Title
Essays in Development Economics

Author
Lee, Diana Kim

Publication Date
2014

Peer reviewed|Thesis/dissertation

Essays in Development Economics

by

Diana Kim Lee

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Ethan Ligon, Chair
Associate Professor Jeremy Magruder
Professor Elisabeth Sadoulet
Professor Pranab Bardhan

Spring 2014

Essays in Development Economics

Abstract

Essays in Development Economics

by

Diana Kim Lee

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Associate Professor Ethan Ligon, Chair

Poverty and health are two topics in the field of development economics that are of critical importance to both researchers and policy-makers. Despite advances in poverty alleviation and gains in health outcomes in many developing countries, many challenges remain. Two of these challenges include accurately measuring poverty and improving the quality of health care delivery systems. In this dissertation, I present three essays with theoretical, empirical, and policy-relevant insights into these two challenges.

The first essay addresses the issue of accurate poverty measurement by developing a new asset index that captures long run household economic well-being. The accurate measurement of household well-being is necessary for measuring poverty levels and targeting poverty programs. However, since standard expenditure aggregates are costly to collect, relative well-being in developing countries is often measured using asset indices based on durable goods ownership. Although various methods exist to generate proxies for economic well-being (e.g., principal component analysis), the underlying theories associated with these methods have not been formalized. This makes it difficult to interpret the economic meaning of the resulting indices and can lead to inaccurate targeting and evaluation. In this paper, I develop a new asset index, the utility index, by modeling and structurally estimating household preferences over discrete assets. By drawing from economic theory, the utility index can be more directly interpreted as capturing long run household well-being. In contrast to existing asset indices, the utility index incorporates additional information on prices, demographics, and spatial and temporal variation and can therefore be used for policy simulations that are not otherwise possible. After developing the theoretical model, I describe a strategy to construct the utility index by structurally estimating the marginal utility associated with each asset. I then demonstrate how the utility index can be used by measuring changes in poverty in Nicaragua using data from the Living Standards Measurement Surveys. I also use the model to project changes in poverty under a constant income distribution but changing prices and find that about a third of the poverty decrease measured from 1998 to 2005 can be attributed to decreasing asset prices. In addition, I show through the empirical analysis that traditional asset indices are only moderate approximations for household well-being. Finally, I discuss

and demonstrate the distinctions between asset and consumption measures, which point to the complementary nature of the two strands of measurement.

The second essay presents an alternate approach for improving accurate poverty measurement in developing countries. Although the utility index developed in the first essay presents a method for measuring long run economic well-being, complementary measures of short run welfare are necessary for identifying households which are vulnerable to falling into transitory poverty. Again, given the expenses associated with collecting full consumption data, researchers have developed methods to construct wealth indices based on dichotomous asset and consumption indicators. This work provides guidance on generating such indices by comparing across various methods of construction and variable choices. Specifically, we assess the performance of alternate indices using data from the Living Standards Measurement Surveys in five countries in Sub-Saharan Africa–Ghana, Rwanda, Uganda, Tanzania, and Malawi. We compare indices against a benchmark of household per capita expenditure according to three criteria: rank correlation coefficients, sensitivity to identifying poor households, and accuracy of classifying households as poor or non-poor. Comparing across construction methods, we find that indices generated using principal components analysis correspond most closely with expenditure, though variation across construction methods is small. Comparing across variable inclusion groups, we find that indices generated using a combination of indicators drawn from the categories of staple food consumption, other food consumption, housing quality, semi-durables expenditure, and durables ownership tend to outperform indices generated using variables from only one or two categories. We also assess the various indices in urban and rural subsamples and in analyses of repeated cross-sections and find that index performance is similar to what we find in national, single wave analyses.

The third essay turns to the challenge of improving the quality of health care delivery systems by looking at provider investment decisions. Pay-for-performance (P4P) programs, which aim to increase health service provision and quality using financial incentives, have been recently introduced in a number of developing countries. P4P programs contract directly on outputs without specifying the mechanisms for improvements, allowing providers to innovate and modify different aspects of health care delivery as needed. Characterizing these provider responses can help to identify successful mechanisms for quality improvement and enhance our understanding of the links between P4P and overall health systems strengthening. In this paper, we examine provider input responses to the Rwandan P4P program using facility-level data from the 2007 Demographic and Health Survey Service Provision Assessment (SPA) collected after the randomized program rollout to a subset of districts. We focus on facility-level incentives for institutional deliveries, which, as documented in earlier research, resulted in higher institutional delivery rates. Using the SPA facility data, we find that the program's effect on institutional delivery rates is comparable to results in previous studies that used household surveys. Comparing system inputs, we find positive treatment effects for a general management indicator and the daily presence of staff per capita providing maternity-related services. There are no differences in other delivery-specific and general health care delivery inputs. Additionally, we perform a mediation analysis to assess the link between inputs and outcomes and find that management and staffing differences explain a relatively small fraction of the P4P effect on institutional delivery rates. The small mediation effects indicate the potential importance of unobserved factors, such as recruitment effort, in the provider production function. Furthermore, the null results for the other analyzed inputs suggest a weaker link between P4P and overall health system strengthening.

To my mother

Thank you for being there every day and every night–for bringing me to the library when I was little, for playing with me, for bringing me to church, for cooking for us, for walking with us, for working to put me through college, for being proud of me even when I failed, for praying for me, for loving me, and for being you. You were stronger and more intelligent than you ever knew, and I will continue to carry you with me until we meet again.

Contents

List of Figures

List of Tables

Acknowledgments

My dissertation would never have been possible without the guidance and support of my committee members, colleagues, friends, and family.

I would like to express gratitude to my advisor, Dr. Ethan Ligon, for supporting my idea and for strengthening the economic grounding of my work through his many theoretical insights. I would like to thank Dr. Elisabeth Sadoulet for encouraging me and clarifying the narrative and relevance of my research. I am also thankful for the invaluable comments provided by Dr. Jeremy Magruder, Dr. Pranab Bardhan, and the participants of Berkeley's Agricultural and Resource Economics Development Workshop. I am grateful for my co-authors, Dr. Sebastian Bauhoff, Dr. Tisamarie Sherry, and Dr. Luc Christiaensen, whose abundant ideas and enthusiasm made research enjoyable.

I am forever indebted to the friends and family who carried me through my graduate years. My officemates, Dr. Andrew Dustan, Dr. Lydia Ashton, and Dr. Tiffany Shih, made classes and research bearable, and my community at New Hope Covenant Church and Holy Ground small group gave me new perspective each day. I am especially thankful for my roommates, Sophia Yu and Yaa Bruce, for shared times of work and relaxation and for my mother, father, brother, and extended family for loving me and always believing in me.

Finally, I am thankful to my husband, Federick Ngo, for his infernal optimism and to God, who makes the sun rise each and every day.

Chapter 1

Utility Index

Measuring poverty using asset ownership: Developing a theory-driven asset index incorporating utility and prices

Poverty alleviation for households in developing countries remains one of the central goals of development economists and practitioners. The accurate measurement of household economic well-being is a prerequisite for poverty alleviation and is necessary for monitoring poverty changes and identifying beneficiaries for poverty programs. Although economic well-being has traditionally been measured using consumption expenditure aggregates, the high costs of data collection have led to the development of alternate measures of household well-being. One alternate measurement strategy uses data on durable goods ownership to generate asset indices for the purpose of rank ordering households. Various methods for constructing asset indices exist, but the underlying economic theory associated with current constructs is somewhat informal, making it difficult to interpret the resulting indices as valid measures of well-being. This is a practical problem because policy-makers currently use these asset indices to target and evaluate poverty alleviation programs. The use of measures which do not have a clear economic meaning can potentially result in wasted financial resources due to mis-targeting or erroneous conclusions about program effectiveness.

I therefore pose a new strategy for generating an asset index with a clear economic meaning. This index, which I call the utility index, is based directly on a theory of household demand for assets. The utility index offers an improvement over existing strategies because it is structurally estimated and meaningfully incorporates information on price, household demographics, location, and temporal variation—data not used in existing asset indices. Furthermore, the utility index can be directly interpreted as expected household well-being (i.e., utility) and can be used for policy simulations that are not possible with existing asset indices. It also offers the potential for rescaling household well-being onto a cardinal scale using the mapping from income to utility provided by the indirect utility function.

To develop the utility index, I model household utility and estimate the parameters of that model. Instead of using the monetary value of asset stocks as the relevant measure of well-being, I introduce a model of utility for two main reasons. First, the ultimate goal is measuring household well-being and not monetary asset values. To this end, I am interested in measuring the benefits to households provided by different assets, and these benefits are imperfectly captured by price. This is particularly the case when prices are determined primarily by supply-side factors such as changes in technology. Second, by introducing a model, I can make projections about how asset stocks respond to price and income changes. This is important in characterizing the extent to which changes in well-being are driven by technology, changes in income, or both.

I develop the utility index using the basic economic theory of utility maximization. Specifically, I assume that observed asset stocks are the optimal solutions to a household utility maximization problem which allows for asset resale.[1] In this problem, a representative member of the household chooses which discrete assets and consumption goods to purchase this period according to the utility she receives from the various items. I specify a within-period utility function which accounts for the club good nature of the discrete assets and the subsistence consumption requirements of the household. The income available to the household this period is determined by the sum of household wealth (i.e., the value of the accrued asset stock) and additional income received this period. To solve the problem, the member maximizes utility subject to the budget constraint and chooses to hold the bundle of discrete goods which gives her the highest utility, taking into account the consumption losses associated with asset expenditures.

I use this model to describe an empirical strategy to measure household well-being. To do so, I first estimate the marginal utilities associated with each of the discrete assets. To estimate the model, I allow prices to vary spatially and temporally, but I assume that the utility parameters are constant. I estimate the utility parameters by maximizing the derived log likelihood function using data on household size, asset ownership, and price. This maximum likelihood estimation involves computationally solving the nested nonlinear discrete choice problem. The marginal utility parameters are identified by population ownership levels relative to price, with frequent ownership of relatively expensive items consistent with higher marginal utility. I then use the estimated parameters to calculate the expected utility of each household conditional on bundle ownership, price, and household size. This expected utility is the measure of household well-being.

I demonstrate the empirical characteristics of the utility index using the 1998, 2001, and 2005 rounds of the Nicaraguan Living Standards Measurement Surveys. I find that the estimated marginal utilities of this set of discrete assets are broadly, but not strictly, increasing in price. Exceptions include televisions, which provide highly valued services relative to price, and items such as VHS players and toasters, which provide lowly valued services relative to price. I then demonstrate how the utility index adjusts for spatial and temporal price variation. By accounting for the utility provided by each asset, the utility index ranks the same bundle similarly across time and space, in contrast to a strategy which directly uses asset prices, which have dropped substantially over time.

[1] I provide some evidence for the validity of the asset resale assumption in the context of the empirical demonstration in Nicaragua. I also discuss the model's validity when the assumption fails.

I also demonstrate the predictive power of the model by assessing the extent to which price decreases in Nicaragua have contributed to improved well-being. I do so by projecting asset ownership holding constant the 1998 inferred income distribution while lowering prices to 2005 levels. I find that decreases in asset prices account for about a third of the decrease in poverty. This exercise highlights the additional insights about household well-being which are generated using a theoretical model.

I then empirically test the ability of traditional asset indices to approximate household wealth. I use the utility index as the benchmark since it is theoretically derived and exploits additional information on demographics and spatial and temporal price variation. I find that three traditional asset indices (the inverse frequency weighting method, a standard factor analysis index, and an expanded factor analysis index incorporating household size and urban location) are only moderate empirical approximations for expected utility. They generate weights with little relation to marginal utility, sometimes resulting in negative weights with questionable economic interpretation. Furthermore, the inferences about poverty and inequality generated using the traditional asset indices only weakly approximate the findings associated with the utility index.

Finally, I discuss and demonstrate how the utility index differs from the standard consumption aggregate. Although asset measures are often discussed as proxies for consumption aggregates, assets are stocks while consumption aggregates capture a flow. The two measures therefore often respond differently to temporal shocks. I show this distinction empirically, finding that, despite the moderate correlation between the two measures within each survey round, the associated poverty trends move in opposite directions.

The rest of the paper proceeds as follows: I discuss the existing literature on asset indices and asset ownership in Section 1.1. Sections 1.2 and 1.3 describe the model and empirical strategy. Section 1.4 describes the data and empirical results. This includes the decomposition of poverty into price effects, the assessment of traditional asset indices, and the theoretical and empirical discussion of distinctions between asset and consumption measures. Section 1.5 concludes.

1.1 Background: Existing literature

Researchers and policy-makers interested in issues of poverty need accurate measures of household economic well-being. These measures are necessary for many applications, such as monitoring poverty, identifying poor households for assistance programs, and controlling for socioeconomic status in empirical analyses. Although household economic well-being has traditionally been captured using household per capita expenditure aggregates, the data collection efforts associated with expenditure aggregates are time consuming and expensive. This has led to the development of alternate measures of household well-being.

