
Using Global Positioning Systems in Household Surveys for Better Economics and Better Policy

John Gibson • David McKenzie

Distance and location are the important determinants of many choices that economists study. Economists often rely on information about these variables that is self-reported by respondents in surveys, although information can sometimes be obtained from secondary sources. Self-reports are typically used for information on distance from households or community centers to roads, markets, schools, clinics, and other public services. There is growing evidence that self-reported distance is measured with error and that these errors are correlated with outcomes of interest. In contrast to self-reports, global positioning systems (GPS) can determine location within 15 m in most cases. The falling cost of GPS receivers makes it increasingly feasible for field surveys to use GPS to more accurately measure location and distance. This article reviews four ways that GPS can lead to better economics and better policy by clarifying policy externalities and spillovers, by improving the understanding of access to services, by improving the collection of household survey data, and by providing data for econometric modeling of the causal impact of policies. Several pitfalls and unresolved problems with using GPS in household surveys are also discussed. JEL codes: C81, O12, R20

Distance and location are important determinants of many choices that economists study. For example, in the von Thünen model, distance to market determines landowners' decisions about what crop is most profitable to produce. Studies of child labor market activity find distance from urban areas to be an important determinant of both schooling and work decisions (Fafchamps and Wahba 2006). In migration models, greater distance between origin and destination implies larger migration costs and reduced migration flows (Borjas 2001).

Although information on location and distance can sometimes be obtained from secondary sources, economists often rely on information that is self-reported by respondents in surveys. These self-reports are typically used for data on the distance from households or community centers to roads, markets, schools, clinics, and other public services. Evidence is presented here showing that self-reported distances and areas are measured with considerable error and are often correlated with outcomes of interest. In contrast, global positioning systems (GPS) can determine location within 15 m most of the time.

GPS locations are determined from satellites (currently 30) with precise atomic clocks. The satellites orbit about 20,000 km above the earth and send unique radio signals with a time-stamp. A GPS receiver uses the time delay between transmission and reception to calculate the distance to each satellite and to calculate the latitude and longitude of the location by triangulation. More precise calculations, including elevation, can be made if four satellites are in view (El-Rabbany 2006). Accuracy depends partly on the GPS receiver's unobscured view of the sky and partly on the quality of the receiver used to process the satellite signal. Consumer-grade GPS receivers are accurate to within 15 m 95 percent of the time,¹ with further improvements in accuracy to about 3 m achievable by using differential GPS, which augments satellite information with information from a local reference station.²

Two principal factors have dramatically increased the feasibility and usefulness of collecting GPS information in household surveys. First, on May 1, 2000, the U.S. military turned off selective availability, which had introduced random errors of up to 100 m in the civilian signal. The removal of selective availability allowed more accurate measurement, increasing the range of possible applications. Second, the cost of a basic GPS receiver has fallen to under \$100, bringing it within the budget of most household surveys.³ Coverage and precision will improve even further with the launch of the European GALILEO system, expected to be operational by 2008.⁴

The Demographic and Health Surveys (since 1997) and the Indonesia Family Life Survey are the two surveys that are well-known to economists who have used GPS to geo-reference the locations of community centers (and hence cluster of households, given the sample design). GPS has been used to locate individual households (and enterprises) in a few recent World Bank surveys, including the Rural Investment Climate Surveys in Indonesia and Sri Lanka and the Living Standard Measurement Study surveys in Albania and Tanzania. Still, the majority of household surveys in developing economies do not geo-reference communities or households, in part because of a lack of information about the benefits.

An alternative to using GPS is to use secondary sources of data on locations. In developed countries, postal addresses are widely used. For example, the United Kingdom has 2.1 million post codes for 26 million addresses. Post codes are a

very accurate proxy for household location since it is possible to get map-grid references to the nearest 100 m for most post codes.⁵ Duranton and Overman (2005) use these very detailed location data to examine the location patterns of manufacturing. However, few developing economies have detailed post codes. In developing economies, face-to-face interviewing predominates (telephone interviewing is the norm in developed countries), making it quite feasible for field teams to gather GPS data as a part of their usual survey workload.

Another source of secondary data is remote sensing, which gathers data from a sensor mounted on an aircraft or satellite. These data are typically used in studies of land cover, dealing with topics such as deforestation (Deininger and Minten 2002) and urban sprawl (Burchfield *et al.*, 2006). The unit of analysis is the pixel or picture element, which determines the size of the smallest landscape feature that can be distinguished and mapped. Typical sizes are 30×30 m or 1×1 km grids. However, these grid cells are not individual agents, and using data at this level may involve aggregating across decision-makers and may result in the ecological fallacy of drawing inferences about the behavior of individuals from analyses based on grouped or area-level data (Freedman 2001). A better matching of the spatial scale of the decision process and the scale at which measurement is carried out (Anselin 2002) may come from surveying individual decision-making agents and using GPS to link them to other spatial data.

Geographic information systems (GIS) enable such linking of different layers of data. As GIS can be considered simply a tool for combining, manipulating, and displaying spatial information captured in a variety of ways, including through GPS, a broader view sees it as a part of an emerging geographic information science (Goodchild 1992) that may enable researchers to discover new relationships for geographically referenced information. Some of the literature reviewed here relies more heavily on GIS than on GPS but still shows the types of analyses that could be facilitated by the more frequent use of GPS in household surveys.⁶

This article reviews four ways that GPS can lead to better economics and better policy by clarifying policy externalities and spillovers, by improving the understanding of access to services, by improving the collection of household survey data, and by providing data for econometric modeling of the causal impact of policies. The article also discusses some pitfalls, unresolved problems, and ongoing research.

Four Ways Using GPS can Lead to Better Economics and Better Policy

The use of GPS can lead to better economics and better policy in at least four ways.

Using GPS can Clarify Policy Externalities and Spillovers

The spatial proximity of one household to another may be of direct interest, particularly for understanding interactions between households, the role of social networks, and the potential spillovers from policies that treat some households and not others.

One example of interactions between households that researchers might want to study is the possibility that households learn from each other's actions. Conley and Udry (2005) study learning in the context of the decision to cultivate pineapple in Ghana and to determine how much fertilizer to apply. They note that the classic identification problem is that a farmer's greater likelihood of adopting a new technology soon after neighbors have done so might be a consequence of some unobserved variable that is spatially correlated, such as soil type, pests, or topographic features, rather than a result of genuine learning. They use GPS to define the geographic neighbors of a given plot as those within 1 km of the center of the plot, and they collect data on whom farmers talk to (informational neighbors). Controlling for the deviation of a farmer's input from that of his geographic neighbors, Conley and Udry could then identify learning through the impact of informational neighbors' choices. They also find evidence of positive spatial correlation in unobserved shocks to the productivity of fertilizer, highlighting the importance of controlling for geographic effects when examining learning.

McKenzie, Gibson, and Stillman (2007) also use GPS to study learning from neighbors, examining how emigrants' negative employment experiences affect the expectations of would-be emigrants. The would-be emigrants were all unsuccessful in a random ballot in Tonga that offers an opportunity for ballot winners who obtain employment to move to New Zealand. When subsequently interviewed about their employment and income expectations had they moved to New Zealand, the would-be emigrants greatly understated employment rates and incomes compared with actual outcomes for emigrants. One explanatory factor is that many ballot winners who did move to New Zealand found that their initial job offers were no longer available, and news of this negative outcome appears to have flowed back to the would-be emigrants in Tonga. Specifically, if all ballot-winning emigrants within a 6 km circle (based on GPS measurements) of the ballot losers did not take up their initial job offers in New Zealand, the employment expectations of the ballot losers were 19.6 percentage points lower.

