necessarily zero.

Le Sage and Pace (2009) suggest the following summary measures of aggregate impacts:

Average Direct Impact: The impact of changes in the ith observation of x_r on y_i could be summarized by averaging over the direct impact associated with all observations i. This is somewhat analogous to the standard non-spatial linear regression coefficient interpretations that represent the average response of the dependent variable to changes in the independent variables.

Average Total Impact to an Observation: The sum across the ith row of $S_r(W) = (I_n - \rho W)^{-1}(I_n \beta_r)$ would represent the total impact on individual observation y_i resulting from changing the explanatory variable x_r by the same amount across all n observations.

[4] To see the derivation of this expression, consider the following transformation of the standard SAR model (derived following LeSage and Pace 2009):

$$y = \delta \boldsymbol{W} y + \boldsymbol{X\beta} + u$$
$$(\boldsymbol{I}_N - \delta \boldsymbol{W}) y = \boldsymbol{X\beta} + u$$
$$y = \sum_{r=1}^{k} (\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} (\boldsymbol{I}_N \beta_r) x_r + (\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1} u$$

Average Total Impact from an Observation: The sum across the jth column of $S_r(W) = (I_n - \rho W)^{-1}(I_n \beta_r)$ represents the total impact over all y_i from changing the rth explanatory variable by some amount in the jth observation.

Spatial weights matrix

The weighting matrix $\boldsymbol{W}$ may be defined in various ways. Because our dataset has geographic coordinates assigned at the household level, the weights I emphasize in this study are based on geographic distances between observations. In particular, I define neighbors as households residing within 20km from one another, with the relationship between neighbors weighted by inverse distances. This corresponds to a scale of plausible interaction and shared conditions for rural households, whereby households will be neighbors to all others within the village, but also (possibly) to households residing in nearby villages.[5]

[5] Tests of alternative specifications of the weighting matrix W wew carried out using Kelejian's extension of the J-test (Kelejian 1998, Kelejian and Piras 2009; see Elhorst 2010a, and Le Sage and Pace 2009 for alternative testing approaches based on Bayesian posterior model information). I used the J-test to evaluate several alternative weighting matrices. Based on these results, I conclude that the best fit for this application is a simple weights matrix derived from inverse distances, using geographic household coordinates, and a threshold value of about 20 kilometers. However, it is worth noting that when spatial dependence is pronounced, the specification of the weights matrix is fairly robust to misspecification (Elhorst 2010). In this case, given the strength of spatial dependence in my estimating equations, as I show in the results section, the choice of weights is fairly robust to misspecification and I do not elaborate further on the specification tests for reasons of conciseness.

5. Discussion of results

Table 2 presents estimation results of the basic (non-spatial) household growth model. The dependent variable here is log per capita household income.[6] The first two columns show results from first difference (FD), with the endogeneity of the lagged dependent variable controlled for with Control Function approach (in column 1) and IV-2SLS approach (in column 2).Similarly, the fixed effects (FE) estimation results are shown in columns 3 and 4 and the long-difference results are shown in columns 5 and 6. For each method, the differences in estimates under alternative endogeneity controls are most pronounced in the lagged outcome variable; for other covariates the choice of CF or IV does not make a large difference. Although the CF method is less robust than IV-2SLS, these results suggest that any bias arising from failing to meet the CF assumptions (which essentially are that the endogenous variable has a linear expectation which is correctly specified by the control function auxiliary regression) is minimal with respect to the independent variables of primary interest, namely the access/public goods indicators.

Access indicators are measured in log terms (after first converting to meters to preserve positive values). Thus, their coefficients may be interpreted as elasticities. generally have the expected negative sign, indicating that better proximity to infrastructure and services has beneficial impacts on income growth trajectories (the only positive estimates are not distinguishable from zero). In comparing the magnitude and significance of the FD, FE and LD coefficient estimates, we may clearly observe that magnitudes of estimated coefficients are increasing and that standard errors decrease (note: the table shows p-values, not standard errors, but the direction of the changing magnitudes of standard errors is easily deduced from the significance of the estimates).