Over the past two decades, researchers have begun to use information on asset ownership to generate proxies for household economic well-being. Durable asset ownership provides summary information on the long run economic situation of households and therefore provides a natural source of information for measuring economic well-being. Since assets are accrued over time, the current stock of assets provides a measure of household savings over the preceding periods. In the case of the poorest households, basic assets such as furniture and appliances often constitute the majority of household savings, particularly in areas with low financial literary or restricted access to formal savings instruments. Asset ownership also provides a reasonable measure of future welfare, since the assets provide services which can be enjoyed by households in future periods and are also relatively insensitive to temporary negative shocks. It is important to note that the durability of assets imply that they are not well suited to measuring short term welfare.

Data on asset ownership offer additional practical benefits for the measurement of household well-being. They are relatively inexpensive to collect, consisting of a set of ten to fifteen yes or no questions and associated questions on asset values. As dichotomous variables, assets are arguably subject to lower measurement error than the standard consumption measures. For these reasons, data on asset ownership are widely available and are often collected in survey instruments such as the Demographic and Health Surveys (DHS) which are not focused on economic well-being. These surveys are designed to measure non-economic dimensions of well-being but include data on asset ownership to capture basic information on economic well-being.

Researchers have recognized the benefits of using asset ownership to proxy for wealth and have proposed a variety of strategies for constructing proxies for well-being from asset ownership. I describe the three main strands of strategies for constructing asset indices: assigned weights, latent variable models, and statistical models. Within the strategy of assigned weights, the simplest scheme applies equal weights to all assets (Bollen, Guilkey, and Mroz 1995; Gorbach et al. 1998; Guilkey and Jayne 1997; Havanon, Knodel, and Sittitrai 1992; Jensen 1996; Razzaque et al. 1990). Others have attempted to determine weights using policy analysis and consultation with national experts (Navajas et al. 2000). Within the category of latent variable models, ownership propensities estimated using logit or probit models are used to generate estimates of the underlying latent variable (argued to be wealth) (Ferguson et al. 2003; Montgomery et al. 2000). Similarly, using the rationale that less commonly owned goods are generally more valuable, Lyte et al. (2001) and Morris et al. (2000) apply weights based on the inverse of population ownership levels. The statistical strategies for constructing asset indices are most common, and rely on methods such as factor analysis and principal component analysis (Filmer and Pritchett 2001; Vyas and Kumaranayake 2006). Aimed at reducing data into fewer dimensions, both procedures derive weights based on the variance structure of the underlying data. Although some rationalize the use of such methods under the assumption that wealth drives the underlying processes determining the correlation between asset ownership (Vyas and Kumaranayake 2006), the lack of corresponding economic theory is explicitly noted by most researchers (Filmer and Pritchett 2001; Vyas and Kumaranayake 2006).

To date, these measures have been and continue to be used to measure changes in poverty (Sahn and Stifel 2000; Stifel and Christiaensen 2007) and to identify beneficiaries for pro-poor programs in places such as Ecuador (Schady and Araujo 2008), Armenia, Brazil, Colombia, and Indonesia (Caldés, Coady, and Maluccio 2006). As such, it is critical that the chosen indices are valid measures of well-being since measures which may not be strongly related to household well-being will result in errors in poverty measurement and program targeting. Due to the lack of formal economic theory underlying existing asset indices,[2] it is possible that the resulting measures are only weakly related to household well-being. This is especially possible since the existing strategies do not exploit data on prices and demographics, which are important components of asset demand.

At the same time, there is a rich body of literature characterizing demand for durable assets. Deaton and Muellbauer (1988) present an overview, listing special features associated with durable assets which require careful modelling. These features fall into three categories-durability, market complexities, and discrete choice. First, because assets are durable and contribute to utility through the provision of services, distinctions need to be made between purchases which contribute to asset stocks and flows of service consumption. Durability also introduces additional factors which affect demand. These factors include the effect of existing stocks on purchasing constraints and future preferences, the importance of future expectations on advancing and postponing purchasing decisions, and the incorporation of adjustment and transaction costs into new and replacement demand. Second, market complexities arise due to significant information imperfections. Second-hand markets for durable assets can be incomplete or non-existent due to the information asymmetries existing between buyers and sellers regarding value and quality. Furthermore, demand may be slow to diffuse through populations for assets associated with new technologies. Finally, durable assets are associated with the choice between ownership and non-ownership, which requires special treatment in estimations using cross-sectional microeconomic data (Deaton and Muellbauer 1988).

Various models have arisen to address these issues. I focus my discussion here on classes of models relevant to estimating uncovering underlying preference parameters using cross-sectional microeconomic data.[3] The basic discrete choice model has one indirect utility function associated with each asset ownership level, zero or one, for the case of a single discrete asset. Households choose the ownership level associated with highest indirect utility, given their income level. The applications of this model are extensive and have included the estimation of quasi-Engel curves linking ownership to income levels (Bonus 1973). One application which is particularly pertinent to the problem of generating an asset index is the estimation of preference parameters for durable assets. Berry, Levinsohn, and Pakes (1995) provide an estimation strategy for estimating differentiated product demand using aggregate data in static settings with extensions to dynamic settings (Carranza 2007). Additional extensions incorporate microdata to match consumer characteristics with product choices

[2] The structure underlying the latent variable models assumes that households with an underlying level of wealth above a certain threshold have an increased probability of owning the asset. Although this can be consistent with a theoretical model where households have hierarchical preferences, the underlying assumptions are particularly strong and unlikely to hold in reality.

[3] Deaton and Muellbauer (1988) provide background on other approaches developed to identify rates of investment and costs of adjustment using stock adjustment models and discretionary replacement models.

(Goldberg 1995).

In this paper, I contribute to the literature of poverty measurement by drawing from the body of work on durable asset demand. Because the number of unique features associated with durable assets is sizable, I only address a subset which is particularly relevant to the context of generating household rankings. I directly address the discrete nature of durable assets but abstract away from the dynamic issues associated with durability. I later provide empirical evidence that dynamics less of a concern in this particular application context. I briefly discuss the implications of market complexities but leave serious consideration of these issues to future work.

1.2 Model

As discussed above, the goal of this work is to construct a proxy for economic well-being that can be estimated using data available to practitioners. Specifically, the task at hand is to use data on asset stocks and prices observed in cross-sectional data to rank order households according to their economic well-being. In developing the theoretical model with this goal in mind, I also discuss the assumptions necessary to use asset stocks as proxies for current economic well-being.

I use utility as the relevant measure of well-being, since utility is standardly used to capture how various goods and services contribute to household well-being. The strategy for developing the model and estimation is as follows: I will use a model of utility maximization to create a mapping between household asset stocks and utility. Using this mapping, I can infer unobserved household utility from observed asset ownership.

When thinking about mapping current utility to current asset stocks, it is important to be precise about the relevant utility maximization problem associated with asset purchases. The primary assumption which I make is that observed asset stocks are the optimal solution to the current period's maximization problem using accrued wealth as the relevant budget constraint. To make this assumption, I draw from the standard results of two-stage budgeting and an additional assumption. Under the theory of two-stage budgeting, the solution to the lifetime utility maximization problem can be broken into two stages. In the first stage, the household member decides how to allocate resources across periods; in the second, she decides how to spend her resources within each period. This result follows from the observation that the within-period marginal rate of substitution conditions continue to characterize resource allocation within each period, meaning the within-period maximization problem can be treated as a standard static maximization problem provided the within-period budget constraint is correctly adjusted (Blundell and MaCurdy 1999).

To use this result, I assume that observed resource allocations are the solution to the second stage of a two-stage budgeting problem. By doing so, I can treat current period allocation decisions within the framework of a within-period utility maximization problem without accounting for what has occurred in previous periods. Because of the durability of assets, I make the additional assumption that households are free to buy and sell assets every period with minimal transaction costs, reoptimizing when appropriate. Within the context of two-stage budgeting theory, this implies that households take current asset values into account when determining their resource allocation across periods. The assumption of the existence of asset resale may or may not be reasonable in different contexts. I therefore present some evidence that the assumption is reasonable in the context of the empirical application and discuss the implications of the method when the assumption is violated in Section 1.4.

Under these assumptions, I write down the lifetime utility maximization problem and constraints following Blundell and MaCurdy (1999). I discuss the utility function in terms of a representative household member. The member chooses over a set of discrete assets, $x_1, ..., x_J$, and one consumption good, c, the numeraire. Let $u(c, x)$ denote the within-period utility function and $V(m, p)$ denote the value function. To make the model tractable and to apply two-stage budgeting, I assume utility is separable across time.

$$V(m,p) = \max_{c,x_1,...,x_J} u(c, x_1, ..., x_J) + \beta \int V(m', p')dF(y', p')$$

The intertemporal budget constraint is given by the time path of savings, s, which are affected by the return on savings, r, other income, y, and asset prices, p. The relevant asset prices are the prices for which the assets, taking appreciation and depreciation into account, can be bought and sold locally.

$$s' = (1 + r')(s + y - c) + p'_1 x_1 + ... + p'_J x_J$$

Under two-stage budgeting theory, the solution to the problem can be found by backwards induction. Denoting m as the resources allocated to the current period, utility can be maximized for the period conditional on m. This results in an indirect utility function $v(m, p)$ for each period. The solution for m can be found by substituting the indirect utility functions into the lifetime utility function, giving the optimal m as a function of prior period savings, the current interest rate, other current income, and prior period asset stock holdings.

$$m = M(s^*, r, y, x^*)$$

where s^* and x^* represent savings and asset ownership from the previous period, respectively. Using m, the choice over consumption and asset purchases can then be correctly characterized as

$$\max_{c,x_1,...,x_J} u(c, x_1, ..., x_J)$$

subject to

$$c + p_1 x_1 + ... + p_J x_J \leq m.$$

Within-period utility maximization

I turn now to defining and solving the within-period maximization problem. In specifying the problem, I characterize the preference parameters that I will later estimate using observed data. These parameters can then be used to infer household well-being from observed asset ownership. Although I focus on one particular parametrization of utility, one benefit of introducing a model is that researchers can modify the model to identify parameters which may be particularly relevant for specific applications.

In the utility function I choose, I assume that the marginal utility associated with each asset is independent of total household size. The assets included in such indices are items such as televisions and refrigerators, for which this assumption is generally reasonable. I also assume that the representative household member takes subsistence consumption for all household members into account. For simplicity, I assume that there is no complementarity between the discrete goods and between the discrete goods and consumption.

As with the lifetime utility function, the within-period utility function is defined over one consumption good c (the numeraire) and J discrete goods, each denoted x_{jt}. The bundle of discrete goods is denoted by x. The discrete goods have an associated price vector, p, and the budget constraint is given by household wealth, m, as defined previously. A constant relative risk aversion (CRRA) utility function is assumed for the consumption good, subject to a subsistence level of consumption, ϕ, which is scaled by household size. Ownership of each discrete good is associated with a scalar, ψ_j, which captures the marginal utility gained from the services rendered by the discrete good over the relevant time period. There are 2^J potential bundles of discrete goods. The within-period utility maximization problem is given by

$$\max_{c,x_1,\ldots,x_J} \left[\frac{(c-\phi h)^{1-\gamma}-1}{1-\gamma}\right] + \psi_1 x_1 + \ldots + \psi_J x_J$$

$$s.t.\ c + p_1 x_1 + \ldots + p_J x_J \leq m$$

$$x_j \in [0,1].$$

The solutions follow Hanemann's discrete choice model of consumer demand, using a specified utility function (Hanemann 1984). To solve the problem, the household will spend whatever income that is not spent on the discrete goods on the consumption good ($c = m - p_1 x_1 - \ldots p_j x_j$). To determine the optimal bundle of discrete goods, the household calculates the indirect utility associated with each potential bundle of discrete goods and chooses the bundle with the highest conditional indirect utility given the household's income. Defining the conditional indirect utility for bundle x^k as

$$\upsilon(m,p|x^k) = \left[\frac{(m - p_1 x_1^k - \ldots p_J x_J^k - \phi h)^{1-\gamma}-1}{1-\gamma}\right] + \psi_1 x_1^k + \ldots + \psi_J x_J^k$$

the general solution is then given by the bundle, x^n, such that

$$v(m,p|x^n) \geq v(m,p|x^k) \qquad \forall x^k \ s.t.\ k \neq n.$$

That is, the optimal bundle is the bundle associated with the highest conditional indirect utility. Using the case of two discrete goods with parameters $\phi = 1$, $\gamma = 0.5$, $\psi_1 = 2$, $\psi_2 = 6$, $p_1 = 5$, and $p_2 = 16$, I illustrate the within period utility maximization problem. Again, the two goods can be imagined to be a radio and television. Under these parameters, the radio is cheaper than the television and also provides less utility in each period. Figure 1.1 shows the four indirect utility functions associated with each potential bundle. The solid black line represents the indirect utility associated with owning neither of the two goods; it is increasing with income at a decreasing rate due to the functional form of the utility function. The gray dotted line represents the indirect utility associated with owning only the radio, the cheaper good. It appears at the level of income at which the radio becomes affordable (taking into account subsistence consumption) and is translated upwards due to the utility associated with radio services. The other two utility bundles can be similarly described.