The standard approach for evaluating the impact of a policy is to compare outcomes for those subject to the policy with outcomes for a comparable group not subject to that policy. However, as Miguel and Kremer (2004) point out, this can give misleading results when there are externalities. They investigate the impact of a deworming treatment in schools in Kenya. Using GPS distances at the school level, they control for the number of primary school pupils within a certain

distance of the school and use the number of treated pupils within this distance to measure health spillovers. They find that naïve estimates that fail to take externalities into account would underestimate the program treatment effects, leading to the mistaken conclusion that deworming is not cost-effective. Such an approach could be extended by obtaining GPS locations of the residence of each child, which could then be used to construct a child-specific measure of exposure to treated and nontreated children. This would provide more variation in the extent of spillover, which could be used to examine the heterogeneity in treatment effects. A recent example of research examining spillovers at the individual level is provided by de Mel, McKenzie, and Woodruff (2007). They conduct a randomized experiment in which grants of capital stock are given to a randomly selected sample of Sri Lankan firms. They then estimate the impact of this treatment, controlling for the number of treated firms within 500 m and 1 and 5 km of each surveyed firm.

Information on the Spatial Distribution of Population and Services is Essential to Improving Understanding of Access to Services

One of the most common uses of GPS information in developing economies has been to measure access to infrastructure and social services, particularly health care. For example, Perry and Gessler (2000) use GPS to measure access from communities to primary health-care facilities in Andean Bolivia and use the results to propose an alternative model of health distribution in the study area.

In addition to providing purely descriptive measures of access, GPS data on distance and travel time can identify barriers to the use of services. In examining the influence of accessibility to family planning on choice of contraceptive device, Entwisle *et al.* (1997) demonstrate two advantages of GPS over survey-based measures of access. First, surveys often collect data on family planning accessibility only for certain political or administrative boundaries, establishing, for example, whether there is a facility in the village. However, facilities in neighboring administrative units may be closer. Using geo-referenced data allows more flexible specification of boundaries unconstrained by administrative definitions. Second, respondents in surveys often report travel times to health facilities in 30-minute increments, whereas GIS gives no time clumping, allowing better specification of functional form.

Gibson *et al.* (forthcoming) examine the use of different channels for receiving remittances in Tonga. Transaction costs on money transfers are much higher for Western Union than for withdrawals from an automated teller machine (ATM). There are eight ATMs on the main island of Tongatapu and five Western Union branches, so a simple measure of branches and ATMs per capita would suggest that ATMs are more accessible. However, when GPS coordinates of the ATMs and

Western Union branches are combined with village-level population information from the census and a digitalized map of the road network to measure the share of the population within different travel distances of the two financial channels, the Western Union branches are shown to be more dispersed and to offer better access. The branches cover 97 percent of the population within a 10 km travel distance, whereas the ATMs cover only 77 percent of the population within this distance (figure 1). This combination of GPS data collection and mapping software can be particularly effective in illustrating access in a form that policy-makers can readily use, as shown in figure 1.

Recent health applications combine distance with measures of health infrastructure quality. Hong, Montana, and Mishra (2006) use the 2003 Demographic and Health Surveys in Egypt to look at the relationship between the use of IUD contraceptives and the quality of family planning services. They link each household to the nearest family planning clinic within 10 km and then use detailed survey data to measure the quality of the facility. Rosero-Bixby (2004) uses GPS data on census tracts and locations of health facilities in Costa Rica to assess improvements in access following health reforms, measuring access through a combination of distance and services provided by the facility. He notes that households may not necessarily use the nearest facility, particularly if it is of low quality, and that using GIS enables calculating such measures as the density of services that meet a standard quality within a specified radius.

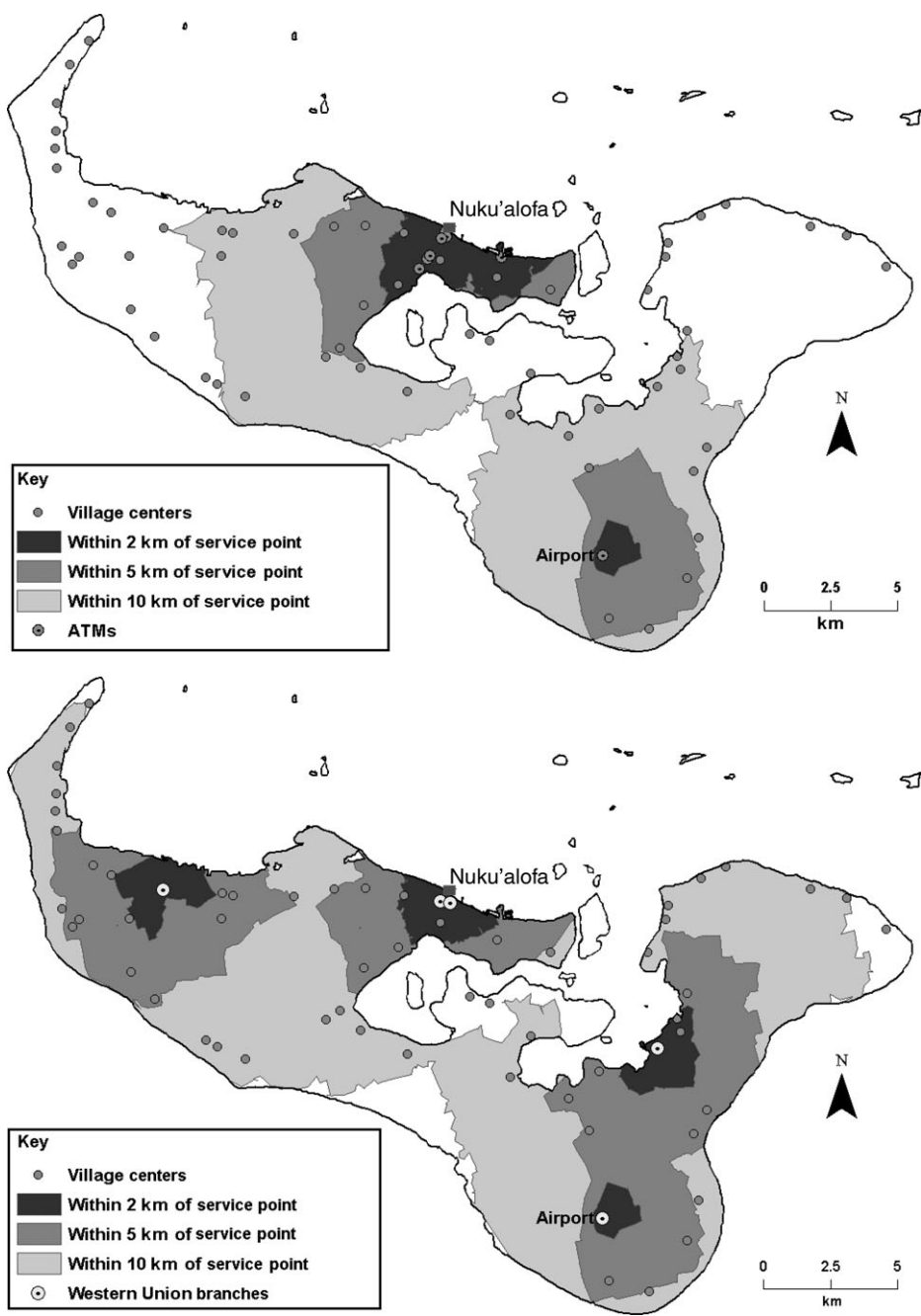
A limitation of these health studies is that they measure distance only at the community level, whereas households on opposite sides of a village or town may each be closer to a different facility, either of which may be in another community. A second limitation is that distance to health facilities could be correlated with a host of other unmeasured factors, such as poverty, disease environment, and other infrastructure, which could also affect health decisions.