Griliches and Hausman (1986) summarize general results for alternative estimators in the presence of measurement error: FE and LD estimators will suffer less from bias than FD; the LD estimator will suffer less from bias than the FE estimator (with the magnitude of that difference depending upon the structure of the data, number of periods, length between them, etc.). McKinnish (2008) suggests using these observations to diagnose the presence of attenuation bias

[6] All values are in nominal terms. One reason for this was to preserve a positive direction of growth (recall that there is a net decline in real incomes across the period). Since all specifications include price controls which are also in nominal terms.

resulting from mis-measurement and/or conflation of transitory and sustained changes. She notes that, under this condition, we would expect that FD estimates will be the most attenuated, as much of the signal is differenced out, relative to the noise of short term volatility. Longer differences, such as FE and especially LD, will capture proportionally more signal relative to noise, and therefore be less attenuated. This is exactly what we observe in the table. For this reason, my preferred estimator is the LD estimator.

The LD estimates using the CF indicate that electricity, extension and feeder roads all play significant roles in conditioning income growth. Of these, the magnitude of access to extension is the largest: a 50% reduction in distance to extension office (about 2.4km at the sample mean of 4.7km) is associated with nearly 9% larger rates of income growth. A 50% reduction in distance to electricity (about 1.8km at the sample mean) is associated with a 2.5% increase in income growth. At the sample mean of 0.8 km from the nearest feeder road, a reduction of 0.4 km is associated with a 1.1% increase in income growth.

The household characteristics are largely insignificant. This has to do with the fact that there is little variation over time: there is some change in household heads, which prevents these characteristics from falling out after differencing, but too little variation to contribute meaningfully as explanatory factors. Of the contemporaneous shocks, only fertilizer price is consistently important, with the expected negative sign. Maize wholesale prices approach significance with a negative sign, suggesting that maize prices affect incomes more through their role in household consumption expenditures than through output sales. Rain shocks are not significant, although this may have more to do with the greater periodicity of rainfall variation: it may not be meaningful to compare individual seasons which are separated by multiple unobserved periods.

Table 4 Determinants of income growth (log per capita KSh)

	First-difference		*Fixed effects*		*Long-difference*	
	(1)	(2)	(3)	(4)	(5)	(6)
	CF	**IV**	**CF**	**IV**	**CF**	**IV**
Lagged endogenous variable:						
log income	-1.195	-1.335	-0.935	-1.306	-0.479	-1.243
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Access:						
km electricity	-0.006	-0.009	-0.021	-0.026	-0.050	-0.051
	(0.691)	(0.549)	(0.228)	(0.316)	(0.068)*	(0.187)
km extension	0.004	0.012	-0.077	-0.074	-0.173	-0.180
	(0.944)	(0.847)	(0.262)	(0.164)	(0.081)*	(0.024)**
km tarmac	-0.048	-0.054	-0.025	-0.021	0.057	0.047
	(0.269)	(0.196)	(0.605)	(0.728)	(0.749)	(0.757)
km feeder	-0.008	-0.009	-0.011	-0.011	-0.022	-0.015
	(0.323)	(0.276)	(0.170)	(0.237)	(0.076)*	(0.313)
Household characteristics:						
farm size	-0.005	-0.002	0.003	0.011	0.020	0.035
	(0.513)	(0.787)	(0.766)	(0.195)	(0.199)	(0.002)***
female	0.078	0.066	0.037	0.039	-0.041	0.005
	(0.337)	(0.419)	(0.653)	(0.550)	(0.690)	(0.954)
age	-0.001	-0.002	-0.005	-0.006	-0.010	-0.012
	(0.910)	(0.716)	(0.257)	(0.152)	(0.069)*	(0.045)**
education	0.011	0.012	0.019	0.021	0.024	0.033
	(0.146)	(0.165)	(0.021)**	(0.044)**	(0.062)*	(0.042)**
Shocks:						
rain	0.002	0.008	0.008	0.017	0.003	-0.038
	(0.937)	(0.801)	(0.786)	(0.615)	(0.986)	(0.790)
maize price	-0.105	-0.042	-0.206	-0.268	-0.594	-0.774
	(0.648)	(0.882)	(0.447)	(0.342)	(0.191)	(0.031)**
DAP price	-0.665	-0.778	-0.881	-0.846	-0.869	-0.879
	(0.003)***	(0.007)***	(0.000)***	(0.004)***	(0.116)	(0.107)
R-squared	0.731	0.728	0.645	0.641	0.491	0.444
N	2465	2465	3698	3698	2465	2465