Figure 1.2 shows the optimal bundles associated with each level of income. At the lowest levels of income, it is optimal for households to purchase neither good and spend all their income on consumption. At slightly higher levels of income, it is optimal to purchase the cheaper radio but not the more expensive television. For households with moderate levels of income, it becomes optimal to forego purchasing the radio and to instead purchase the more expensive television. This occurs because households enjoy the services provided by the television more than the services provided by the radio and should therefore purchase the television when it is affordable. Finally, the bundle with both the radio and the television becomes optimal at the highest levels of income. It is not shown in the figure, but larger households require slightly higher income before purchasing the same non-zero bundle as a smaller household due to the need to satisfy subsistence consumption.

As a note, more expensive assets may not provide more highly valued services. For example, the utility associated with a more expensive good such as a CD player might be lower than that associated with a cheaper good like a radio. In this case, the bundle with a CD player but no radio would be dominated everywhere by the bundle with a radio but no CD player, and households would only choose to own CD players if they also own a radio.

Before moving to the empirical estimation strategy, I highlight one strength of using discrete assets to infer well-being. As illustrated by the graphical solutions, each optimal bundle maps to a distinct interval of utility. Thus, a household's level of well-being can be directly inferred from asset ownership without data on consumption expenditures if asset purchases are optimal. In this way, a fair amount of discrimination across households can be achieved using data on a relatively small set of assets.

1.3 Empirical strategy

Returning to the goal of generating a measure of well-being, the asset demand model offers a natural way of ranking households. This can be done if the parameters of the utility function are known. In this section, I provide an overview of the empirical strategy for estimating the parameters of the utility function. The details for the empirical estimation can be found in the Appendix.

To estimate the model, I use a latent variable maximum likelihood estimation strategy, where I assume a distribution for the underlying unobserved total income, derive the likelihood of observing a set of preference parameters, and then find the parameters which maximize the likelihood function under a set of constraints. The estimation strategy takes observations on price, household size, and asset stocks from cross sectional data and infers the subsistence level of consumption, ϕ, and the marginal utility parameters, ψ, in the utility function specified above.

The data consists of a set of households, indexed by i, which exist in various regions, indexed by r, and time periods, indexed by t. Prices are observed and are measured as the average reported current resale price within each region and time period. Quality and depreciation are accounted for through the use of reported resale prices. Asset stocks are measured as dichotomous variables.[4] Empirically, each bundle is a vector where each component indicates ownership of a different asset. The length of each vector is the total number of assets.

$$x = \{x_1, x_2, ...x_J\}$$

where $x_j \in [0, 1]$. I use superscripts to denote distinct bundles. The set of potential bundles is denoted by so x^k, where $k \in 1, ..., 2^J$.

Household preferences are unobserved and the goal of the empirical estimation. Household preferences are distributed around the average population preferences, $\bar{\psi}$, which I assume are constant across time and space. Although average population preferences may vary spatially and temporally, the direction of the variation is ambiguous and therefore difficult to build into the model without additional information. For instance, marginal utilities for certain goods may increase over time as new technology improves the quality of services provided by goods such as computers. At the same time, the marginal utilities for other goods may decrease over time as improved substitutes make items such as VHS players obsolete. Similarly, the services provided by items such as cars may be valuedly differently in urban and rural areas. In one scenario, cars may be less highly valued in rural areas because usable roads may not exist. Alternatively, cars may be less highly valued in urban areas because of the existence of functioning public transportation systems. As illustrated by these examples, the direction of preference variations is ambiguous and will differ for each asset depending

[4] If households own more than one asset (i.e. two televisions), the additional assets can be separately modeled as distinct assets. This is consistent with the idea that the marginal utility associated with a second television would differ from that of a first television.

on the context. Without additional information to identify spatial and temporal variation in average utility parameters, I maintain the simplifying assumption that the average marginal utility parameters are constant. As a note, it is possible to estimate the spatial and temporal variation in utility parameters by estimating the model separately for each region-period. Although this is straightforward to implement, a comparison of households in different regions and time periods requires placing the parameters onto a common scale. Thus, this strategy involves the development of additional methodology which can be addressed in future work.

The marginal utility for asset j and household i at time t is given by

$$\psi_{jit} = \bar{\psi}_j + \epsilon_{jit}.$$

The indirect utility for bundle x_{it}^k is

$$\begin{aligned} \upsilon(m_{it}, p_{rt}|x_{it}^k, h_{it}, \phi, \psi_{it}) = & \left[\frac{(m_{it} - p_{1rt}x_1^k - ... - p_{Jrt}x_J^k - \phi h)^{1-\gamma} - 1}{1-\gamma}\right] \\ & + (\bar{\psi}_1 + \epsilon_{1it})x_1^k + ... + (\bar{\psi}_J + \epsilon_{Jit})x_J^k. \end{aligned}$$

Household i will choose bundle x_{it}^k if the indirect utility for that bundle exceeds that of every other potential bundle at the household's given total income m_{it}. The probability that an arbitrary household with total income m_{it} chooses bundle x_{it}^k is given by

$$Pr(x_{it}^k|m_{it}, h_{it}, p_{rt}, \phi, \psi_{it}) = Pr\left(\upsilon(m_{it}, p_{rt}|x^k, h_{it}, \phi, \psi_{it}) \geq \upsilon(m_{it}, p_{rt}|x^n, h_{it}, \phi, \psi_{it})\ \forall x^n \neq x^k\right).$$

This discrete choice probability is the basis for multinomial logit and probit models, where the relevant choice set is given by the set of potential bundles, not single assets. The parameters cannot be estimated by a direct application of a multinomial logit or probit, though, due to the nonlinearity of unobserved wealth in the conditional indirect utility functions. This requires additional integration over the wealth distribution, which can be incorporated into a maximum likelihood strategy. Although it is possible to ignore wealth effects when utility is assumed to be quasilinear in wealth, this is unlikely to hold in reality.

The parameters can be estimated using maximum likelihood given distributional assumptions for unobserved household preferences and wealth. Assuming that household preferences for each asset are independent and distributed normally,

$$\epsilon_{jit} \sim N(0, \sigma_j),$$

the probability that a household with total income m_{it} chooses bundle x_{it}^k can be rewritten as

$$\begin{aligned} & Pr(x_{it}^k|m_{it}, h_{it}, p_{rt}, \phi, \psi_{it}) \\ & \qquad = \prod_{x^n \neq x^k} Pr\left(\upsilon(m_{it}, p_{rt}|x^k, h_{it}, \phi, \psi_{it}) \geq \upsilon(m_{it}, p_{rt}|x^n, h_{it}, \phi, \psi_{it})\right) \end{aligned}$$

$$= \prod_{x^n \neq x^k} Pr\left(\upsilon(m_{it}, p_{rt}|x^k, h_{it}, \phi, \bar{\psi} + \epsilon_{it}) \geq \upsilon(m_{it}, p_{rt}|x^n, h_{it}, \phi, \bar{\psi} + \epsilon_{it})\right)$$

$$= \prod_{x^n \neq x^k} Pr\left(\upsilon(m_{it}, p_{rt}|x^k, h_{it}, \phi, \bar{\psi}) - \upsilon(m_{it}, p_{rt}|x^n, h_{it}, \phi, \bar{\psi}) + \epsilon_{it}(x^k - x^n) \geq 0\right)$$

$$= \prod_{x^n \neq x^k} \left[1 - \Phi\left(0, \upsilon(m_{it}, p_{rt}|x^k, h_{it}, \phi, \bar{\psi}) - \upsilon(m_{it}, p_{rt}|x^n, h_{it}, \phi, \bar{\psi}), \sigma^{kn}\right)\right]$$

where $\sigma^{kn} = \sqrt{\sigma_1^2|x_1^k - x_1^n| + ... + \sigma_J^2|x_J^k - x_J^n|}$.

By assuming a distribution for unobserved wealth, I calculate the probability of observing bundle x_{it}^k for household i at time t as

$$Pr(x_{it}^k|h_{it}, p_{rt}, \phi, \bar{\psi}, \sigma) = \int_{m_{it}} Pr(m_{it})Pr(x_{it}^k|m_{it}, h_{it}, p_{rt}, \phi, \bar{\psi}, \sigma)$$

where $m_{it} \sim \ln N(0, 1)$.

The log likelihood function for the utility parameters is then given by the probability of observing each bundle, given the price, household size, and population utility parameters, and then summed over all the observed bundles. The log likelihood function can be written as

$$\ln L(\phi, \bar{\psi}, \sigma) = \sum_{it} \ln Pr(x_{it}^k|h_{it}, p_{rt}, \phi, \bar{\psi}, \sigma).$$

Since theory implies that the subsistence parameter ϕ, the marginal utility associated with each good ψ_j, and the variance terms σ_j are non-negative, the log likelihood function can be maximized subject to a set of constraints. Additional constraints are to identify unique point solutions because the log likelihood function is flat in certain regions. This can be seen in the case of one disrete asset. The predicted population ownership share can be increased by raising the marginal utility associated with the asset. When the marginal utility associated with the asset is increased to a high enough level, asset ownership dominates non-ownership at all wealth levels for which the asset is affordable. The ownership share and associated likelihood are therefore constant at some threshold value of marginal utility and above. With more than one asset, analogous flat regions exist in the log likelihood function because the utility function is ordinal. The details for the construction of the additional constraints can be found in the Appendix.[5]

[5] An alternative estimation strategy without additional constraints would involve estimating level set solutions instead of point estimates.

I assume a lognormal distribution for income m_i and estimate the parameters which solve the following problem:[6]

$$\max_{\phi,\psi,\sigma} \ln L(\phi, \bar{\psi}, \sigma)$$
$$s.t.$$
$$0 \leq \phi$$
$$0 \leq \sigma_j \quad \forall j = 1, ..., J$$
$$0 \leq \psi_j \leq \psi_{jrt,max} \quad \forall j = 1, ..., J$$

After estimating the parameters ϕ^*, ψ^*, and σ^*, the predicted utility associated with each household is the expected value of utility as a function of household size and the observed bundle, taken over the values of m_{it} for which that bundle is affordable. The predicted utility for a given household can be written as

$$U(h_{irt}, x_{irt}) = E\left(\left[\frac{(m_{it} - p_{1rt}x_{1irt} - ... - p_{Jrt}x_{Jirt} - \phi^* h_{irt})^{1-\gamma} - 1}{1-\gamma}\right]\right.$$
$$\left. + \psi_{1rt}x_{1irt} + ... + \psi_{Jrt}x_{Jirt} \,|\, m_{it} \geq p_{1rt}x_{1irt} + ... + p_{Jrt}x_{Jirt} + \phi^* h_{irt}\right).$$

The resulting utility index is generated primarily from the variation in observed asset stocks. Secondary variation from household size and price variation is also exploited. Differently sized households with the same bundle will differ slightly in their expected utilities due to differences in their expected levels of non-subsistence consumption. Likewise, households which face different prices but own the same bundle will differ slightly in their expected utilities due to different expected levels of non-subsistence consumption. For the most part, though, the expected utilities for households with the same bundle will be relatively similar due to the assumption of homogeneous preferences.

The unconditional indirect utility function offers a direct mapping from utility to underlying wealth. This makes it possible to place the utility index onto a cardinal scale of household wealth. It is important to note, though, that the stock of wealth is not directly comparable to traditional measures of flows of income and consumption. For this reason, I use the utility index to calculate poverty using 1998 relative poverty lines for the purpose of comparison with traditional indices and consumption.

[6] Empirical estimates for the income distribution in the sample can be used instead of a hypothetical distribution. I choose to assume an income distribution since the empirical income distribution may not be readily available for the particular population and time period of interest.

Identification and distinctions from alternative strategies

Following the proposed estimation strategy, the model parameters are identified using the distribution of asset bundles observed in the data. The marginal utility parameters are determined by ownership levels relative to prices. Assets with low relative prices that appear infrequently in the population are associated with low marginal utility and are only purchased after other higher utility assets have been acquired. These low utility assets include items such as toasters and rice cookers which provide supplementary services which are also provided by other multipurpose appliances like stoves. Conversely, certain assets are relatively expensive but are commonly owned, indicating that the associated marginal utilities are relatively high. These high utility assets include items such as stoves and televisions, which households often purchase in lieu of cheaper goods when they can afford them.