Using GPS Can Improve the Collection of Household Survey Data

GPS is also being used to improve the quality and cost-effectiveness of household survey data. GPS is being used in several phases of data collection, from the development of a sample frame to quality control and follow-up surveys. More accurate and cost-effective surveying enables researchers to carry out better analysis and provide better evidence-based advice to policy-makers.

Household surveys require an accurate sample frame. The most common approach involves using a recent census to select enumeration areas. However, censuses may become outdated during periods of rapid urbanization and are of little use in drawing samples in post-conflict countries that have not had a census for decades. Afghanistan plans to complete a census in 2007, its first since 1979,⁷ and Lalasz (2006) reports that 15 countries have not taken a census

Figure 1. Service Areas for ATMs and Western Union Branches for Tongatapu, Tonga



Source: Gibson *et al.*, (forthcoming), figure 4.

since 1990. The traditional solution is to do area sampling, in which enumerators list all households in a well-defined block, such as a village or an urban area bounded by certain streets. These blocks are determined largely by convenience in defining and locating them and can be expensive to enumerate.

Landry and Shen (2005) show how GPS can be used to do area-based sampling quickly and cheaply, since enumeration areas can be defined by spatial coordinates and made arbitrarily small. Landry and Shen consider the problem of surveying in China, where household registration lists are widely used as sample frames. Widespread migration from rural areas means that many households are unlikely to be found on these registration lists. They use GPS to survey randomly chosen $54 \times 54 \text{ m}^2$ (approximately one square second) spatial blocks and find that 45 percent of the households reached were not on household registration lists.

A potential problem with this approach is that the sample size is not known until after data collection, since the number of households within a spatial block will not be known in advance. Landry and Shen use the existing population data to create a rough population model of Beijing. Since the number of dwellings within their spatial units was four times as large as they had budgeted for, they administered their questionnaire to just a quarter of the units.

Aerial photography is likely to alleviate such problems in the future. For example, Cowen and Jensen (1998) extract information on individual dwelling units in a 32-block census area in South Carolina from aircraft multispectral data. They find that the dwelling unit data derived from remote sensing had a correlation of 0.91 with data derived from the census. As satellite imagery continues to improve in resolution and fall in price, it appears likely that the combination of remote sensing and spatial sampling will become the standard for constructing sample frames in situations where reliable census or registrar data are not available.

Kumar (2007) describes combining remote sensing and GPS for drawing samples in a survey of 1,600 households spread across different air pollution zones in Delhi and India. He partitions the study area into strata according to air pollution levels (obtained by remote sensing) and proximity to main point sources of air pollution. Random points are then simulated using GIS techniques (weighting by size of residential area in each strata), and GPS is used to navigate to the households located at each selected point, and the households are asked to participate in the survey. This method of creating a frame and drawing a sample should be more efficient than simply imposing a regular grid on the study area, since air pollution is irregularly distributed over space.

Visualization of the locations at which sampling has occurred can provide a useful form of quality control to ensure that interviewers conduct surveys where they are supposed to and to check whether any dwellings are inadvertently missed. In 2004, Timor-Leste became the first country to use GPS units in its

census to record the locations of all households. USAID/Timor-Leste (2004) reports that survey managers checked the GPS points visited by the census teams against detailed aerial photographs to detect any areas missed in the enumeration. Census undercounting matters for a variety of policy purposes, including the allocation of federal money and political representation. Undercounting can be particularly high in developing economies. Lalasz (2006) reports that the 1991 Census is thought to have undercounted Nigeria's population (officially put at 89 million) by perhaps 20 million people. The use of GPS can show where undercounting has occurred and help survey managers to reduce it.

GPS can also reduce the cost and time to relocate dwellings for follow-up surveys. Follow-ups may be needed to allow field managers to check for errors by enumerators and to collect panel data. In many developing economies, the lack of street addresses, especially in densely populated urban areas, and changes in administrative boundaries between waves of a survey can make identifying the same dwelling or household time-consuming. A pilot study by Dwolatsky *et al.* (2006) tracing patients who left a tuberculosis control program in South Africa shows the potential for using GPS to relocate dwellings. They find that it takes 20–50 percent less time to locate a home using a customized personal digital assistant linked to GPS than using residential addresses. The main limitation is that this was a small pilot study of only 20 houses, so further experiments are needed to confirm the promising results.

When panel surveys attempt the more difficult task of tracking individuals rather than dwellings, GPS can be very useful for tracking people who had previously been co-residents. For example, the 2004 Kagera Health and Development Survey in Tanzania used GPS to record the locations of 2,700 households containing members who had been in the baseline sample of 900 households first interviewed in 1991–1994 (Beegle, De Weerd, and Dercon 2006). Measures such as how far people have moved from their baseline village center or from households with members who had been co-residents in the baseline surveys can be related to various socioeconomic characteristics.

Finally, collecting GPS data for households enables linking the household data set to other surveys and data sets. There is considerable option value in doing this, since many potential uses of the data will not be known at the time the survey data are collected.

GPS Can Be Used to Provide Data for Econometric Modeling of the Causal Impacts of Policies

Most empirical work in development economics aims to identify the effect of a particular variable of interest, X , on a particular outcome, Y . A standard concern is that there are other variables that are correlated with X and that also affect Y .

Failure to control for these variables gives biased results. One of the most basic uses of GPS is to allow researchers to better control for geographic and locational characteristics in their regressions. Such characteristics are increasingly found to be relevant to outcomes of interest for development economists and practitioners. For example, Deininger and Minten (2002) obtain data from a GIS on soil quality, rainfall, elevation, slope, and other geographic features and find that higher levels of poverty are statistically associated with greater likelihoods of deforestation. However, when they re-estimate the model without GIS data, they find poverty to be associated with lower levels of deforestation. The problem is that poor people live on worse-quality land, which limits the benefits (such as agricultural income) from the deforested land, so controlling and not controlling for land quality give opposite results.

Propensity-score matching has become a popular tool for investigating policy impacts (see Ravallion, forthcoming, for a recent review). The idea is to compare individuals subject to a policy with similar individuals not subject to it. Typical variables used for matching are household socioeconomic characteristics and an often crude set of community-level variables. Brady and Hui (2006) argue that GIS can be used to more explicitly include geography in matching. They present three arguments for doing so: much individual data that would be useful for matching is unmeasured and place can serve as a proxy for unmeasured individual characteristics; nearby places are more likely to share community characteristics, such as culture, trust, and government ability; and geographic matching can be visually persuasive when sudden changes in outcomes occur across administrative borders and a program is in one community and not in its neighbor.

Nevertheless, Brady and Hui acknowledge that in some cases the most comparable places in terms of cultural or socioeconomic characteristics may not be geographically close. Therefore, matching must be based on more than geography. Although the literature on the U.S. labor market emphasizes the importance of comparing participants in training programs with nonparticipants from the same local labor markets (Heckman, Ichimura, and Todd 1997), the literature has generally not explicitly included geographic proximity as a criterion when matching individuals in different communities. As more surveys include GPS coordinates, this will become increasingly possible.