Note: P-values shown in parentheses are based on bootstrapped standard errors, with significance levels denoted by: * p<0.10, ** p<0.05, *** p<0.01

Table 5: LM tests of spatial structure (for LD model)

Hypothesis	LM test statistic	p-value	robust LM test statistic	p-value
H0: no spatial lag	8453	0.000***	797	0.000***
H0: no spatial error	7657	0.000***	0.04	0.842

To evaluate possible spatial structure, I implemented LM and robust LM tests for spatial panels as suggested by Anselin et al (2006) and described in Elhorst (2009). The null hypothesis is no spatial structure, against alternatives of either a spatial lag or a spatial error. (The robust LM tests are called robust because the existence of dependence of one type does not bias detection of the other type.) Results, shown in table 5, indicate that a spatial lag structure is very likely. Support for a spatial error structure is given by the regular LM test, but the robust test suggests that this is not likely.

Elhorst (2010) suggests that, after rejecting a non-spatial model structure, further specification testing should begin with the spatial Durbin model, which contains a spatial lag of the dependent variable and of (some or all) of the independent regressors. (A diagram of spatial models and their relationships is provided in Appendix A.) The spatial Durbin model (SDM) can be tested for reduction to either a spatial lag (i.e. spatial autoregressive, or SAR) model or the spatial error model (SEM), since both SAR and SEM are nested within the SDM.[7]

Table 6 shows LD estimation results for non-spatial and spatial models. Of principal interest is the efficiency gain obtained from any of the spatial specifications, relative to the non-spatial specification, particularly for distance to extension services and feeder roads. For the most part, there are only minor differences in the magnitude of coefficient estimates. The biggest change is for distance to extension advice under the spatial Durbin model, which increases substantially. However, this effect is counterbalanced by the indirect effect of distance to extension in neighboring villages: the exogenous lags are reported under the section labeled *Wx*. The positive coefficient on the lag of *km extension* is a little puzzling, as it suggests that simultaneous improvements in a village and in neighboring villages is not complementary; bear in mind, however, that these coefficients are not directly interpretable as partial effects (I calculate those

[7] This nesting is not intuitively obvious; Elhorst (2010) shows that the exogenous lag coefficients, θ, reduce to the spatial error coefficient ρ, under the following special case, which is testable: $\theta = -\rho\beta$.

below). One possible explanation is that if extension offices tend to be located in non-central places (i.e. in villages rather than in rural towns) and they are fairly well distributed, then a checkerboard pattern may result, whereby in a pair of neighboring villages, if there is an extension office in one of the pair, there is little chance of there being an office within the other. In such cases, a net positive benefit deriving from having an extension office somewhere in the vicinity would be difficult to identify from direct and spillover terms which are simultaneously measured.

The exogenous lag coefficients for roads and electricity are negative, which suggests that direct and spillover effects work in complementary ways, as we would expect. The most significant effect is for feeder roads, which indicates that denser feeder road networks have particularly strong spatial spillovers, whereby community-level improvements percolate out to neighboring areas.

The spatial autoregressive parameters (i.e. the coefficient on the spatial lag of the dependent variable) are highly significant in all models. The estimates of the spatial error term are more variable: in the SEM model, which confines spatial spillover to the error structure, the parameter estimate is highly significant. However, in the SARAR model, which allows for both a spatial autoregressive term and an autocorrelated error term, the error term is not significant. This is consistent with the LM results shown earlier. If, *a priori*, we are unwilling to accept the existence of an endogenous spatial lag, then we should accept a spatial error specification, as it offers a better fit and more efficient estimates than the non-spatial model. However, tests and specification results both indicate that the spatial lag is more likely to be the true specification. What about the endogenous lags (in the SDM)? A Wald test rejects the null that the endogenous lags are jointly equal to zero at the 95% level ($F(4,1212) = 2.96$; p-value=0.019). (Tests also reject collapsibility into the SEM or a non-spatial model.) This suggests that the spatial Durbin specification fits the model best, although the differences from the SAR and SARAR model estimates are very small.