This estimation strategy differs from the traditional latent variable models by incorporating price information. The traditional latent variable models are operationalized using variants of logits and probits and are identified solely from population ownership levels.[7] The traditional latent variable models assume that the probability of ownership for each asset increases with underlying wealth, which is consistent with a model where marginal utility is decreasing in price. This implies that households should always purchase goods in a specified order, for instance always buying radios first, televisions second, and computers third. Yet, even for a population with identical preferences, this ordering may not always occur. The two good solved example of the theoretical model shows a case where it is optimal for households to purchase a television without first purchasing a radio, if they can afford to do so. Thus, compared to the traditional latent variable models, the proposed estimation strategy directly incorporates price information and allows for a more flexible relationship between price and marginal utility.

Although the incorporation of price information represents one key improvement over traditional asset constructs, the proposed asset index differs from simply summing the prices of all owned assets. The ultimate goal is to measure household well-being, not the monetary value of household assets. As discussed above, marginal utility is not the same as price, as more expensive assets may not necessarily make households better off than cheaper assets which provide more fundamental services. This is consistent with empirical observations that ownership levels of assets with similar prices may vary widely, indicating differing levels of marginal utility. The distinction between price and marginal utility is most notable when the asset supply functions are highly elastic and market prices are therefore determined primarily by production factors.

Furthermore, there are issues associated with using the total monetary value of assets stocks when there is substantial spatial and temporal price variation. A meaningful wealth measure should not penalize households for decreases in asset prices associated with technological advances. For instance, a household with a television and cellular phone now is better off than a household with only a television five years ago, even though the total cost of the

[7]See Ferguson et al. (2003) and Montgomery and Hewett (2005) for detailed descriptions of two latent variable models.

television/cellular phone bundle now may be much less than the cost of just the television five years ago. The introduction of the utility model provides a meaningful way to address such price variation under the assumption that a given asset provides the same service to a household regardless of cost.

In addition, the structural estimation underlying the utility index allows the method to be used for policy simulations. Specifically, the model can be used to generate predictions about how asset ownership patterns will respond to variations in price and income as well as demographic shifts. Furthermore, the identification of marginal utility parameters also allows policy makers to identify and potentially focus on the assets which contibrute most highly to household well-being.

1.4 Data and Results

To estimate the model, I use three rounds of nationally representative data from the Nicaraguan Living Standards Measurement Surveys (LSMS), from 1998, 2001, and 2005.[8] Classified by the World Bank as one of the poorest countries in Latin America with half of the population falling below the poverty line in 1993, Nicaragua represents an area where poverty monitoring is vital and where ownership of basic goods can provide meaningful differentiation between households.

Although this period of transition towards democracy and free market policies has generally been characterized by growth, it has also been punctuated by recession, volatile export prices, and natural disasters. In addition to being variable, growth was likely also uneven across economic sectors and geographic regions, as regional disparities accounted for substantial variations in poverty in the early 1990s. This context thus presents a natural setting in which to apply the household utility model to measure differences in household well-being over time and across regions.

The LSMS datasets are comprehensive household surveys collected by country institutions in collaboration with the World Bank to capture household well-being. They therefore contain extensive information on household assets. Furthermore, the instruments include detailed modules on household consumption and expenditure, which will allow me to compare inferences made using the household utility model with those found using the standard per capita consumption measures.

To estimate the model, I use data on the twelve most commonly observed assets: irons, radios, fans, blenders, toasters, bicycles, stoves, sewing machines, VHS players, televisions, refrigerators, and cars. Although some of the goods may double as productive assets for home businesses, all the items are listed under the heading of common household equipment and provide services that can be directly enjoyed by the households. I exclude livestock from the estimation strategy as the assumption that livestock provides the same utility to both urban and rural households is likely invalid. The price of each asset is given by the

[8] An earlier LSMS was also conducted in Nicaragua in 1993 but is excluded from this analysis because it lacks data on ownership of durable goods.

average value of the good, reported by households which own it. Since households are asked to report the value of the good in its current condition, depreciation and quality have been broadly accounted for in the price.

As a note, the LSMS are rich datasets which offer the potential to explore interesting behaviors such as how households acquire and sell off different assets. For the purposes of this work, I exploit only a small fraction of the data to demonstrate the potential use of the proposed asset index. As asset indices are used primarily in datasets with information on asset ownership and prices but little additional information on living standards, I restrict my analysis to this subset of data. I also use sampling weights to allow the data to be used as a nationally representative cross section (ignoring the panel feature of the data).

Model estimates and poverty changes

Table 1.1 shows the ownership levels of the goods by year and urban/rural location. Ownership levels are broadly decreasing in price, with high levels of ownership for inexpensive items such as irons and lower levels of ownership for expensive items such as cars. This is not strictly the case, though. Specifically, relatively expensive items such as televisions are commonly owned while relatively inexpensive items such as toasters and VHS players are rare. This indicates that there is important variation in the utility provided by the different assets which can be used to construct a meaningful asset index.

Ownership for most goods increased over the observed time period in both urban and rural areas, although there were slight declines in ownership for certain goods from the 1998 to 2001 time frame. Irons in rural areas and sewing machines in urban areas are the exception, with small declines in ownership of those goods due perhaps to substitution towards other goods. Increases in ownership were particularly large, exceeding 10 percentage points for radios and bicycles in rural areas and stoves, televisions, and refrigerators in urban areas. Ownership levels of all goods, with the exception of radios, is lower in rural areas. Although iron and television ownership is relatively high in both urban and rural areas, fan and blender ownership is low relative to other goods when comparing rural areas to urban areas. This indicates that the assumption of homogeneous preferences across location is a strong assumption that should be addressed in future work.

The left panels of Table 1.2 show the prices for the included assets by year and urban/rural location. The reported values for the different assets range from 79 (for irons) to 49767 (for cars) cordobas, or 1-800% of the 1998 per capita yearly expenditure.[9] The prices for all assets declined over the eight year period in both urban and rural areas, though price increases were observed for some assets between 1998 and 2001. Price declines were extremely high for VHS players due to changes in technology and were also large (over 20% in both urban and rural areas) for radios, toasters, sewing machines, and refrigerators. Reported prices in rural areas are generally lower than the corresponding urban prices in 1998, but larger

[9]Deflation into 1998 Nicaraguan cordobas is done using consumption expenditures from the World Development Indicators.

urban price decreases caused rural prices to exceed urban prices in 2005 for five of the assets. Asset prices are similar for most goods across urban and rural locality, with the exception of televisions, which are substantially more expensive in urban areas.

I now proceed to estimate the model. The last two columns of Table 1.2 present the estimated average marginal utility parameters and the corresponding standard deviations capturing the distribution of utility parameters. The estimated subsistence parameter for the model is 172 cordobas. Average marginal utility is broadly increasing in price, indicating that households often forego cheaper goods in order to purchase more expensive ones. This is exemplified by the fact that some households (1.7% of the pooled sample) own only televisions out of all the items. Exceptions to this pattern include toasters, VHS players, televisions, and irons. The low average marginal utilities for price associated with toasters and VHS players is consistent with the fact that few households choose to own toasters or VHS players despite their affordability. Conversely, the estimated average marginal utilities for televisions and irons are high relative to their prices, consistent with the high population ownership levels. This is consistent with the fact that iron ownership exceeds radio ownership by 25 percentage points, in spite of the low price differential of one to two USD.

The estimated standard deviations, σ^*, indicate that there is a relatively wide distribution of marginal utilities across households, with higher standard deviations associated with higher average marginal utilities. This implies that the services provided by irons, radios, and fans are similarly valued. Furthermore, the estimated average marginal utilities will generate households rankings associated with a fair amount of uncertainty, particularly among household which own the cheaper asset bundles. This uncertainty is expected given the existence of heterogeneous preferences.

To show how these estimates translate into a measure of well-being, I present the rankings for the utility index for a selected set of bundles in Table 1.3. For comparison purposes, I also present the price rankings generated using the real monetary value of the bundles. The displayed price and utility rankings are averaged across urban and rural areas and all household sizes, and the ownership level is presented for the pooled sample. Looking across bundles, higher priced bundles (which have higher associated price rankings) are generally associated with higher utility rankings. This is not always the case, though, as highlighted by a few pair-wise comparisons. Households which only own fans are ranked as less well off than households which only own radios, although fans cost slightly more, on average. In the same way, households which own the bundle with only a television are ranked as better off than households which own the bundle with an iron, a stove, a fan, a bicycle, and a sewing machine despite the fact that the latter bundle consistently costs slightly more. These rankings are consistent with the fact that the slightly more expensive bundles appear much more frequently in the population. This shows that the services provided by these slightly more expensive bundles are more highly valued.

Table 1.3 also highlights the ability for the utility index to reconcile substantial temporal price changes. For example, due to decreasing prices, the monetary value of an iron, a stove, a fan, and a bicycle in 2005 is lower than the total cost of a bundle with a stove and a bicycle in 1998. Although the strictly larger bundle in the later time period is worth less money, the utility index captures the fact that the additional iron and fan provide a utility boost and therefore make households with the larger bundle in 2005 better off in terms of well-being.

I now demonstrate how the proposed utility index can be used by assessing how poverty has changed over time. Table 1.4 shows the poverty headcount over time using the utility index for various relative poverty lines. For example, the 10% poverty line for the utility index is defined as the index value below which 10% of the 1998 national sample falls. Comparing urban and rural areas, the headcounts show that poverty is highly rural across the distribution, with an estimated rural poverty rate that is four times as high as the corresponding urban rate at the lowest poverty line. The poverty trends show that poverty has been decreasing over time at all poverty lines with the exception of the 90% poverty line in urban areas from 1998 to 2001. These decreases have occurred in both urban and rural areas and are consistent with the summary statistics which show a general increase in asset ownership over the time period.

Table 1.5 presents the percent changes in poverty as well as the ratio of rural to urban poverty headcounts over the time period. The declines in poverty measured using the utility index are large, with national decreases of over 50% for the lowest two poverty lines. Large declines in urban poverty have occurred throughout the distribution, with declines exceeding 35% even at the 60% relative poverty line. Declines in rural poverty have been much more concentrated at the low end of the distribution, with decreases of 52 and 61% for the lowest two poverty lines but more modest changes (less than 12%) for the 40% relative poverty lines and higher. These differential changes in rural and urban areas have caused poverty to become more and more concentrated in rural areas over time, with rural poverty rates over six times as high as urban poverty rates under the 30% and 40% relative poverty lines.

Characterizing model performance

The model estimates household preferences in the absence of data on household wealth. To characterize the model's performance, I predict household asset stocks using data on household wealth available in the LSMS. I use the sum of annual consumption and the asset stock value as the measure of household wealth. I predict asset ownership under the assumption that household preferences are equal to the average marginal utility estimates, $\bar{\psi}^*$. Table 1.6 shows the prediction errors at the household level, by location and year.

In rural areas in 1998, predicted toaster, VHS, and car ownership matches true asset ownership for over 90% of households. Model error is higher for the less expensive assets, with accurate predictions for 50%, 55%, 49%, and 53% of households for iron, radio, fan, and blender ownership, respectively. Similar patterns are seen for asset ownership in rural areas in 2001 and 2005, with slightly lower accuracy in the later years.

Model accuracy in urban areas in 1998 tends to be slightly higher than accuracy in rural areas in 1998. Model accuracy is highest for car ownership (91%), and lowest for radio and blender ownership (55% and 55%). In urban areas in 2005, accuracy is particularly low for radios, VHS players, and sewing machines—assets with the largest average price changes over the time period, with decreases of 44%, 68%, and 37%, respectively. The assumption of constant average marginal utilities generates predictions that ownership for these items will increase substantially in response to decreasing prices. In contrast, ownership levels for radios, VHS players, and sewing machines remained relatively constant, indicating that the price decreases may have associated with simultaneous inward shifts in supply and demand.

The prediction results indicate that the model is able to predict asset ownership with moderate accuracy when data on household wealth is available. This demonstrates the power of employing a simple utility framework with relatively few parameters to explain household purchasing behaviors. At the same time, the prediction errors show that the model fails to capture all the important aspects of household asset ownership. The model tends to over-predict asset ownership for all years, regions, and assets, with the exception of cars. Calibration using observed wealth will be necessary if the model is to be used for predictions of asset ownership. In addition, the results indicate the likely existence of systematic differences in average preferences across regions and time, particularly for assets with substantial price differentials. This heterogeneity can be incorporated into an expanded model to allow for higher model accuracy.

Model validity without asset resale

In the theoretical model, the assumption that assets can be freely bought and sold each period was necessary to model household decision-making as a single period utility maximization problem. This assumption may or may not be reasonable in different contexts. If assets are sticky and households are not free to reoptimize their asset allocations each period, observed asset allocations may be sub-optimal solutions to the current period utility maximization problem. For example, suppose a household purchases a cheaper asset such as a radio in a prior period. With the additional income available to the household this period, it is optimal for the household to sell its radio and purchase a more expensive, higher marginal utility asset such as a television. Without asset resale, the household is unable to trade up; it keeps the radio and spends its additional income on consumption. A measure of well-being based solely on asset ownership would therefore underestimate the true level of current household utility. An asset measure can also overestimate well-being if households are unable to reallocate asset ownership towards consumption or less expensive assets in response to negative income shocks. Without asset resale, then, an asset index can result in the mis-rankings of household well-being.