The two examples above highlight the ability of GPS to help researchers better control for (potentially) observable characteristics. More controversial is the use of distance or other geographic variables as instruments in instrumental variables estimation.⁸ Oster (2006) uses distance to the Democratic Republic of Congo as an instrument for HIV prevalence when examining the response of sexual behavior to HIV prevalence rates in Africa. McKenzie, Gibson, and Stillman (2006) use GPS-measured distance from a household in Tonga to the New Zealand immigration office in Tonga where application forms must be deposited as an

instrument for migration, when looking at the effect on income of migration to New Zealand. Olken (2006) uses GIS data on community locations and geography to study the impact of television and radio on social capital in Indonesian villages. He argues that geography leads to differences in signal strength in different villages because of mountains located between some villages and the transmission towers, but after controlling for other variables, he concludes that this geography has no independent effect on social capital.

However, using distance as an instrument is subject to potential problems. One is that distance to borders and major cities is also likely to determine access to markets, schools, health facilities, and other infrastructure, which can influence economic behavior. Also, people, villages, and cities are not randomly allocated in space. As a result, distances usually incorporate the results of behavioral choices, some of which may affect outcomes. The standard response to such concerns is to include as many other geographic controls as possible. For example, Olken (2006) controls for elevation, district fixed effects, and distance and travel time to major cities and uses a physical model of radio transmission that predicts how signal strength should vary with topography. But as with all instrumental variables, even after including such controls, a case needs to be made for why the exclusion restriction should hold—why should one believe that unobserved geographic features are not also influencing the outcomes of interest?

A second potential concern with the use of distance as an instrument arises when the response of interest varies across individuals. Even if the exclusion restriction holds, the instrumental variables estimator will identify only the local average treatment effect (LATE) in this case. As Heckman (1997, p. 451) notes about the use of distance to the nearest school as an instrument for schooling, “LATE estimates the effect of variation in distance on the earnings gain of persons who are induced to change their schooling status as a consequence of commuting costs that vary within a specified range.” Whether estimation of such a parameter interests policy-makers is a matter of some doubt.

There is less concern about this issue when most individuals respond to distance in a similar manner. McKenzie, Gibson, and Stillman (2006) find that 98 percent of individuals who did not apply for the migration lottery in Tonga gave lack of information as the main reason. Living closer to the consulate office results in better information for most individuals; so distance might be expected to change migration status for most individuals in the sample. Indeed, they find that using distance as an instrument gives an estimated income gain from migration of within 2 percent of that obtained from the experimental estimator provided by a migration lottery. Thus, this is an example where using distance as an instrument provides reliable results.⁹

Similarly, it may be that shocks to local environments, as captured by remote sensing data in two time periods, provide a more defensible identification

strategy. For example, households can be linked to areas of flooding, earthquakes, tsunamis, and other such shocks. One practical constraint is that converting the satellite images to usable data is costly and time-consuming with the current manual techniques.

A final use of GPS is in spatial econometric models. Many unobserved variables, such as climate and soil in agricultural settings, are spatially correlated, leading to spatial autocorrelation in the error term of regression equations.¹⁰ Failure to account for this structure in the error terms will lead to the use of incorrect standard errors for inference, possibly leading to the conclusion that a policy has a significant effect when it does not, or vice versa. Distances between observations obtained through GPS can be used to account for spatial autocorrelation in the error term of the regression equation. Case (1991) and Conley (1999) provide procedures for doing this.

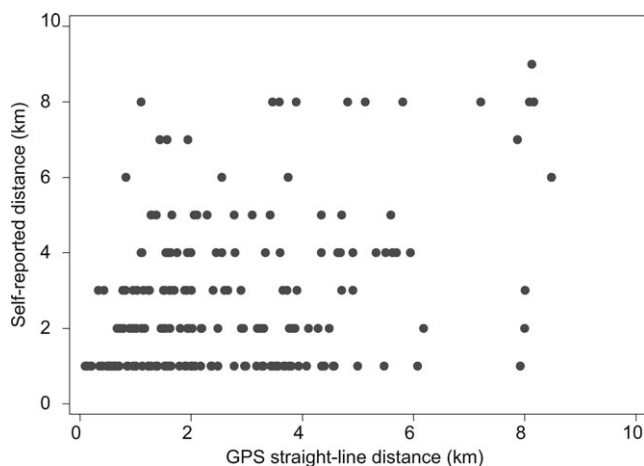
How Much Improvement Does GPS Give Over Self-Reports, and Is a Straight Line Good Enough?

A natural question that arises when distances from households and communities to other households, communities, or infrastructure need to be known is whether GPS should be used to measure these or whether self-reports in household surveys are accurate enough. A follow-up question is whether a simple straight-line distance (as the crow flies) is sufficient or whether the GPS coordinates should be integrated with GIS information on transport routes and topography to measure travel distances and travel times.

The consequences of mismeasuring distance depend on how distance is going to be used and on how badly it is mismeasured. If measurement errors are classical (independent across individuals and over time and uncorrelated with individual characteristics), then when distance is used as a regressor, as in studies of access, the effect will be to understate the impact of distance (attenuation bias). Using distance as an instrument with classical measurement error will lower the power of the instrument, potentially giving rise to weak-instrument concerns, but it will still result in consistent estimates.

However, there are strong reasons to believe that measurement errors are not random. Entwisle *et al.* (1984) note as an example that if people are asked to report travel times to a health provider, those who currently use that health resource will have more accurate knowledge than those who do not. Thus, the measurement error is likely to be correlated with usage patterns, a problem for investigating the impact of distance on use. Indeed, Andrabi *et al.* (2007) report in their survey in Punjab and Pakistan that many households do not even know

Figure 2. The Low Correlation between Reported Distance to the Tax Office in Urban Bolivia and the GPS-Measured Straight-Line Distance



Source: McKenzie and Seynabou Sakho 2007.

the name of the nearest school, let alone its location. If the measurement error is correlated with socioeconomic variables that also affect the outcome of interest, then the mismeasured distance will also give inconsistent instrumental variable estimates.

Few studies systematically compare self-reports of distance and travel times to GPS measurements, particularly in developing economies. For this study, a recent World Bank survey of owners of microenterprises and small enterprises in Bolivia was used to make the first known comparison of self-reports of physical distance to GPS-measured straight-line distances.¹¹ Firm owners are required to register at the local branch of the national tax system, but only 30 percent of the firms in the sample had registered. Firm owners were asked the distance in kilometers to the nearest tax office, which in heavily urban areas could be compared with the straight-line distance taken from GPS coordinates of the firm and tax office. More than half the firms said that they did not know the distance, with lack of a response strongly correlated with whether a firm had registered: 68 percent of unregistered firms answered that they did not know, whereas only 25 percent of registered firms did not know.

Figure 2 shows a scatterplot between the reported and measured distances for firms within 10 km of the tax office, conditional on firms also reporting the distance as 10 km or less. Pearson's correlation between reported and actual distance is still only 0.39, and the Spearman's is 0.31. The degree of measurement error conditional on giving a self-report is not significantly related to whether the

firm is registered or to the age, gender, marital status, or education of the firm owner. However, men, more educated individuals, and registered firm owners are more likely to report a distance.

Self-reports of time can also contain systematic measurement error. Escobal and Laszlo (2005) compare the self-reported time for agricultural producers in Peru to get to the nearest population center with the true travel time. The true time is measured by having surveyors walk with a random sample of respondents and time their journeys, following the same route and pace as the respondent and using GPS to measure latitude, longitude, altitude, and distance. GIS is then used to account for terrain and compute travel time for those in the survey who were not accompanied by the surveyor. They find that respondents consistently underreport the time to reach the center. For example, among coffee producers in the Selva, mean self-reported time was 6.7 minutes, compared with a mean true time of 13.0 minutes. The correlation between self-reported time and true travel time is only 0.28 for coffee producers, 0.29 for potato farmers and -0.08 for rice farmers in their sample. Furthermore, Escobal and Laszlo find that measurement errors are correlated with socioeconomic variables. Not surprisingly, individuals who own a watch give more accurate reports of travel times. They also find a negative correlation between measurement error and education so that more educated people have less measurement error.