For the non-spatial and spatial error models (columns 1 and 4) the coefficient estimates may be interpreted as partial effects. However, as discussed earlier, if we accept an endogenous and/or exogenous spatial lag terms, then partial effects require additional calculation. Table 7 shows coefficient estimates of the access variables from all models, along with the direct, indirect and total impacts for the SDM, SARAR and SAR models. Not surprisingly, the largest changes in the estimated impacts are for distance to feeder roads: the non-spatial estimate of the impact of a 50% reduction in distance is a 1% increase in the income growth rate. The estimated impact of such a change under the SDM specification, which includes both endogenous and exogenous lags, is nearly 11%. Even under the SAR model, which excludes the exogenous lags, the estimated impact is 2.4%, more than double the non-spatial impact.

The changes in impacts for the other access variables are less pronounced, but operate in the same general way, i.e. have larger magnitudes than non-spatial estimates. The impact of distance to extension services is the most significant of these other effects: the estimated elasticity increases from -0.173 (in the non-spatial model) to -0.146 in the spatial Durbin model and -0.222 in the spatial lag model.

It is likely that the role of access is different in the growth trajectories of households pursuing different livelihood strategies. For example, households who are heavily engaged in agricultural markets may benefit more from agricultural extension services. Similarly, households which earn a majority of their income from non-farm sources may benefit more from expansion of electricity and other investments that facilitate growth in the non-farm rural economy. To investigate this, I ran the basic model on two subsets of the sample: marketing specialists, defined as those who sold more than 50% of total value of output in 2000, and non-farm specialists, defined as households with more than 80% of total household income from off-farm sources in 2000. Results are shown in Table 8: model 1 is identical to the non-spatial model shown earlier; model 2 includes marketing specialists only; and model 3 is restricted to off-farm specialists.

Results indicate that income growth in market specialists has been benefited most from the expansion of agricultural extension services (relative to other access indicators in the model). The role of infrastructure in the growth of marketers' incomes is small relative to that of the total sample. This makes sense: access to new technologies through extension services is of most value to agricultural specialists, and output market participants are more likely to be able to obtain returns on investments in new technologies (as opposed to subsistence-oriented farmers). For non-farm specialists, on the other hand, the expansion of the rural electricity grid has been especially beneficial to income growth. Although we do not observe more details of the non-farm sector in local areas, the expansion of electricity has presumably enabled expansion and diversification within the sector, which should allow local labor market participants to earn returns on human capital investments, such as education, specialized training and experience.

Table 6 Comparison of non-spatial and spatial models, estimated with LD estimator

		(1)	(2)	(3)	(4)	(5)
		NS	**SDM**	**SARAR**	**SEM**	**SAR**
Lagged dependent variable:	log income	-0.479***	-0.507***	-0.499***	-0.490***	-0.501***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	CF residual	-0.518***	-0.473***	-0.485***	-0.499***	-0.480***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Access variables:	km electricity	-0.050**	-0.009	-0.038	-0.044	-0.036
		(0.046)	(0.832)	(0.154)	(0.132)	(0.155)
	km extension	-0.173*	-0.282***	-0.180***	-0.195***	-0.176***
		(0.087)	(0.000)	(0.001)	(0.001)	(0.000)
	km tarmac	0.057	0.241	0.087	0.112	0.082
		(0.746)	(0.116)	(0.405)	(0.331)	(0.419)
	km feeder	-0.022*	-0.016*	-0.020**	-0.018**	-0.021**
		(0.091)	(0.090)	(0.028)	(0.049)	(0.023)
Household variables:	farm size	0.020	0.020***	0.018***	0.018***	0.018***
		(0.234)	(0.004)	(0.006)	(0.007)	(0.006)
	female	-0.041	-0.043	-0.042	-0.055	-0.039
		(0.691)	(0.446)	(0.462)	(0.335)	(0.493)
	age	-0.010*	-0.010***	-0.009**	-0.009**	-0.009**
		(0.065)	(0.009)	(0.017)	(0.014)	(0.018)
	education	0.024*	0.028***	0.026***	0.027***	0.026***
		(0.064)	(0.006)	(0.009)	(0.007)	(0.010)
Shocks:	rain	0.003	0.009	-0.002	0.012	-0.005
		(0.986)	(0.927)	(0.980)	(0.916)	(0.955)
	maize price	-0.594	-0.379*	-0.480**	-0.521*	-0.471**
		(0.162)	(0.083)	(0.034)	(0.058)	(0.029)
	DAP price	-0.869	-1.099***	-0.827**	-0.845*	-0.825**
		(0.134)	(0.003)	(0.028)	(0.061)	(0.023)
Wx	km electricity		-0.041			
			(0.427)			
	km extension		0.180*			
			(0.099)			
	km tarmac		-0.280			
			(0.195)			
	km feeder		-0.069***			
			(0.007)			
Spatial	lambda		0.192***	0.172**		0.202***
			(0.000)	(0.012)		(0.000)
	rho			0.044	0.226***	
				(0.612)	(0.000)	
R-squared		0.491	0.385	0.382	0.390	0.380
Log likelihood		-3096	-3071	-3077	-3079	-3077
N		2465	2464	2464	2464	2464