To address this concern, I first present empirical evidence that this assumption holds in the context of Nicaragua and then use a simulation to demonstrate the validity of the utility index in contexts without asset resale. The survey instrument does not include an item on asset resale, but I identify the fraction of households which sold or discarded assets owned in earlier periods using the panel aspects of the data. If households are unable to sell their assets, I expect this fraction to be relatively low and due solely to asset disposal after substantial depreciation. The left panels of Table 1.7 show that the fraction of households which have sold or discarded assets owned in earlier periods is moderate to substantial. For example, approximately 40% of urban households which own bicycles in 1998 have sold or discarded them by 2001. To provide stronger evidence that households are actually reselling their assets and not simply discarding old assets, I repeat the analysis restricting the sample to new assets (less than 2 years old) under the rationale that households are more likely to resell newer assets instead of simply discarding them. These results are shown in the right panels of Table 1.7. Again, I find that the fraction of households which have sold or discarded relatively new assets is moderate to substantial. This is true across the range of items and in both urban and rural areas, indicating that there are at least partially functioning resale markets for all the goods in both urban and rural areas. This indicates that the asset resale assumption is reasonable in the context of Nicaragua. I note that measurement error could cause incorrect identification of asset divestiture. This could overestimate the amount of asset divestiture if there is a high rate of false negative ownership errors in later periods or false positive ownership errors in earlier periods. Alternatively, false negative errors in earlier periods and false positive errors in later periods will result in an underestimation of asset divestiture rates. Since there is little reason to suspect a systematic bias in the types of measurement errors, this is unlikely to affect the validity of the divestiture analysis.

In other contexts where second-hand markets are thin or nonexistent, the lack of asset resale may be a more substantial concern. To assess the validity of an asset index when assets cannot be freely sold, I run a simulation. In the simulation, households live for multiple periods. In each period, household i is endowed with income y_{it} and purchases assets according to the utility function specified above, conditional on prior asset ownership. For example, a household chooses to purchase a radio in the first period according to its first period income. In the second period, the household makes its purchasing decisions based on its second period income, ignoring the value of the previously purchased radio, and chooses from all bundles which contain a radio. I then compare the true current utility, which takes consumption and asset ownership into account, with the utility index based solely on asset ownership.

I run various simulations, which differ along four dimensions. The first dimension of variation is the total number of assets. A larger number of assets will provide greater discrimination between households in general. Without asset resale, a larger number of assets may partially alleviate the mis-rankings associated with not being to trade up. This will occur if households, which are unable to trade up, spend their additional resources this period on purchasing inexpensive assets as opposed to allocating the resources towards unobserved consumption. Thus, a household which already owns a radio and is unable to trade up to a television can purchase a fan instead and achieve an asset stock which more closely reflects it overall wealth, despite being suboptimal relative to a scenario with asset resale. I allow for three scenarios with six, ten, and fifteen assets ranging in price and marginal utility, reported in Table 1.8. I use prices and parameter estimates from the Nicaraguan data for the simulations, adding three additional fictitious assets for the fifteen asset simulation. I treat the marginal utility parameters as perfectly known and pass over estimation issues to isolate the concerns associated with the lack of asset resale. That is, I use the given parameters to simulate household ownership and to generate the utility index. In generating the data, I assume that the six, ten, or fifteen assets are the only assets available and that each household has identical preferences.

The second simulation dimension is intertemporal income variability. The asset resale assumption is irrelevant in the context of relative household rankings if all households live for the same number of periods and are endowed with the same income each period. Without asset resale, households may choose bundles that are suboptimal relative to the optimal bundles with asset resale, but this will affect all households similarly. That is, households which are unable to trade up to a television have consistently lower income and lower utility than households which own a television. Asset ownership will therefore correctly identify relative household rankings. This relationship breaks down when income is variable across periods; without asset resale, income shocks are imperfectly translated into asset stocks. To assess the importance of intertemporal income variability, I endow households with a base income $\bar{y}_i$ drawn from a lognormal distribution. Income for each period is drawn from a normal distribution with mean $\bar{y}_i$ and standard deviation $\sigma\bar{y}_i$, where σ is a fraction of mean income and takes on the values of 0, 0.1, 0.2, and 0.5. Higher values of σ capture contexts with higher intertemporal income variability.

The third and fourth dimensions are simulation length and intertemporal price trends. I generate simulations which last for two, five, and ten periods. In the five period simulation, 1000 households live for five periods, 1000 households live for four periods, 1000 households live for three periods, and so on. Even when incomes are constant over time, an asset index can incorrectly rank households of different ages. Again, an old household with a radio which is unable to trade up to a television can have higher utility than a new household with a television because of utility from consumption. Along the dimension of intertemporal price trends, I allow for two scenarios. In the first, asset prices are constant over time. In the second, asset prices decrease by 5% each period.

Table 1.9 shows the Pearson rank correlation coefficients, ρ, comparing the true current utility to the asset measure. As illustrated above, in a single period model, households will be perfectly ranked using information solely from asset ownership if preferences are homogeneous. (Due to tied rankings using an asset measure, the rank correlation coefficient will be less than 1.) The top panel shows the results when prices are held constant over time. In the five asset, two period simulation where household income is constant across periods ($\sigma = 0$), $\rho = 0.94$. The correlation coefficient remains high when the simulation is extended to five and ten periods ($\rho = 0.95$ and $\rho = 0.95$). When incomes are highly correlated over time, low income households will hold lower utility asset stocks compared to high income households regardless of asset stickiness. Holding the number of goods and periods constant, higher levels of intertemporal income variation are associated with lower rank correlation coefficients. When income varies over time ($\sigma \neq 0$), the five and ten period simulations are associated with lower rank correlation coefficients. The five asset, ten period simulation where $\sigma = 0.5$ is therefore lower, with $\rho = 0.80$.

The results for the simulations with more goods and decreasing prices show similar patterns, with lower correlation coefficients associated with higher intertemporal income variation and longer simulation lengths. Compared to the five good simulation, the correlation coefficients associated with the fifteen good simulations are higher, ranging from $\rho = 0.87$

to $\rho = 0.96$. Compared to the simulations with constant asset prices, the simulations with decreasing asset prices are associated with lower correlation coefficients.

Overall, the simulation results indicate that in the absence of asset resale, rankings generated using asset stocks remain highly correlated with current utility (accounting for both consumption and asset ownership). Although performance is lower in the presence of high intertemporal income variability and asset price changes, the correlation coefficients are still high, with values above 0.80 in the ten and fifteen good simulations. This implies that asset stocks remain reasonable approximations for current utility when the asset resale assumption is violated.

Model application: Decomposing poverty

In addition to generating an improved measure of household well-being, the model is valuable because it has predictive power. Using parameter estimates, I can make projections about how asset ownership evolves with price and income changes. These projections can be used to predict how poverty might respond to a particular program or to decompose observed changes in well-being into price and income effects. In this section, I demonstrate the predictive power of the model by assessing the extent to which price decreases in Nicaragua have contributed to improved well-being.

Specifically, I use the parameter estimates to predict the level of poverty in 2005 under the scenario where income and household size are constant at 1998 levels but prices decrease to 2005 levels. To do so, I use the measures of expected utility from 1998 to calculate the 1998 income distribution using the unconditional indirect utility functions to map utility to income. I then project asset ownership under the predicted 1998 income distribution, observed 1998 household sizes, and the decreased 2005 prices. I calculate the expected utility associated with the projected asset stocks and use this utility distribution to predict the new 2005 poverty headcount.

Table 1.10 shows the actual changes in poverty, the predicted changes in poverty associated with decreased prices, and the fraction of the actual poverty change that can be accounted for by changing prices. The price decreases from 1998 to 2005 are associated with increases in asset ownership and corresponding increases in utility. This is true in both urban and rural areas and at all levels of the distribution, with the exception of the poorest households. The projected income for the poorest households in 1998 was too low to allow for any asset purchases even under the lower 2005 prices, causing the predicted 2005 poverty headcount under the lowest poverty line to remain constant.

Excluding the poverty changes at the highest and lowest poverty lines, on average, the decrease in prices accounts for about a third of the decrease in poverty in both urban and rural areas. The additional decreases in poverty that are not accounted for by decreasing prices are likely due to increases in per capita income and reductions in credit constraints. This exercise demonstrates the benefits of using a model to characterize asset ownership using economic theory—it provides additional information on understanding the underlying causes of measured changes in well-being.

Performance of traditional asset indices

I turn now to an assessment of traditional asset indices. As discussed above, by drawing from economic theory, the utility index provides a measure that can be more directly interpreted as capturing household well-being and offers an improvement over existing asset indices on purely theoretical grounds. Empirically, it also meaningfully incorporates information on price, household demographics, and geographical variation that are not currently exploited by traditional indices. These improvements come at the cost of being computationally intensive, raising the question of whether or not traditional asset indices are sufficient approximations for well-being. Using the Nicaraguan data, I show that traditional asset indices are only moderate approximations for well-being, generating weights with little relationship to marginal utility. Furthermore, poverty and inequality measurements generated using traditional asset indices are only moderate approximations for findings from the more complete model with prices, demographics, and geographic variation.

Table 1.11 compares the asset weights derived using the utility index, the inverse frequency method, a standard factor analysis index, and an expanded factor analysis index. The weights have been rescaled to have a maximum value of one for the purposes of comparison. The inverse frequency method generates weights based on the inverse of the population ownership levels, while the factor analysis index generates weights based on the variance structure of the underlying data. The 2001 and 2005 inverse frequency and factor analysis indices are constructed using weights from the 1998 sample. To the extent that the weights represent marginal utilities, the use of the 1998 weights isolates the effect of changes in asset ownership under the assumption that the marginal utilities have remained constant. The weights for the traditional asset indices are generated using the national sample, as the methods do not allow for urban/rural variation. I construct the first factor analysis index using data on solely asset ownership, as is standard. In addition, I construct a second factor analysis index incorporating household size and urban/rural location. Although it is not clear from theory whether or not the inclusion of such additional information is appropriate in a statistically generated asset index, I include this index to provide a comparison for the utility index which does incorporate additional information on demographics and location.

The weights generated using the various methods differ widely and are only moderately related to the estimated marginal utilities. The inverse frequency index places the highest weight on VHS players, which are the least common items in the population. Compared to the utility model, the inverse frequency index places less weight on items such as televisions which are commonly owned despite the relatively high price and places more weight on items which are relatively cheap but less common (such as toasters). The standard factor analysis method on this sample results in low weights for items such as cars, bicycles, and sewing machines and high weights for items such as blenders, stoves, televisions, and fans. Because of the lack of theory associated with factor analysis, it is difficult to determine whether the method creates weights that differ from the utility model in a systematic manner. Compared to the utility index, factor analysis results in lower variation in the magnitude of weights since it does not incorporate information on price variation. The corresponding weights in the expanded factor analysis index are similar to the weights derived using solely asset ownership. Household size is associated with a negative weight, while being in an urban area is associated with a positive weight.

One particular discrepancy of interest involves the factor analysis weight associated with radios, which is negative. Within the context of Nicaragua, radio ownership decreases with household consumption, indicating that radios may be inferior goods at higher levels of income. This may be due to the fact that households substitute towards televisions when they can afford to do so. If it is the case that there are important substitutes for radios other than televisions that are not included in the model, the utility index may result in inaccuracies in classifying households at the higher end of the wealth distribution.[10] Although radios may be inferior goods at higher levels of income and could be excluded from the analysis, radio ownership provides important discrimination between households at the low end of the wealth distribution. In this case, the negative factor analysis weight on radios proves problematic for classifying households at the lower end of the distribution, as households which own radios are ranked lower than households with no goods according to the factor analysis index. Thus, although the inclusion of radios may be problematic for both the utility asset index and the factor analysis index, the negative factor analysis weight is arguably more concerning given the use of asset indices in identifying the poorest households.

Although it is clear that the weights associated with traditional methods are only weak approximations for marginal utility, it is possible that the resulting indices may still generate similar conclusions about the ultimate measures of interest. I show that this is not the case—the poverty and inequality statistics generated using traditional asset indices are only moderate approximations for the findings which directly account for household preferences. Table 1.12 shows the poverty changes and rural/urban poverty headcount ratios using the set of traditional indices explored above.

Compared to the utility index results shown in Table 1.5, the inverse frequency index and standard factor analysis index also show that poverty has been decreasing across the distribution and in rural and urban areas. There are substantial discrepancies in the magnitudes of the changes, though. The utility index, which accounts for prices and household preferences, shows a 56% decrease in poverty for the national sample over the time period using the 10% relative poverty line. In contrast, the inverse frequency and standard factor analysis indices show decreases of 63% and 22%, respectively. Moreover, the expanded factor analysis index incorrectly shows that poverty has increased by 11% due to the negative weights associated with radios and the large increase in radio ownership. Comparing across the various poverty lines, the discrepancies in poverty changes continue to be relatively large for the national, urban, and rural samples at the lower poverty lines, indicating again that the traditional asset indices are only moderate approximations for a more extensive model of household wealth.