These results strongly suggest that self-reported distances will be misleading, with measurement errors correlated with outcomes of interest. GPS coordinates can be used to give more accurate measurement. The simplest approach is to calculate the straight-line distance between points. This has the advantage of computational simplicity and does not require additional geo-coded information on transport networks or topography. Alternatively, users can combine GPS point coordinates with information on the location of transportation routes and perhaps with information on road quality and topography to measure exact travel distances and predict travel times.

The correlation is likely to be much higher between GPS straight-line travel distances and exact travel distances than between self-reported distance and GPS distances. For example, McKenzie, Gibson, and Stillman (2006) compute the distance in meters from each household in their sample to the New Zealand immigration office in Nuku'alofa. The Pearson correlation between the straight-line and exact travel distance based on road networks is 0.82, and the Spearman correlation is 0.78. They find the absolute percentage measurement error to be correlated with whether an individual migrates and with income from work in Tonga. The measurement error is greater for individuals located in more remote areas (on the other side of a lagoon), and this remoteness in turn is correlated with economic behavior. Nevertheless, the size of this error is small enough in application that there is no difference in the instrument variable estimates

between straight-line distances and road distances. The income gains from migration are estimated to be \$280 (standard error \$122) using the straight-line distance as an instrument for migration and \$281 (standard error \$101) using the road distance.

More generally, the difference between straight-line and road distance measures will be larger when geographic features such as mountains, lakes, lagoons, and rivers lie between a household or village and the location of interest. Thus, the difference between straight-line and road distances will be correlated with the remoteness of a location, which in turn is likely to affect many variables of interest. As a result, road distances are preferred to straight-line distances where possible.

Furthermore, since a curve between two points is always longer than a line, travel distances will be longer than straight-line distances. As a result, measures of access based on straight-line distances will overestimate the proportion of the population that is covered by a given service. This is demonstrated by Noor *et al.*, (2006), who examine health coverage in Kenya, where the government has set a target of ensuring that no one lives more than one hour away from effective health services by 2010. Estimates based on straight-line distances—the standard approach to coverage—indicate that 82 percent of the population is within one hour of government health services. Adjusting for the travel network drops coverage rates to 63–68 percent. This would mean that 19 million people rather than 25 million are currently covered.

Pitfalls and Unresolved Problems

Several problems need to be avoided or overcome in using GPS in household surveys.

Interviewer Error

Although taking GPS readings is straightforward, it requires good training to avoid creating another source of measurement error. One method of improving accuracy in readings is to have interviewers take multiple readings for the same location and then use their average. Some GPS receivers have a built-in function for doing this. The guidelines for collecting GPS data in Demographic and Health Surveys recommend taking multiple readings within a five-minute period and averaging them (Montana and Spencer 2004). The fieldguides of Spencer *et al.*, (2003) and Montana and Spencer (2004) provide a good starting point for researchers planning to use GPS in household surveys. The guides recommend at least 60 minutes of outdoor hands-on training with the GPS units.

Datum and Coordinate Projection Problems

A spheroid approximation of the shape of the earth is used to solve geodetic problems for point location, and this surface, its origin, and the orientation of its latitude and longitude lines make up a “geodetic datum.” GPS receivers typically use the World Geodetic System 1984 (WGS84) datum, which is a geocentric datum (its origin coincides with the center of the Earth) designed for making worldwide measurements. However, there are hundreds of other datums that may use a different center, spheroid, or reference point on the Earth’s surface in order to be locally more accurate. Coordinate values resulting from interpreting latitude, longitude, and height values based on one datum as though they were based on another can cause position errors of up to 1 km (Ramachandran 2000), although the discrepancy will typically be less. Bennett (2006) gives the example of walking around Tiananmen Square in Beijing with a GPS receiver and then importing the measurements into Google Earth, which shows a path offset by approximately 14 m because of the difference between WGS84 datum and the datum used by Google Earth.

Another common source of error comes from mixing geographic and projected coordinate systems. Projected coordinates overcome the problem that latitude and longitude are not constant units, so Cartesian geometry cannot be used to measure either distances or areas when work is being done with latitude and longitude coordinates.¹² Projected coordinate systems convert latitude and longitude coordinates from the Earth’s three-dimensional surface onto a two-dimensional map. Consequently, if location data from a GPS (which are for a three-dimensional surface, and so use a geographic coordinate system) are combined with two-dimensional map data (in a projected coordinate system), lining up the various data layers requires conversion. This is easily done in a GIS. Building up different layers of data for households, villages, and features of interest such as roads, coastlines, rivers, and public services adds value to the information in each layer. But it is surprisingly easy to have unmatched coordinate systems, because metadata, which should tell users about the coordinate system and datum used, are not always included with existing geographic data. For example, one of the authors digitized a road around the edge of the island of Tongatapu in Tonga by driving on it with a GPS receiver turned on. He obtained an existing base map of the coastline, but no metadata were available to show the projection used. He chose the most likely coordinate system for the base map, but the road and the coastline were misaligned; on one side of the island, the road that the author had driven on appeared to be in the ocean.

A more general issue is that in many developing economies there is not much off-the-shelf geographic information available at the resolution needed to merge it with village- or community-level data. Information is more often available at

a coarser scale, making it difficult to link household locations to local geographic features. Even when information is available at high resolutions, it may not match up with the household survey because of differences in coordinate systems. It is therefore important for countries to have a spatial data infrastructure that coordinates collection activities so that different geographic data sets can be matched.¹³

Road Network Problems

Practical difficulties increase when constructing a road network for measuring distances or travel times. The algorithm used in a GIS for calculating the shortest distance requires good alignment between the lines and junctions of the digitized road network. For example, if digitized road segments at a junction do not line up, the algorithm will back-track and seek another path. These problems may be especially apparent when roads have been digitized for another reason, such as cartographic display, so once again metadata about the origin and purpose of geographic layers used in conjunction with GPS data are very useful. It is also sensible to budget considerable research assistance time to clean a road data set since misalignment problems are common. For example, the road network data set underlying the service areas for money transfer facilities in figure 1 took more than a week to clean. The digitized roads had been obtained from an earlier cartographic project, so even though it looked like a digital road network, it was more like a picture of a network and considerable work was required to convert it into the continuous lines and junctions needed for calculating travel distance and time.

One way to reduce the effort required to obtain a usable roads network data set is to digitize only the main roads and then to assume a network of feeder roads, which will automatically have nicely aligned junctions. This approach is used by Staal *et al.* (2002, 2003) who study market access and its effect on market participation and technology adoption for smallholder Kenyan dairy farms. To build a road network linking their sample of farms to Nairobi and other urban areas, they use topographic maps to digitize three classes of roads: all weather, bound surface; all weather, loose surface; and dry weather only roads. Since this classification leaves many of the surveyed farms off the actual road network, they add a 4 km grid of assumed feeder roads to fill in the areas between existing roads. It is not clear how much error is introduced by using this combination of actual and assumed roads.