* p<0.10, ** p<0.05, *** p<0.01

Table 7 Impacts of non-spatial and spatial models

		(1) NS	(2) SDM	(3) SARAR	(4) SEM	(5) SAR
Coefficient estimates	km electricity	-0.050**	-0.009	-0.038	-0.044	-0.036
		(0.046)	(0.832)	(0.154)	(0.132)	(0.155)
	km extension	-0.173*	-0.282***	-0.180***	-0.195***	-0.176***
		(0.087)	(0.000)	(0.001)	(0.001)	(0.000)
	km tarmac	0.057	0.241	0.087	0.112	0.082
		(0.746)	(0.116)	(0.405)	(0.331)	(0.419)
	km feeder	-0.022*	-0.016*	-0.020**	-0.018**	-0.021**
		(0.091)	(0.090)	(0.028)	(0.049)	(0.023)
Wx	km electricity		-0.041			
			(0.427)			
	km extension		0.180			
			(0.199)			
	km tarmac		-0.280			
			(0.195)			
	km feeder		-0.069***			
			(0.007)			
Spatial	rho		0.192***	0.172**		0.202***
			(0.000)	(0.012)		(0.000)
	lambda			0.044	0.226***	
				(0.612)	(0.000)	
Direct	km electricity		-0.007	-0.036		-0.035
			(0.869)	(0.205)		(0.205)
	km extension		-0.281***	-0.181***		-0.177***
			(0.000)	(0.000)		(0.000)
	km tarmac		0.267*	0.109		0.104
			(0.069)	(0.285)		(0.296)
	km feeder		-0.016	-0.018**		-0.019**
			(0.101)	(0.048)		(0.040)
Indirect	km electricity		-0.051	-0.007		-0.009
			(0.358)	(0.294)		(0.221)
	km extension		0.135	-0.038*		-0.045***
			(0.281)	(0.057)		(0.005)
	km tarmac		-0.335	0.022		0.026
			(0.147)	(0.344)		(0.301)
	km feeder		-0.092***	-0.004		-0.005*
			(0.001)	(0.142)		(0.055)
Total	km electricity		-0.058	-0.043		-0.044
			(0.120)	(0.205)		(0.205)
	km extension		-0.146	-0.218***		-0.222***
			(0.109)	(0.001)		(0.000)
	km tarmac		-0.068	0.131		0.130
			(0.711)	(0.283)		(0.294)
	km feeder		-0.108***	-0.022*		-0.024**
			(0.000)	(0.051)		(0.041)
	R-squared	0.491	0.385	0.382	0.390	0.380
	Log likelihood	-3096	-3071	-3077	-3079	-3077
	N	2465	2464	2464	2464	2464