[10] Questions on CD player ownership are included in the 2005 survey but are excluded from the earlier rounds. Ownership of CD players is under 4% in 2005 and unlikely to fully account for the decrease in radio ownership at higher wealth levels.

As an aside, because urban households receive a positive weight under the expanded factor analysis index, the urban poverty headcount associated with the expanded factor analysis index is 0 under the 10% and 20% relative poverty lines. The respective changes in urban poverty are therefore undefined.

The bottom panel of Table 1.12 shows the rural/urban poverty headcount ratios using the traditional indices. Although the traditional indices also show that poverty is highly rural, the measured levels of inequality differ from the findings using the utility index. For example, the utility index shows that under the 10% 1998 relative poverty line, rural poverty in 2005 is 5.24 times as high as urban poverty, while the ratios found using the inverse frequency and standard factor analysis indices are 3.82 and 9.10, respectively. The factor analysis bias against rural households is due in part to the large increase in radio ownership which occurred disproportionately in rural areas. Again, the rural/urban headcount ratios are undefined at the lower poverty lines using the expanded factor analysis index due to the zero urban poverty headcount.

Finally, asset indices are also used for proxy-means targeting, to identify the poorest households for the purpose of targeting program benefits. I demonstrate that the traditional asset indices often classify households into groups which differ substantially from the more theoretically valid measure of well-being captured by the utility index. When comparing the resulting ranks associated with the different methods, the correlations are high, with rank correlation coefficients of 0.95, 0.91, and 0.91 between the utiliy index and the inverse frequency index, the standard factor analysis index and the expanded factor analysis index, respectively. This statistic masks variations in rankings which are important in classifying households into different groups.

Table 1.13 shows how well the deciles groupings created using the different asset indices correspond to those created using the utility model. Between 34 and 45% of households are categorized in the matching utility index decile using the traditional asset indices, and similar fractions of the sample (40-48%) are classified in deciles which differ from the corresponding utility index groupings by one decile. Discrepancies of two or more deciles compared to the utility index groupings exist for over 10% of the sample for all of the traditional indices. This shows that targeting errors are potentially very large when using methods without a clear economic meaning.

Together, the results indicate that traditional asset indices are not only unable to capture household preference parameters of interest, they are also somewhat weak proxies for household well-being when considering practical applications of such indices.

Distinctions between asset indices and consumption measures

Finally, I discuss and demonstrate the distinctions between asset indices and standard consumption measures. Although many asset indices originally arose as cheaper alternatives to consumption expenditure measures (Filmer and Pritchett 2001), the two measures differ in the temporal aspects of well-being that they capture. The reasons for these differences, which I discuss below, suggest that consumption aggregates may not be the appropriate benchmark for asset indices as substantive discrepancies will exist between the measures even when they are perfectly constructed.

The primary difference between consumption aggregates and asset measures is due to the fact that consumption is a flow which occurs each period while assets are a stock generally accrued over multiple periods. Carefully constructed consumption measures include the value of durables purchased this period and often include an imputed measure of the value of services (such as housing) received this period. They do not commonly include the total value of assets brought into the period. In this way, consumption aggregates capture a measure of contemporaneous well-being. In contrast, asset measures are constructed using assets which households have purchased both in this period and in all previous periods. They therefore represent accrued wealth, a longer term sense of economic well-being. The discrepancy between consumption and discrete good ownership may be a measure of current shocks to households and has been used to determine whether local government targeting can better identify households facing negative shocks (Alderman 2002). In this way, the consumption aggregates and asset indices capture slightly different temporal dimensions of well-being and should be discussed with these differences in mind.[11]

I demonstrate these differences empirically by comparing the poverty inferences found using per capita consumption with the analogous results using the utility index. Household per capita consumption measures are constructed using expenditure data on food, non-durables, durables, and housing. Non-durables expenditures include expenditures on household supplies, transportation, communication, clothing, and education. Expenditure on durables captures expenditure on housing improvements and purchases of furniture, appliances, and vehicles this period. Housing expenditure is given by reported rental values or estimated rental values if the housing is owned. As with the values of owned durable goods, consumption values are deflated into 1998 Nicaraguan cordobas using household consumption levels from the World Development Indicators. Per capita adjustments are done using raw household sizes where each child is treated as one adult equivalent. Table 1.14 shows annual per capita consumption levels by category and in aggregate. Nicaraguan households spend almost two-thirds of their yearly expenditure on food, slightly less than a third on non-durable purchases, and relatively small fractions on durables and housing. Although average household expenditure showed a small increase from 1998 to 2001, the period from 2001 to 2005 was associated with a decrease in average household expenditure. This trend is seen in all categories except for housing, which saw an increase in spending across both time periods.

Using the three rounds of data, I first assess the correlation between the consumption aggregate and the utility index. The Spearman rank correlation coefficients are presented in

[11] The distinctions between asset indices and consumption aggregates have been demonstrated in existing empirical work. While some studies find that asset indices correspond reasonably to per capita expenditure (Ferguson et al. 2003; Filmer and Pritchett 2001; Montgomery and Hewett 2005; Morris et al. 2000) recent reviews caution that asset-based measures are relatively weak proxies for consumption expenditure (Howe et al. 2009) and are more strongly associated with particular expenditure subcomponents such as non-food items (Filmer and Scott 2012).

the last row of Table 1.15. Per capita consumption is moderately correlated with the utility index, with rank correlation coefficients ranging from 0.54 to 0.59. The correlations within urban and rural areas are lower, ranging from 0.52 to 0.53 in urban areas and from 0.37 to 0.40 in rural areas.

Although there is moderate correspondence between the two measures within each survey, the trends in poverty differ drastically. The poverty headcounts associated with per capita expenditure are shown in the top of Table 1.15. As with the asset indices, the poverty lines are determined relative to the 1998 population distribution. The gains in household consumption from 1998 to 2001 benefitted households at the moderate to high end of the consumption distribution, with slight decreases in the poverty headcount for poverty lines at 40% and above. Slight increases in poverty headcounts are seen at the lowest poverty lines. From 2001 to 2005, poverty headcounts increased across the distribution with increases of 2 to 5.5 percentage points above 1998 levels depending on the poverty line.[12]

Partitioning the sample by urban and rural areas, both the consumption and utility measures show that poverty is highly rural. The utility index shows higher rates of rural poverty in 1998 compared to the per capita consumption measure for all poverty lines except for the 20% line. The consumption poverty trends within the rural population show a consistent increase in poverty across the time period, while poverty decreased moderately in urban areas from 1998 to 2001 before returning to 1998 levels in 2005. The consumption poverty trends stand in stark contrast to the changes seen using asset indices. Although per capita consumption was stagnant and then decreasing over the time period, there were large improvements in household well-being according to asset ownership.

Although theory does not preclude consumption and asset measures from being more similar, these findings underscore the important differences between asset and expenditure measures. It also demonstrates how the use of the two measures in conjunction can generate a more complete understanding of changes in different temporal dimensions of well-being.

1.5 Discussion

This paper demonstrates how asset information can be used to create a measure of household well-being in a manner that is consistent with household demand for assets. The introduction of a theoretical model marks one important step in developing an index which can be meaningfully interpreted, but substantive issues remain.

One outstanding issue is the choice of assets to include in an index. This analysis focused on durable goods such as electronics, vehicles, and other appliances but failed to incorporate

[12]It is important to note that consumption-based poverty headcounts are sensitive to the categories included into total consumption as well as the use of different price deflators. The estimates presented are based on consumption aggregates constructed by the author. Trends in poverty provided using the preconstructed LSMS consumption aggregates and poverty headcounts are relatively similar. Extreme poverty rates changed from 12.2% to 11.0% to 12.3%, while general poverty rates changed from 39.0% to 37.0% 39.2%.

any information on financial assets which are increasing in importance and availability in developing country contexts. In addition, assets such as land, livestock, and housing may be important for differentiating between households in developing countries. Information on these assets is included in many surveys, but their inclusion in an asset index will require further work to address the additional intricacies associated with durable assets which I have not explored in this paper.

As mentioned in the background section, market imperfections introduce a layer of complexity to the issue of estimating demand for durable assets. In addition to missing resale markets, there are market issues associated with asset prices. Current asset values may be difficult to measure when households have minimal experience with resale. Price measurement for demand estimation is also complicated by the issue of endogeneity. Although it is reasonable to assume that prices are exogenous to a given household, aggregate demand affects equilibrium prices. The bias on the resulting marginal utility estimates is unclear. This can be addressed in future work by using instruments for price which capture supply-side factors. Furthermore, prices are not a good indicator of the true market value of certain assets. For instance, assets such as land are subject to complex issues such as inheritance and regulations. Again, this may lead to biased marginal utility estimates.

Other market complexities include the existence of asset distribution programs and dynamics associated with information diffusion. Many governments and non-governmental organizations have programs which subsidize and distribute assets such as toilets, televisions, and housing improvements. This will lead to incorrect marginal utility estimates since free or highly subsidized distribution causes assets to appear more highly demanded than they truly are. Conversely, recent introduction of assets may result in underestimation of the marginal utilities associated with new assets. This will occur when households are not fully aware of the features of new assets and therefore choose not to purchase them even though they have a high value for the services provided by the new assets. Researchers measuring well-being in either of these scenarios should consider these issues when applying any asset indices.

In addition to these general issues, I discuss the specification limitations associated with my model. I parameterize utility using certain simplifying assumptions which are arguably restrictive. First, I assume that there are no complementaries between different assets. This assumption can be relaxed by specifying marginal utilities for each asset which enter in multiplicatively. Complementarities can also be allowed by estimating the marginal utility associated with each bundle as opposed to each asset. Second, I assume that the marginal utilities associated with each asset are independent of household size. Though this assumption may be reasonable for most of the included assets, it may not be reasonable for all. This can be addressed by allowing the marginal utilities of each asset to vary with household size.[13] Third, I assume that average household preferences are constant across urban

[13]These proposed modifications have not been implemented at this time due to limits in computational capacity. The maximum likelihood estimation is particularly computationally intensive due to the large set (2^J) of available bundles.

and rural areas and over time. Additional work is necessary to allow for heterogeneous preferences.

Despite the limitations of the model, the introduction of a theoretical model continues to offer an improvement over the traditional asset indices. Compared to the traditional methods, the assumptions of the model are explicit and can therefore be addressed and changed as appropriate. With the model, the "weight" (marginal utility) of each good can be interpreted as the relative benefit afforded to households by the good, and the utility index can be interpreted as the expected utility of each household conditional on prices, bundle ownership, and household size. The introduction of a theoretical model also allows researchers to simulate how well-being would change with changes in income, price, and household demographics. Such simulations can help to identify the underlying causes of observed changes in well-being and are not possible without a well-formed theoretical model.

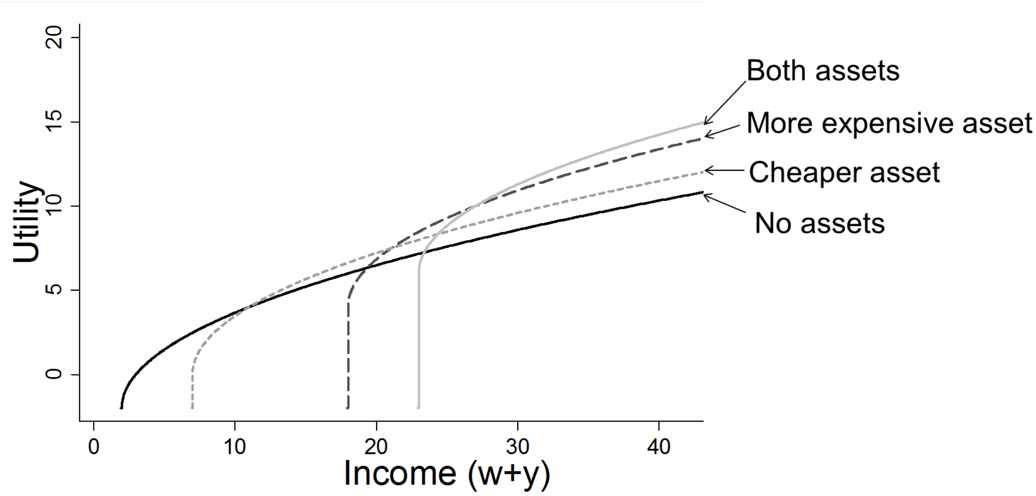
The empirical analyses in this paper which compare the traditional asset indices to the utility asset index indicate that household rankings do indeed depend on the method used. The discrepancies in quantiles within a survey and trends across surveys underscore the need for caution in using traditional indices, particularly for use in proxy-means targeting and poverty monitoring. The analogous comparisons with per capita expenditure again highlight the need for caution when using asset indices as proxies for consumption expenditure. Because asset ownership and expenditure capture different temporal concepts of well-being, they may only be weakly related and may result in substantive differences in poverty assessments, as shown in the Nicaraguan case.