Confidentiality Issues

The accuracy of GPS in measuring household and community locations also poses a challenge for maintaining the privacy of survey respondents. VanWey *et al.*

(2005) discuss the potential conflicts between the ethical need to ensure the confidentiality of information collected about human research subjects and the desire to link the characteristics and actions of individuals or households with a particular location. Uncertainty about how to proceed may mean that spatially explicit data are underutilized, undermining the role of data sharing and data preservation in advancing science or that researchers inadvertently disclose information that can identify survey respondents. These conflicts affect not only the original producers of data but also any data archivists charged with maintaining the database and providing it to other researchers while honoring the commitments to confidentiality made when the data were collected.

Moreover, this conflict between the confidentiality and the usability of GPS data is not limited to the sharing of data. It also affects the reporting and displaying of results based on geo-referenced data. For example, to show the confidentiality challenges posed by mapping point data, Curtis, Mills, and Leitner (2006) reverse address match back to individual residences (re-engineer), a newspaper map showing the locations in New Orleans where deaths occurred during Hurricane Katrina. Each location mark in the newspaper map covered approximately one and a half city blocks, and the map showed no roads and had few reference points. Nevertheless, over 30 percent of the re-engineered locations fell within 25 m of the actual residence where a death occurred. (Validation for the actual locations where the deaths occurred came from the search and rescue markings on dwellings, which were recorded during a field survey.) In several cases, the re-engineered location matched the field-verified residence where the death occurred. The authors scatter a series of random coordinates throughout the study area to show that chance alone would not give the same level of discovery. They conclude that “[t]he fact that many of the re-engineered coordinates could be used to identify an actual address, or an address within the immediate vicinity, should sound a note of caution for academics publishing maps displaying human cases as points” (p. 53).

Typical approaches for maintaining confidentiality are to allow access to GPS data only to approved researchers who promise not to identify respondents, to convert point data to surfaces or distances to avoid revealing individual locations, to aggregate and report data only for larger areas, and to use a geographic masking procedure that adds stochastic or deterministic noise to the geographic coordinates for sampled households and communities. Many surveys use a combination of these four approaches. Human-subject panels (which review the benefits and risks to subjects of research projects) can play an important role in protecting confidentiality, but researchers need to be aware of the costs of the different approaches.

Approval procedures are generally stricter for obtaining geo-referenced data than for ordinary household survey data. For example, researchers have to provide

additional justifications and commitments before they can obtain (masked) GPS data on community locations in Demographic and Health Surveys. Stricter yet are the rules for access to data enclaves. Data enclaves are usually located within a survey organization, and accredited researchers are required to come to the enclave to run their analysis. All output is checked for disclosure risk before release, and there are typical restrictions on the linking of data sets and on the identity and location of individual respondents. There are often entry fees to use data enclaves, and the limited number of enclave locations may act as a barrier to research. Remote (or virtual) data enclaves are being explored in some countries to overcome these barriers. Rules on minimum cell sizes or the size of spatial units to be mapped can ensure that no information identifying individuals is provided to researchers. Cromley, Cromley, and Ye (2004) describe a system in which user queries yield results only if the cell contains at least six records and the minimum population in the smallest mapping unit is about 1,000.

Point data on household locations can be converted to a continuous surface to represent the spatial distribution of characteristics or outcomes without identifying respondents. For example, geographers have a variety of spatial interpolation methods, such as spatial variants of the kernel density estimators increasingly used by economists (Bithell 1990). Alternatively, point location coordinates can be replaced by distances to various features of interest in any public-release data set. However, these methods are not very flexible and are likely to limit future research use of the data. Surfaces do not provide the micro-data needed for studies that seek to measure causal impacts. The features of interest for which distances are reported are likely to vary from one study to another, and as distances to more features are included, the possibility of using triangulation to identify household locations increases. Moreover, distances are often needed for more than features of interest. For example, knowing the position of households relative to each other is important in studies of learning from neighbors (Conley and Udry 2005) and improves the modeling of spatial autocorrelation (Gibson and Olivia 2007).

Aggregating groups of observations into larger reporting units is widely used for maintaining the confidentiality of survey and census records. With GPS coordinates, the locations of individual households could be aggregated into larger areal units such as census blocks or census tracts so that what is reported is an administrative code for the larger area or a polygon showing the boundaries of the area where the household is located. These techniques can also be applied to the visual display of data by using dot points on a map that are sufficiently large to prevent disclosure risk. For example, VanWey *et al.* (2005) show how the size of the required buffer around the locations of sampled schools in a U.S. survey would need to vary from 6 km in a city to more than 50 km in the countryside to hold disclosure risk to 5 percent when mapping the sample points. But aggregation

seriously degrades the analyses that can be conducted, since so much detailed spatial information is lost. For example, Fefferman, O'Neil, and Naumova (2005) provide an example of how areal aggregation can yield little benefit of additional privacy and large costs in impaired ability of statistical tools to analyze patterns of disease prevalence.

Geographic masking methods work by modifying the geographic coordinates linked to each household or community. Either random perturbations or affine transformations can be used. If the relationship of sample points to each other is important, whereas the relationship to another data layer (say, a base map or a road network) is not, then simply moving all points by a given distance and direction or rotating them about a chosen point may preserve confidentiality and not greatly degrade the usability of the data. In a variant of this approach, the Rural Investment Climate Survey in Indonesia rotates points in each sample cluster. Normally, however, point locations obtained with a GPS are merged onto other data layers so these affine transformations will introduce error that reduces the usability of the GPS data.

Another option is to introduce random perturbation around the original point with the radius of the perturbation circle chosen by the data custodian, possibly weighting the size of the circle by population density at each point to take into account the effect of population density on the risk of disclosure (Kwan, Casas, and Schmitz 2004). For example, many recent Demographic and Health Surveys include information on HIV infection. Because of confidentiality issues, up to 2 km of random error in any direction is added to cluster locations in urban areas and up to 5 km in rural areas. Additionally, one point in each survey is displaced up to 12 km in any direction.¹⁴ Only limited research has been conducted on the effect of random perturbation on disclosure risk or the accuracy of results. Kwan *et al.* (2004) show that the accuracy of results diminishes as the size of the perturbation circle increases. Zimmerman and Pavlik (2006) show that releasing meta-data about perturbation methods and having different masked versions of the same data set (for example, a spatially aggregated data set and a randomly perturbed data set) can considerably increase disclosure risk.

Conclusions

The removal of selective availability and the falling costs of GPS receivers have made the collection of GPS information increasingly feasible in household surveys, yet many household surveys still do not use this technology. This article argues that the collection of GPS coordinates should become a routine part of household surveys, since doing so can lead to better economics and better policy advice. In particular, the article has shown how GPS is being used to better

measure and understand the causal impacts of policies, policy externalities, and access to services and how using GPS can improve the quality of the data collected.

One of the greatest arguments to collect GPS information is the option value it gives for unforeseen future applications.¹⁵ As the stock of geo-referenced data increases within developing economies, a number of innovative applications are likely to emerge, which combine household survey data with other geographic information. In addition, interesting new research questions in econometrics and sampling methodology will emerge from the use of GPS. For example, typical household surveys often involve population-based clusters, which are not randomly spread across geographic areas (Montana and Spencer 2004). As a consequence, more research is needed to determine how best to estimate spatial models using a sample that is not spatially representative at the local level and how best to sample within communities to allow both spatial and nonspatial uses of the data.

The increasing prevalence of GPS data will have greater research value if current practices are improved. For example, the provision of accurate, clear and timely metadata about the data layers in existing GIS collections would allow more seamless and reliable merging of such data with GPS data. Ensuring respondent privacy while maximizing the research value of GPS data may be best achieved by having surveys approved by human-subject panels and by releasing data only to valid researchers who sign confidentiality provisions.