* p<0.10, ** p<0.05, *** p<0.01

Table 8 Determinants of income growth for sub-groups in the sample

		(1) **full sample**	(2) **marketers**	(3) **non-farm workers**
Lag dependent variable:	log income	-0.485	-0.361	-0.310
		(0.000)***	(0.002)***	(0.011)**
Access variables:	km electricity	-0.050	-0.012	-0.111
		(0.072)*	(0.700)	(0.032)**
	km extension	-0.173	-0.336	0.004
		(0.080)*	(0.039)**	(0.977)
	km tarmac	0.059	0.124	0.363
		(0.742)	(0.699)	(0.161)
	km feeder	-0.015	-0.039	-0.007
		(0.492)	(0.206)	(0.824)
Household variables:	farm size	0.020	0.011	0.050
		(0.181)	(0.648)	(0.246)
	female	-0.037	0.038	-0.065
		(0.719)	(0.765)	(0.644)
	age	-0.010	-0.006	0.000
		(0.075)*	(0.432)	(0.953)
	education	0.024	0.041	-0.011
		(0.057)*	(0.020)**	(0.610)
Shocks:	rain	0.005	0.009	0.054
		(0.974)	(0.959)	(0.775)
	maize price	-0.578	-0.859	0.084
		(0.193)	(0.208)	(0.874)
	DAP price	-0.889	0.180	-0.614
		(0.102)	(0.756)	(0.372)
	R-squared	0.490	0.398	0.335
	Log likelihood	-3098	-1338	-904
	N	2465	1136	803

Notes: Marketers sold more than 50% of output in 2000. Non-farm workers earned more than 80% of total household income from off-farm sources in 2000. P-values shown in parentheses: * p<0.10, ** p<0.05, *** p<0.01

6. Conclusions

This work has shown that access to infrastructure and public goods are important conditioners of rural household income growth. I have used a variety of indicators of rural infrastructure and public services: household reported distances to the nearest tarmac roads, feeder roads, electricity and agricultural extension services. Although expansion of these services has been very gradual over the period studied, this expansion has had positive effects on rural income growth rates. In particular, the expansion of agricultural extension services, of the electrical grid and of rural feeder roads have all had positive impacts which are statistically significant and of important magnitudes.

The role of transportation infrastructure is often emphasized in rural accessibility studies, given its theoretical role in mediating the costs of market participation. Results of this analysis indicate that the expansion of feeder roads, rather than further extension of the all-weather system, may be the most important element of an investment agenda. Furthermore, the pronounced role of rural services and electricity in growth outcomes indicate that these are also important components of the rural access landscape. Policies that emphasize roads over all other infrastructure investments may fail to maximize rural investment potentials. More effective investment strategies will emphasize a range of infrastructure types as well as the provision of public services.

This study also provides evidence that rural income growth processes, as well as the role of public goods in those processes, are explicitly spatial in nature. I show evidence in support of a model of household income growth in which household growth outcomes are spatially dependent upon growth outcomes of neighboring households. Household growth has significant and relatively large positive spillovers on neighboring outcomes. This model has implications for the evaluation of the impacts of rural infrastructure investments: under a spatial lag model, the estimated impacts of such investments are considerably larger, especially for feeder roads. This magnification of impact estimates is even more pronounced when an exogenous lag of access conditions is included. This latter result suggests that it is important to consider the spillover impacts of access investments even if the process under consideration were not characterized by spatially dependent outcomes. In the case of household income growth, outcomes and the determinants of those outcomes are both characterized by spatial dependence, which results in positive synergies and larger net impacts than non-spatial models would suggest.

These results are important for a number of reasons. First, failure to correctly specify the spatial structure of household models will, at a minimum, affect the efficiency of estimators. As we have seen, this may be a particularly important issue when focusing on independent variables which vary little over time or suffer from measurement error. Furthermore, if the process of interest is characterized by endogenous spatial dependence, wherein household outcomes are not independent of their neighbors', then ignoring that structure carries a high cost: all estimates will be inconsistent. Finally, under exogenous and endogenous dependence structures, both direct and indirect effects should be incorporated into impact estimates. Failure to do so may seriously underestimate the role of model covariates, as in the case study presented here.

7. References

Alene, A. D., Manyong, V. M., Omanya, G., Mignouna, H. D., Bokanga, M., & Odhiambo, G. (2008). Smallholder market participation under transactions costs: Maize supply and fertilizer demand in Kenya. Food Policy, 33(4), 318-328.

Anselin L., A. K. Bera, R. Florax and M. J. Yoon. 1996. Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics 26(1):77-104.