Additional implications for researchers include the importance of collecting asset price information for durable goods. Basic economics suggests that prices are critical pieces of information in determining household purchases, and the empirical estimates show that prices are strong indicators for the utility benefits associated with a good. Thus, meaningful asset indices should incorporate such information. Although prices are currently not collected in instruments such as the DHS, the marginal costs of collecting this additional data should be relatively low given a concise set of relevant durable goods and the potential to collect the information at the community level.

In conclusion, this paper demontrates the potential for a theoretically driven asset index to be used as a tool for monitoring poverty and suggests some avenues for continued work in improving measures of well-being.

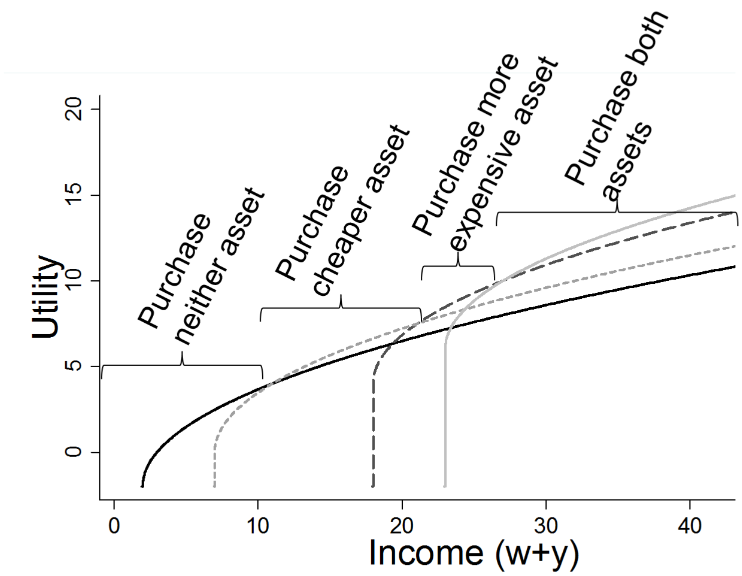
1.6 Figures and Tables

Figure 1.1: Model illustration: Indirect utility functions for two discrete assets



The solutions are shown in the case where the more expensive asset provides more marginal utility, but this is not required to be the case in the model.

Figure 1.2: Model illustration: Optimal bundles as a function of income



The solutions are shown in the case where the more expensive asset provides more marginal utility, but this is not required to be the case in the model.

Table 1.1: Summary statistics: asset ownership

	Real price 1998 cordobas (Pooled average)	Population ownership level					
		Urban areas			Rural areas		
		1998	2001	2005	1998	2001	2005
Iron	93	0.80	0.81	0.84	0.38	0.35	0.33
Radio	100	0.29	0.28	0.29	0.34	0.50	0.62
Fan	113	0.49	0.49	0.55	0.11	0.11	0.12
Blender	166	0.40	0.39	0.46	0.06	0.07	0.09
Toaster	178	0.08	0.09	0.08	0.01	0.00	0.01
Bike	392	0.30	0.35	0.37	0.19	0.27	0.29
Sewing machine	626	0.18	0.14	0.17	0.08	0.07	0.08
Stove	686	0.52	0.57	0.68	0.10	0.11	0.11
VHS	968	0.07	0.09	0.07	0.01	0.01	0.01
TV	1166	0.75	0.77	0.86	0.29	0.31	0.36
Fridge	2288	0.29	0.30	0.40	0.07	0.06	0.09
Car	38196	0.08	0.08	0.10	0.02	0.02	0.02

Notes: Average prices are calculated using pooled data from 1998, 2001, and 2005 national samples. Nationally representative sampling weights are applied.

Table 1.2: Utility model estimation and spatial and geographical price variation

	Population ownership level (Pooled)	Real price: 1998 cordobas						Marginal utility estimates 1998 cordobas	
		Urban areas			Rural areas				
		1998	2001	2005	1998	2001	2005	$\bar{\psi}^*$	σ^*
Iron	0.62	105	100	90	79	82	73	132	64
Radio	0.37	137	106	76	107	108	80	106	12
Fan	0.34	124	118	102	122	108	96	132	18
Blender	0.27	186	177	140	177	176	162	184	31
Toaster	0.05	204	191	138	211	177	165	68	19
Bike	0.31	471	382	371	429	391	350	365	53
Sewing machine	0.13	817	631	514	705	533	517	505	145
Stove	0.38	801	764	562	579	758	531	716	146
VHS	0.05	1408	1037	438	1916	1427	646	403	46
TV	0.59	1286	1386	1190	680	742	720	2346	1571
Fridge	0.22	2868	2399	1886	2456	2252	1877	2887	1567
Car	0.06	35781	49767	35760	30331	30447	26563	5276	2357
Subsistence ϕ^*								172	

Notes: Prices are calculated as the average value of the assets as reported by households. Sample statistics and model estimation are done using nationally representative sampling weights. Average ownership levels are calculated using pooled data from 1998, 2001, and 2005 national samples.

Table 1.3: Utility index rankings and comparisons with price rankings: Selected bundles

Bundle	Own. Level	Price rank 1998	Price rank 2001	Price rank 2005	Average utility index rank 1998	Average utility index rank 2001	Average utility index rank 2005
Iron	0.0308	141	131	124	109	86	101
Fan	0.0028	194	189	137	127	122	141
Radio	0.1258	195	146	79	168	148	184
Iron, Radio	0.0274	278	258	210	249	243	251
Bike	0.0150	393	326	324	344	338	347
Iron, Fan, Radio, Bike	0.0007	746	633	567	593	592	592
Stove	0.0025	715	683	481	634	637	640
Stove, Radio	0.0013	819	770	565	688	686	688
Iron, Bike, Sewing machine	0.0015	1234	936	853	766	761	760
Stove, Bike	0.0013	1097	975	817	823	818	820
Iron, Stove, Fan, Bike	0.0007	1330	1185	944	861	859	856
Iron, Stove, Fan, Bike, Sew.	0.0002	2063	1812	1442	911	911	909
TV	0.0166	1112	1209	1019	944	945	948
Iron, TV, Fan, Blender	0.0030	1503	1597	1352	1250	1248	1250
Fridge	0.0003	2491	2152	1698	1446	1426	1429
Iron, TV, Radio, Bike	0.0114	1820	1778	1534	1509	1494	1499

Notes: Average ownership levels are calculated using pooled data from 1998, 2001, and 2005 national samples. Price ranks are calculated using the total monetary value of bundles in 1998 cordobas. Utility index rankings are averaged over national samples; variation by urban/rural location and household size is not shown.

Table 1.4: Poverty headcount using utility index

	Poverty headcount								
1998 Relative poverty line	National sample			Urban areas			Rural areas		
	1998	2001	2005	1998	2001	2005	1998	2001	2005
10%	0.09	0.06	0.04	0.04	0.03	0.02	0.17	0.12	0.08
20%	0.20	0.11	0.07	0.10	0.06	0.04	0.33	0.20	0.13
30%	0.30	0.25	0.19	0.14	0.11	0.06	0.51	0.47	0.37
40%	0.40	0.35	0.30	0.20	0.18	0.10	0.66	0.64	0.59
50%	0.50	0.46	0.38	0.30	0.29	0.17	0.76	0.74	0.68
60%	0.60	0.56	0.48	0.41	0.38	0.26	0.84	0.83	0.79
70%	0.70	0.66	0.60	0.55	0.52	0.41	0.90	0.89	0.86
80%	0.80	0.79	0.72	0.70	0.70	0.59	0.93	0.93	0.91
90%	0.90	0.90	0.89	0.84	0.86	0.83	0.97	0.97	0.97

Notes: The X% 1998 relative poverty line is determined by the index threshold classifying the bottom X% of the 1998 national sample as poor.

Table 1.5: Changes in poverty and rural/urban poverty headcount ratios: Utility index

1998 Relative poverty line	% change in poverty 1998-2005			Rural/urban poverty headcount ratio		
	National Sample	Urban Areas	Rural Areas	1998	2001	2005
10%	-0.56	-0.62	-0.52	4.14	4.52	5.24
20%	-0.63	-0.65	-0.61	3.25	3.12	3.59
30%	-0.37	-0.59	-0.27	3.53	4.13	6.36
40%	-0.24	-0.51	-0.11	3.33	3.60	6.09
50%	-0.23	-0.43	-0.10	2.51	2.58	3.96
60%	-0.20	-0.37	-0.07	2.03	2.17	3.03
70%	-0.15	-0.25	-0.05	1.66	1.71	2.11
80%	-0.10	-0.16	-0.03	1.34	1.34	1.55
90%	-0.01	-0.01	-0.01	1.15	1.13	1.16

Notes: The X% 1998 relative poverty line is determined by the index threshold classifying the bottom X% of the 1998 national sample as poor.

Table 1.6: Model performance: Asset ownership predictions–Fraction of sample

	Urban areas								
	1998			2001			2005		
	$x_{predicted} - x =$			$x_{predicted} - x =$			$x_{predicted} - x =$		
Asset name	1	0	-1	1	0	-1	1	0	-1
Iron	0.15	0.75	0.07	0.17	0.80	0.03	0.13	0.83	0.04
Radio	0.29	0.55	0.14	0.47	0.41	0.12	0.61	0.34	0.05
Fan	0.30	0.59	0.09	0.39	0.56	0.05	0.34	0.61	0.05
Blender	0.37	0.55	0.06	0.43	0.52	0.05	0.43	0.54	0.02
Toaster	0.16	0.78	0.03	0.28	0.70	0.02	0.50	0.49	0.01
Bike	0.27	0.58	0.13	0.43	0.46	0.11	0.40	0.49	0.11
Sew. mach.	0.23	0.66	0.09	0.36	0.59	0.05	0.51	0.45	0.03
Stove	0.23	0.64	0.11	0.22	0.69	0.09	0.24	0.71	0.05
VHS	0.13	0.82	0.02	0.26	0.72	0.02	0.55	0.44	0.01
TV	0.11	0.76	0.11	0.14	0.80	0.06	0.10	0.87	0.03
Fridge	0.18	0.72	0.07	0.30	0.65	0.04	0.39	0.58	0.03
Car	0.00	0.91	0.06	0.00	0.94	0.06	0.00	0.93	0.06
	Rural areas								
	1998			2001			2005		
	$x_{predicted} - x =$			$x_{predicted} - x =$			$x_{predicted} - x =$		
Asset name	1	0	-1	1	0	-1	1	0	-1
Iron	0.44	0.50	0.06	0.52	0.45	0.03	0.56	0.43	0.01
Radio	0.23	0.55	0.21	0.21	0.48	0.31	0.21	0.49	0.30
Fan	0.46	0.49	0.04	0.65	0.35	0.01	0.63	0.36	0.01
Blender	0.44	0.53	0.02	0.48	0.51	0.01	0.48	0.51	0.01
Toaster	0.08	0.91	0.00	0.11	0.89	0.00	0.21	0.79	0.00
Bike	0.16	0.72	0.11	0.24	0.62	0.14	0.25	0.63	0.12
Sew. mach.	0.14	0.80	0.05	0.30	0.67	0.03	0.30	0.66	0.04
Stove	0.39	0.58	0.02	0.26	0.71	0.04	0.56	0.43	0.01
VHS	0.04	0.95	0.00	0.07	0.93	0.00	0.21	0.79	0.00
TV	0.39	0.57	0.03	0.43	0.54	0.03	0.47	0.51	0.02
Fridge	0.16	0.81	0.02	0.24	0.74	0.02	0.35	0.64	0.01
Car	0.00	0.97	0.02	0.00	0.98	0.02	0.00	0.98	0.02

Notes: Predictions done using model estimates and household current wealth, measured as sum of asset stock values and annual consumption.

Table 1.7: Flexibility of asset stocks: Asset divestitures (panel households)

	Fraction of households selling/discarding item over time							
	Durables of any age				Newer durables (< 2 years old)			
	Urban areas		Rural areas		Urban areas		Rural areas	
	'98-'01	'01-'05	'98-'01	'01-'05	'98-'01	'01-'05	'98-'01	'01-'05
Iron	0.09	0.08	0.43	0.34	0.07	0.08	0.38	0.31
Radio	0.64	0.56	0.34	0.22	0.60	0.52	0.31	0.21
Fan	0.28	0.25	0.42	0.39	0.26	0.26	0.37	0.47
Blender	0.29	0.25	0.45	0.33	0.32	0.25	0.45	0.36
Toaster	0.62	0.73	0.75	0.40	0.67	0.65	1.00	0.00
Bike	0.40	0.39	0.29	0.31	0.40	0.41	0.32	0.34
Sew. Mach.	0.53	0.55	0.63	0.55	0.50	0.50	0.85	0.90
Stove	0.11	0.10	0.33	0.37	0.15	0.15	0.37	0.39
VHS	0.42	0.79	0.67	0.75	0.50	0.78	1.00	1.00
TV	0.08	0.05	0.18	0.16	0.07	0.04	0.15	0.23
Fridge	0.32	0.23	0.46	0.34	0.30	0.24	0.43	0.38
Car	0.57	0.48	0.41	0.44	0.63	0.78	0.60	0.67

Notes: The denominator of each column is the sample of households in urban/rural areas which report item ownership in the earlier wave. Degenerate rates of divestiture for newer toasters and VHS players in rural areas is due to very low base ownership.