Funding

Financial support from Marsden Fund grant UOW0503 is gratefully acknowledged.

Notes

John Gibson is a Professor of Economics at the University of Waikato; his email address is jkgibson@mngt.waikato.ac.nz. David McKenzie (corresponding author) is a senior economist in the Development Research Group at the World Bank; his email address is dmckenzie@worldbank.org. The authors are grateful to Kathleen Beegle, Chris Bennett, Piet Buys, Geua Boe-Gibson, Alan de Brauw, Uwe Deichmann, John Hoddinott, Ben Olken, Duncan Thomas, John Thyne, three anonymous referees, and participants at the 2007 Stanford Institute for Theoretical Economics Conference for helpful comments and advice. This article is dedicated to the memory of Piet Buys, who championed global information system use at the World Bank through his knowledge, enthusiasm, and eagerness to help others use this emerging technology.

1. www.garmin.com/support/faqs/faq.jsp?faq=582&webPage=Main%20web%20page

2. Specifically, a “base station” GPS receiver is set up on a precisely known location and used to compare position based on the satellite signals with this known location. The difference is then

applied to other GPS receivers in the area to correct their calculations of their unknown locations.

3. The Garmin eTrex GPS unit, used in a number of household surveys, could be purchased for \$88 at Walmart.com and \$93 at Amazon.com on February 21, 2007.

4. See http://ec.europa.eu/dgs/energy_transport/galileo/index_en.htm for more details [accessed March 12, 2007].

5. www.xyzmaps.com/NewPostcode.htm

6. Recent reviews of the use of GIS in economics that are based largely on developed countries are Overman (forthcoming) and Bateman *et al.* (2002).

7. <http://afghanistan.unfpa.org/projects.html> [accessed December 28, 2006].

8. Of course, researchers can use geography to create instruments without using GPS, through manual map work. Recent examples include Woodruff and Zenteno (2007), who used distance from the capital of the state in which an individual was born to the nearest station on the north–south railway lines as they existed in the early 1900s as an instrument for migration in Mexico. Hoxby (2000) used the number of stream mouths in a metropolitan area in the United States as an instrument for the number of school districts in examining the impact of school choice. GPS can make such applications more accurate and less time-consuming.

9. However, note that this application involves using distance within Tonga as an instrument for migration to another location (New Zealand). As a result, one is less concerned in this application about other geographic features in Tonga affecting the outcome of interest, since this outcome is in New Zealand.

10. See Anselin (2002) for an accessible review of spatial econometrics. Note also that in the agricultural example given here, the omitted climate and soil variables are likely to be correlated with the regressors of interest, so one will wish also to include detailed spatial variables as controls in the regression.

11. Roberts *et al.* (2006) report on a survey in Bukoba, Tanzania, where self-reported distance was compared with distances calculated by using pedometers and an estimate of average step length. They find that more than 60 percent of self-reported distances were more than twice the calculated distances.

12. A degree of latitude is 110.6 km at the equator and 111.7 km at the poles. A degree of longitude is 111.3 km at the equator, 55.8 km at 60 degrees latitude and only 16.9 km at 80 degrees latitude. Instead, the great circle distance should be used (<http://mathforum.org/library/drmath/view/51711.html>). Stata code that implements this formula is available in the “globdist” ado files written by Ken Simons (available through Stata’s online search function).

13. The Global Spatial Data Infrastructure Association provides a codebook on how to build a spatial data infrastructure. See www.gsdi.org [accessed February 21, 2007].

14. For more details see www.measuredhs.com/topics/gis/methodology.cfm

15. Turner (2006) offers some intriguing ideas about the visual representation of geo-coordinates.

References

- Andrabi, Tahir, Jishnu Das, Asim Khwaja, Tara Vishwanath, and Tristan Zajonc. 2007. “The Learning and Educational Achievement in Punjab Schools (LEAPS) Report.” World Bank, Washington, D.C.
- Anselin, Luc. 2002. “Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models.” *Agricultural Economics* 27(3):247–67.
- Bateman, I., A. Jones, A. Lovett, I. Lake, and B. Day. 2002. “Applying Geographical Information Systems (GIS) to Environmental and Resource Economics.” *Environmental and Resource Economics* 22(1):219–69.

- Beegle, Kathleen, Joachim De Weerd, and Stefan Dercon. 2006. "Kagera Health and Development Survey 2004 Basic Information Document." World Bank, Washington, D.C., www.worldbank.com/lsm/country/kagera2/docs/KHDS2004%20BID%20feb06.pdf (accessed March 13, 2007).
- Bennett, Christopher. 2006. "Using Google Earth on World Bank Projects." World Bank, East Asia and Pacific Transport Unit, Washington, D.C.
- Bithell, John. 1990. "An Application of Density Estimation to Geographical Epidemiology." *Statistics in Medicine* 9(5):691–701.
- Borjas, George. 2001. "Economics of Migration." In Neil J. Smelser, and Paul B. Baltes, eds., *International Encyclopedia of the Social and Behavioral Sciences*. Oxford: Pergamon.
- Brady, Henry, and Iris Hui. 2006. "Is It Worth Going the Extra Mile to Improve Causal Inference? Understanding Voting in Los Angeles County." Department of Political Science, University of California, Berkeley.
- Burchfield, Marcy, Henry Overman, Diego Puga, and Matthew Turner. 2006. "The Determinants of Sprawl: A Portrait from Space." *Quarterly Journal of Economics* 121(2):587–633.
- Case, Anne. 1991. "Spatial Patterns in Household Demand." *Econometrica* 59(4):953–65.
- Conley, Timothy. 1999. "GMM Estimation with Cross-sectional Dependence." *Journal of Econometrics* 92(1):1–45.
- Conley, Timothy, and Christopher Udry. 2005. "Learning about a New Technology: Pineapple in Ghana." New Haven, CT: Yale University.
- Cowen, David, and John Jensen. 1998. "Extraction and Modeling of Urban Attributes Using Remote Sensing Technology." In D. Liverman, E. Moran, R. Rindfuss, and P. Stern, eds., *People and Pixels: Linking Remote Sensing and Social Science*. Washington, D.C.: National Academy Press.
- Cromley, Ellen, Robert Cromley, and Yanlin Ye. 2004. "On-Line Reporting and Mapping of Spatially Aggregated Individual Records Selected by User Queries." *Cartographica* 39(2):5–13.
- Curtis, Andrew, Jacqueline Mills, and Michael Leitner. 2006. "Spatial Confidentiality and GIS: Re-engineering Mortality Locations from Published Maps about Hurricane Katrina." *International Journal of Health Geographics* 5(1):44–56.
- Deininger, Klaus, and Bart Minten. 2002. "Determinants of Deforestation and the Economics of Protection: An Application to Mexico." *American Journal of Agricultural Economics* 84(4):943–60.
- De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2007. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." Policy Research Working Paper 4230. Washington, D.C: World Bank.
- Duranton, Gilles, and Henry Overman. 2005. "Testing for Localisation Using Micro-Geographic Data." *Review of Economic Studies* 72(4):1077–1106.
- Dwolatzky, Barry, Estelle Trengove, Helen Struthers, James McIntyre, and Neil Martinson. 2006. "Linking the Global Positioning System (GPS) to a Personal Digital Assistant (PDA) to Support Tuberculosis Control in South Africa: A Pilot Study." *International Journal of Health Geographics* 5(1):34.
- El-Rabbany, Ahmed. 2006. *Introduction to GPS: The Global Positioning System*. Boston, MA: Artech.
- Entwisle, Barbara, Albert Hermalin, Peerasi Kamnuansilpa, and Apichat Chamrathirong. 1984. "A Multilevel Model of Family Planning Availability and Contraceptive Use in Rural Thailand." *Demography* 21(4):559–74.
- Entwisle, Barbara, Ronald R. Rindfuss, Stephen J. Walsh, Tom P. Evans, and Sara R. Curran. 1997. "Geographic Information Systems, Spatial Network Analysis, and Contraceptive Choice." *Demography* 34(2):171–87.