Anselin L., J. Le Gallo and H. Jayet. 2006. Spatial panel econometrics. In: Matyas, L. and P. Sevestre (eds) The econometrics of panel data, fundamentals and recent developments in theory and practice, 3rd edn. Kluwer, Dordrecht, pp 901-969

Argwings-Kodhek, Gem, T.S. Jayne, Gerald Nyambane, Tom Awuor, and T. Yamano. 1998. How Can Micro-Level Household Information Make a Difference for Agricultural Policy Making? Nairobi: Egerton University/Tegemeo Institute of Agricultural Policy and Development. Available at: http://www.tegemeo.org/viewdocument.asp?ID=28

Baerenklau, K. A. 2005. Toward an Understanding of Technology Adoption: Risk, Learning, and Neighborhood Effects. Land Economics 81 (1): 1-19.

Bandiera, O., and I. Rasul. 2006. Social networks and technology adoption in northern Mozambique. The Economic Journal 116 (514): 869 - 902.

Barrett, Christopher B. 2008. Smallholder market participation: concepts and evidence from eastern and southern Africa. Food Policy, 33(4), 299-317.

Barrett, Christopher B., Mesfin Bezuneh, and Abdillahi Aboud. 2001. Income diversification, poverty traps and policy shocks in Côte d'Ivoire and Kenya. Food Policy 26.4 (2001): 367-384.

Chamberlin, Jordan and T.S. Jayne. 2009. Has Kenyan Farmers' Access to Markets and Services Improved? Panel Survey Evidence, 1997-2007. Tegemeo Working Paper WPS 37/2009. Nairobi: Egerton University/Tegemeo Institute of Agricultural Policy and Development.

Conley, T. G., and C. R. Udry. 2010. Learning about a new technology: Pineapple in Ghana. The American Economic Review 100 (1): 35-69.

Corrado, L. and B. Fingleton. 2010. Where is the economics in spatial econometrics? SIRE Discussion paper SIRE-DP-2011-02.

Deichmann, Uwe, Forhad Shilpi and Renos Vakis. 2009. Urban Proximity, Agricultural Potential and Rural Non-farm Employment: Evidence from Bangladesh. World Development Vol. 37, No. 3, pp. 645–660, 2009.

Dercon, Stefan, Daniel O. Gilligan, John Hoddinott and Tassew Woldehanna. 2009. The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages. American Journal of Agricultural Economics (2009) 91 (4): 1007-1021.

Elhorst, J. P. 2010. Applied spatial econometrics: Raising the bar. Spatial Economic Analysis 5 (1): 9-28.

Elhorst, J. P. 2012. Dynamic spatial panels: models, methods, and inferences. Journal of geographical systems, 14(1), 5-28.

Elhorst, J.P. and S. Freret. 2007. Yardstick competition among local governments: French evidence using a two-regimes spatial panel data model. Paper presented at the European and

North American RSAI Meetings, August 29-September 2, 2007, Paris, November 7-11, 2007, Savannah.

Ertur, Cem, Julie Le Gallo and Catherine Baumont. 2003. The European Regional Convergence Process, 1980-1995: Do Spatial Regimes and Spatial Dependence Matter? International Regional Science Review 2006 29: 3.

Ertur, Cem, Julie Le Gallo and Catherine Baumont. 2006. The European regional convergence process, 1980-1995: Do spatial regimes and spatial dependence matter?. International Regional Science Review, 29(1), 3-34.

Franzese, R. J. and J. C. Hays. 2007. Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. Political Analysis 15(2):140-164.

GOK. 2007. Kenya Vision 2030: Ruwaza ya Kenya 2030 (Popular version). Government of the Republic of Kenya, Nairobi. Available at: http://www.vision2030.go.ke/

Heltberg, R. and F. Tarp. 2002. Agricultural supply response and poverty in Mozambique. Food Policy 27 (2002) 103–124

Kelejian, H.H., and I.R. Prucha. 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157: 53-67.

Kelejian, H. 2008. A spatial J-test for model specification against a single or a set of nonnested alternatives. Letters in Spatial and Resources Sciences 1 (1), 3–11.

Kelejian, H. and I. Prucha. 1998. A generalized spatial two stages least square procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17 (1), 99–121.

Kelejian, H. and I. Prucha. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40 (2), 509–533.

Kelejian, H. H., and I. R. Prucha. 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157 (1): 53-67.

Kelejian, Harry H., and Gianfranco Piras. 2010. An extension of Kelejian's J-test for non-nested spatial models. Regional Science and Urban Economics 41 (2011) 281–292.

Key, Nigel, Elisabeth Sadoulet, and Alain de Janvry. 2000. Transactions Costs and Agricultural Household Supply Response. American Journal of Agricultural Economics, Vol. 82, No. 2 (May, 2000), pp. 245-259

Kim, Chong Wong, Tim T. Phipps and Luc Anselin. 2003. Measuring the benefits of air quality improvement: a spatial hedonic approach. Journal of Environmental Economics and Management, Volume 45, Issue 1, January 2003, Pages 24–39.

Langyintuo, A. S., and M. Mekuria. 2008. Assessing the influence of neighborhood effects on the adoption of improved agricultural technologies in developing agriculture. African Journal of Agricultural and Resource Economics 2 (2): 151-69.

Le Gallo, J., Ertur, C., & Baumont, C. 2003. A spatial econometric analysis of convergence across European regions. European Regional Growth, Springer-Verlag, Berlin, 99-129.

LeSage, J. P. and R.K. Pace. 2009. *Introduction to spatial econometrics*. Boca Raton, US: CRC Press Taylor & Francis Group.

Manski, C. F. 1993. Identification of endogenous social effects: The reflection problem. The review of economic studies, 60(3), 531-542.

Millo, Giovanni, and Gianfranco Piras. 2012. SPLM: Spatial Panel Data Models in R. Journal of Statistical Software, 47(1), 1-38. URL http://www.jstatsoft.org/v47/i01/.

Minten, B. and S. Kyle. 1999. The effect of distance and road quality on food collection, marketing margins, and traders' wages: evidence from the former Zaire. Journal of Development Economics, 60(2), 467-495.

Minten, B., L. Randrianarison and J. F. Swinnen. 2009. Global retail chains and poor farmers: Evidence from Madagascar. World Development, 37(11), 1728-1741.

Mutl, J. and M. Pfaffermayr. 2011. The Hausman test in a Cliff and Ord panel model. Econometrics Journal (2011), volume 14, pp. 48–76.

Pinstrup-Andersen, Per, and Satoru Shimokawa. 2006. "Rural infrastructure and agricultural development." Paper prepared for presentation at the Annual Bank Conference on Development Economics, Tokyo, Japan, May 29-30, 2006 . World Bank, Washington DC.

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

Stifel, D. and B. Minten 2008. Isolation and agricultural productivity. Agricultural Economics 39 (2008) 1–15

Stifel, David, Bart Minten, and Paul Dorosh. 2003. "Transactions costs and agricultural productivity: Implications of isolation for rural poverty in Madagascar." Transactions Costs and Agricultural Productivity: Implications of Isolation for Rural Poverty in Madagascar (February 2003). International Food Policy Research Institute, MSSD Discussion Paper 56 (2003).

Timmer, P. 1988. "The Agricultural Transformation." In H. Chenery and T.N. Srinivasan, eds., Handbook of Development Economics. Vol. 1. Amsterdam: North-Holland, pp. 275-331.

van de Walle, Dominique. 2009. Impact evaluation of rural road projects. Journal of Development Effectiveness,1:1,15 — 36

Wismadi, A., Brussel, M., Zuidgeest, M., Sutomo, H., Nugroho, L. E., & van Maarseveen, M. 2012. Effect of neighbouring village conditions and infrastructure interdependency on economic opportunity: A case study of the Yogyakarta region, Indonesia. Computers, Environment and Urban Systems.

Wooldridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second edition. The MIT Press: Cambridge, MA.

Wooldridge, J.M. 2010a. Correlated Random Effects Models with Unbalanced Panels. Keynote paper presented at the Health Econometrics Workshop, Ann Arbor, MI, October 2010.

Wooldridge, J.M. 2010b. Econometric Analysis of Cross Section and Panel Data. Second Edition. MIT Press: Cambridge, MA.

8. Appendix A: Typology of spatial models (from Elhorst 2010)