- Escobal, Javier, and Sonia Laszlo. 2005. "Measurement Error in Access to Markets." Montreal, Quebec: McGill University.
- Fafchamps, Marcel, and Jackline Wahba. 2006. "Child Labor, Urban Proximity and Household Composition." *Journal of Development Economics* 79(2):374–97.
- Fefferman, Nina, Eileen O'Neil, and Elena Naumova. 2005. "Confidentiality and Confidence: Is Data Aggregation a Means to Achieve Both?" *Journal of Public Health Policy* 26(4):430–49.
- Freedman, David. 2001. "Ecological Inference and the Ecological Fallacy." In Neil J. Smelser, and Paul B. Baltes, eds., *International Encyclopedia of the Social and Behavioral Sciences*. Oxford: Pergamon.
- Gibson, John, and Susan Olivia. 2007. "Spatial Autocorrelation and Non-Farm Rural Enterprises in Indonesia." Paper presented at the 51st Conference of the Australian Agricultural and Resource Economics Society, February 13–16, Queenstown.
- Gibson, John, Geua Boe-Gibson, Halahingano Rohorua, and David McKenzie. Forthcoming. "Efficient Financial Services for Development in the Pacific." *Asia-Pacific Development Journal*.
- Goodchild, Michael. 1992. "Geographical Information Science." *International Journal of Geographical Information Science* 6(1):31–45.
- Heckman, James. 1997. "Instrumental Variables: a Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources* 32(3):441–62.
- Heckman, James, Hidehiko Ichimura, and Petra Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64(4):605–54.
- Hong, Rathavuth, Livia Montana, and Vinod Mishra. 2006. "Family Planning Services Quality as a Determinant of Use of IUD in Egypt." *BMC Health Services Research* 6(1):79.
- Hoxby, Caroline. 2000. "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review* 90(5):1209–38.
- Kumar, Naresh. 2007. "Spatial Sampling for Collecting Demographic Data." Paper presented at the Annual Meeting of the Population Association of America, March 29–31, New York.
- Kwan, Mei-Po, Irene Casas', and Ben Schmitz. 2004. "Protection of Geoprivacy and Accuracy of Spatial Information: How Effective are Geographical Masks?" *Cartographica* 39(2):15–28.
- Lalasz, Robert. 2006. "In the News: The Nigerian Census." Population Reference Bureau, www.prb.org/Template.cfm?Section=PRB&template=/ContentManagement/ContentDisplay.cfm&ContentID=13767 (accessed December 28, 2006).
- Landry, Pierre F., and Mingming Shen. 2005. "Reaching Migrants in Survey Research: The Use of the Global Positioning System to Reduce Coverage Bias in China." *Political Analysis* 13(1):1–22.
- McKenzie, David, and Yaye Seynabou Sakho. 2007. "Does It Pay Firms to Register for Taxes? The Impact of Formality on Firm Profitability." Washington, D.C: World Bank.
- McKenzie, David, John Gibson, and Steven Stillman. 2006. "How Important is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration." Policy Research Working Paper 3906. Washington, D.C: World Bank.
- . 2007. "A Land of Milk and Honey with Streets Paved with Gold: Do Emigrants Have Over-Optimistic Expectations about Incomes Abroad?" Policy Research Working Paper 4141. Washington, D.C: World Bank.
- Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Effects on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1):159–217.

- Montana, Livia, and John Spencer. 2004. "Incorporating Geographic Information into MEASURE Surveys: A Field Guide to GPS Data Collection." MeasureDHS, www.measuredhs.com/basicdoc/gps/DHS_GPS_Manual.pdf (accessed February 21, 2007).
- Noor, Abdusalán, Abdinasir Amin, Peter Gething, Peter Atkinson, Simon Hay, and Robert Snow. 2006. "Modelling Distances Travelled to Government Health Services in Kenya." *Tropical Medicine and International Health* 11(2):188–96.
- Olken, Benjamin. 2006. "Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages." BREAD Working Paper 130. Durham, NC: Bureau for Research and Economic Development.
- Oster, Emily. 2006. "HIV and Sexual Behavior Change: Why Not Africa?" University of Chicago.
- Overman, Henry G. Forthcoming. "Geographical Information Systems (GIS) and Economics." In S. Durlauf and L. Blume eds., *The New Palgrave Dictionary of Economics*. New York: Palgrave Macmillan.
- Perry, Baker, and Wil Gessler. 2000. "Physical Access to Primary Health Care in Andean Bolivia." *Social Science and Medicine* 50(9):1177–88.
- Ramachandran, R. 2000. "Public Access to Indian Geographical Data." *Current Science* 79(4): 450–67.
- Ravallion, Martin. Forthcoming. "Evaluating Anti-Poverty Programs." In R.E. Evenson, and T.P. Schultz eds., *Handbook of Development Economics*. Vol. 4. Amsterdam: North-Holland.
- Roberts, Peter, K.C. Shyam, and Cordula Rastogi. 2006. "Rural Access Index: A Key Development Indicator." Transport Papers 10. Washington, D.C.: World Bank, Transport Sector Board.
- Rosero-Bixby, Luis. 2004. "Spatial Access to Health Care in Costa Rica and Its Equity: A GIS-Based Study." *Social Science and Medicine* 58(7):1271–84.
- Spencer, John, Brian Frizzelle, Philip Page, and John Vogler. 2003. *Global Positioning System: A Field Guide for the Social Sciences*. Oxford: Blackwell Publishing.
- Staal, S., C. Delgado, I. Baltenweck, and R. Kruska. 2003. "Spatial Aspects of Producer Milk Price Formation in Kenya: A Joint Household-GIS Approach." Paper presented at the 24th Conference of the International Association of Agricultural Economists, August 13–18, 2000, Berlin.
- Staal, S., I. Baltenweck, M. Waithaka, T. de Wolff, and L. Njoroge. 2002. "Location and Uptake: Integrated Household and GIS Analysis of Technology Adoption and Land Use, with Application to Smallholder Dairy Farms in Kenya." *Agricultural Economics* 27(2):295–315.
- Turner, Andrew. 2006. "Introduction to Neogeography." *O'Reilly Short Cuts*, December.
- USAID/Timor-Leste. 2004. "East Timor Completes the World's First GPS-Based Census." USAID Timor-Leste Small Grants Program Highlights, <http://timor-leste.usaid.gov/PrintVersion/SGArchive51Print.htm> (accessed December 28, 2006).
- VanWey, Leah, Ronald Rindfuss, Myron Gutmann, Barbara Entwisle, and Deborah Balk. 2005. "Confidentiality and Spatially Explicit Data: Concerns and Challenges." *Proceedings of the National Academy of Sciences* 102(43):15337–42.
- Woodruff, Christopher, and Rene Zenteno. 2007. "Migration Networks and Microenterprises in Mexico." *Journal of Development Economics* 82(2):509–28.
- Zimmerman, Dale, and Claire Pavlik. 2006. "Quantifying the Effects of Mask Metadata Disclosure and Multiple Releases on the Confidentiality of Geographically Masked Health Data." Iowa City: University of Iowa Department of Biostatistics.