Table 1.8: Model validity without asset resale: Simulation parameters

	Price	Marginal utility (ψ)	6 assets	10 assets	15 assets
Asset 1	50	52			X
Asset 2	79	82	X	X	X
Asset 3	90	95			X
Asset 4	107	104	X	X	X
Asset 5	122	95	X	X	X
Asset 6	150	160			X
Asset 7	177	142		X	X
Asset 8	211	62		X	X
Asset 9	429	368	X	X	X
Asset 10	579	724	X	X	X
Asset 11	680	2440	X	X	X
Asset 12	705	512		X	X
Asset 13	1916	406		X	X
Asset 14	2456	2935			X
Asset 15	30331	9358			X
ϕ=164					
Base income distribution$y_t \sim \ln N(0, 3769)$					

Notes: The prices and marginal utility parameters are taken from the empirical model estimates for Nicaragua, with the exception of the assets priced at 52, 95, and 150.

Table 1.9: Model validity without asset resale: Simulation results

	Income variation (sd)	Rank correlation coefficient: True utility versus expected utility based on asset ownership								
		5 assets			10 assets			15 assets		
		No. of periods			No. of periods			No. of periods		
		2	5	10	2	5	10	2	5	10
Constant asset prices	0.00	0.94	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.96
	0.10	0.95	0.94	0.94	0.95	0.95	0.95	0.96	0.96	0.96
	0.20	0.94	0.93	0.92	0.95	0.95	0.94	0.94	0.96	0.95
	0.50	0.90	0.84	0.80	0.91	0.87	0.83	0.93	0.89	0.87
Decreasing asset prices (-5% per period)	0.00	0.94	0.94	0.91	0.96	0.96	0.94	0.96	0.96	0.95
	0.10	0.94	0.94	0.90	0.95	0.96	0.94	0.96	0.96	0.95
	0.20	0.94	0.92	0.88	0.95	0.95	0.93	0.95	0.96	0.94
	0.50	0.89	0.83	0.75	0.91	0.87	0.81	0.92	0.89	0.85

Notes: The two period simulation allows half the households to live for one period and half the households to live for two periods; the five period simulation allows one fifth of the households to live for one period, one fifth of the households to live for two periods, etc. Households are endowed with a mean income; income each period is drawn from a normal distribution with a mean equal to the mean household endowment and a standard deviation equal to a fraction of the mean endowment.

Table 1.10: Predicted poverty changes associated with decreasing prices

	Poverty changes from 1998-2005					
	Urban areas			Rural areas		
1998 Relative poverty line	Actual change in poverty	Predicted change using lower prices	Estimated fraction of change due to lower prices	Actual change in poverty	Predicted change using lower prices	Estimated fraction of change due to lower prices
10%	-0.025	0.000	0.00	-0.086	0.000	0.00
20%	-0.065	-0.010	0.15	-0.200	-0.030	0.15
30%	-0.085	-0.030	0.35	-0.135	-0.161	1.19
40%	-0.102	-0.012	0.12	-0.071	-0.034	0.48
50%	-0.130	-0.070	0.54	-0.078	-0.016	0.21
60%	-0.154	-0.041	0.26	-0.056	-0.010	0.17
70%	-0.138	-0.107	0.77	-0.043	-0.001	0.01
80%	-0.112	-0.041	0.37	-0.026	-0.003	0.11
90%	-0.011	-0.139	12.13	-0.008	-0.015	1.97

Notes: The X% 1998 relative poverty line is determined by the index threshold classifying the bottom X% of the 1998 national sample as poor. Poverty predictions for 2005 are generated by predicting asset ownership using the inferred income distribution from 1998, observed household size in 1998, and 2005 prices.

Table 1.11: Traditional asset index weights: Comparisons with utility index

			Scaled Weights			
	Population ownership level (pooled)	Avg. price (pooled) 1998 cordobas	Utility index	Inverse frequency index	Factor analysis index	Factor analysis index (expanded)
Iron	0.62	93	0.01	0.07	0.84	0.87
Radio	0.37	100	0.01	0.14	-0.05	-0.06
Fan	0.34	113	0.01	0.13	0.93	0.94
Blender	0.27	166	0.02	0.17	1.00	1.00
Toaster	0.05	178	0.01	0.85	0.56	0.55
Bike	0.31	392	0.04	0.17	0.49	0.49
Sewing machine	0.13	626	0.05	0.33	0.55	0.54
Stove	0.38	686	0.08	0.13	0.98	1.00
VHS	0.05	968	0.04	1.00	0.57	0.56
TV	0.59	1166	0.26	0.08	0.93	0.95
Fridge	0.22	2288	0.31	0.23	0.89	0.88
Car	0.06	38196	1.00	0.79	0.53	0.52
Household size						-0.09
Urban						0.74

Notes: Parameters (weights) from each method are scaled to give a maximum of 1. The inverse frequency method uses the inverse of population ownership levels as weights. Factor analysis uses the variance structure of the data to derive weights statistically. Weights for the inverse frequency and factor analysis indices are calculated using the 1998 national sample and then applied to the 2001 and 2005 samples.

Table 1.12: Changes in poverty and rural/urban headcount ratios: Traditional asset indices

	% change in poverty: 1998-2005								
1998 relative poverty line	Inverse frequency index			Factor analysis index			Expanded factor analysis index		
	Full	Urban	Rural	Full	Urban	Rural	Full	Urban	Rural
10%	-0.63	-0.65	-0.60	-0.22	-0.58	-0.07	0.11	NaN	0.15
20%	-0.63	-0.62	-0.63	-0.22	-0.58	-0.07	-0.10	NaN	-0.07
30%	-0.28	-0.53	-0.15	-0.16	-0.52	-0.01	-0.15	-0.58	0.00
40%	-0.29	-0.49	-0.17	-0.18	-0.45	-0.04	-0.20	-0.55	-0.05
50%	-0.24	-0.35	-0.15	-0.18	-0.37	-0.05	-0.16	-0.38	-0.04
60%	-0.18	-0.27	-0.09	-0.16	-0.32	-0.04	-0.14	-0.31	-0.02
70%	-0.14	-0.20	-0.06	-0.12	-0.22	-0.03	-0.13	-0.23	-0.03
80%	-0.09	-0.12	-0.03	-0.09	-0.13	-0.02	-0.08	-0.14	-0.01
90%	-0.02	-0.03	-0.01	-0.03	-0.04	0.00	-0.03	-0.04	0.00

	Rural/urban poverty headcount ratio								
1998 relative poverty line	Inverse frequency index			Factor analysis index			Expanded factor analysis index		
	1998	2001	2005	1998	2001	2005	1998	2001	2005
10%	3.40	3.77	3.82	4.10	5.41	9.10	NaN	NaN	NaN
20%	3.02	2.76	2.93	4.10	5.41	9.10	NaN	NaN	NaN
30%	3.38	3.85	6.13	3.95	4.85	8.21	4.71	6.54	11.20
40%	2.70	2.89	4.37	3.38	3.81	5.85	4.20	4.86	8.79
50%	2.23	2.26	2.90	2.64	2.73	3.99	3.54	3.67	5.46
60%	1.98	1.89	2.47	2.15	2.26	3.02	2.37	2.46	3.37
70%	1.65	1.66	1.96	1.68	1.69	2.08	1.82	1.83	2.31
80%	1.38	1.39	1.53	1.43	1.44	1.61	1.45	1.48	1.66
90%	1.17	1.18	1.20	1.16	1.17	1.21	1.18	1.19	1.23

Notes: The X% 1998 relative poverty line is determined by the threshold of the respective index classifying the bottom X% of the 1998 national sample as poor. The inverse frequency method uses the inverse of population ownership levels as weights. Factor analysis uses the variance structure of the data to derive weights statistically. The expanded factor analysis index includes information on household size and urban/rural location. Weights for the inverse frequency and factor analysis indices are calculated using the 1998 national sample and then applied to the 2001 and 2005 samples. The NaN results using the expanded factor analysis index are due to a 0% poverty headcount in urban areas.

Table 1.13: Decile comparisons with utility index classification

	Fraction of population		
Decile difference	Inverse frequency index	Factor analysis index	Expanded factor analysis index
Same decile	0.433	0.344	0.369
1 decile difference	0.421	0.483	0.456
2 decile difference	0.101	0.156	0.158
3 decile difference	0.035	0.008	0.007
4 decile difference	0.004	0.004	0.004
5 decile difference	0.001	0.001	0.001
6 decile difference	0.000	0.001	0.001

Notes: Decile groupings classify the bottom 10% of the sample according to the respective index in the lowest decile, the next 10% of the sample in the second decile, and so on. The sample is pooled across the 1998, 2001, and 2005 national data. The inverse frequency method uses the inverse of population ownership levels as weights. Factor analysis uses the variance structure of the data to derive weights statistically. The expanded factor analysis index includes information on household size and urban/rural location. Weights for the inverse frequency and factor analysis indices are calculated using the 1998 national sample and then applied to the 2001 and 2005 samples.

Table 1.14: Annual household per capita consumption expenditure (1998 cordobas)

Year	1998	2001	2005
Total expenditure	6,175	6,203	5,366
	(7,893)	(7,728)	(5,604)
Food expenditure	4,031	3,800	3,181
	(5,172)	(3,518)	(2,436)
Non-durables expenditure	1,667	1,895	1,732
	(3,182)	(3,678)	(2,992)
Durables expenditure	347	363	305
	(2,014)	(2,286)	(1,165)
Housing expenditure	129	145	149
	(341)	(376)	(293)

Notes: Education expenditure is included in the non-durables category. Housing expenditure is calculated using rental values or estimated rental values if housing is owned.

Table 1.15: Changes in poverty using consumption; Correlations between consumption and utility index

1998 relative poverty line	Poverty headcount: Per capita consumption								
	Full sample			Urban areas			Rural areas		
	1998	2001	2005	1998	2001	2005	1998	2001	2005
10%	0.100	0.111	0.138	0.051	0.044	0.050	0.165	0.217	0.261
20%	0.200	0.209	0.250	0.119	0.100	0.116	0.306	0.384	0.439
30%	0.300	0.305	0.355	0.191	0.174	0.195	0.443	0.517	0.581
40%	0.400	0.391	0.441	0.283	0.255	0.279	0.553	0.612	0.670
50%	0.500	0.497	0.534	0.376	0.355	0.377	0.662	0.725	0.758
60%	0.600	0.599	0.640	0.490	0.465	0.500	0.744	0.814	0.842
70%	0.701	0.689	0.729	0.596	0.574	0.616	0.838	0.875	0.891
80%	0.800	0.790	0.836	0.718	0.705	0.759	0.908	0.927	0.948
90%	0.900	0.897	0.920	0.848	0.849	0.881	0.967	0.976	0.977
	Correlation with utility index (Spearman rank correlation coefficient)								
	Full sample			Urban areas			Rural areas		
	1998	2001	2005	1998	2001	2005	1998	2001	2005
Correlation	0.543	0.576	0.587	0.528	0.519	0.522	0.369	0.381	0.399

Notes: The X% 1998 relative poverty line is determined by the consumption threshold classifying the bottom X% of the 1998 national sample as poor.

Chapter 2

Consumption Proxies

Proxying for household expenditure using dichotomous variables: A comparison of wealth indices in five African countries

Joint work with Luc Christiaensen[1]

The utility index developed in the first chapter presents a method for measuring long run economic well-being but was shown to differ both theoretically and empirically from the standard consumption aggregate that is typically used to measure short run welfare. Thus, although the utility index provides a more theoretically compelling method for capturing long run well-being, there is still a need to develop less costly measures of short run welfare. In this chapter, we use an empirical analysis of a set of Living Standards Measurement Surveys (LSMS) to assess various poverty measurement strategies across multiple applications. In this work, we focus specifically on measures constructed using dichotomous indicators for durable goods ownership, housing quality, and household expenditures.

Using data from multiple survey rounds in Ghana, Tanzania, Uganda, Rwanda, and Malawi, we compare indices along two dimensions of variation-index construction method and item choice. We assess three methods of index construction-using variable weights equal to the inverse of the mean population indicator level, deriving weights from principal components analysis (PCA), and estimating weights using a latent variable model referred to as the threshold method. We also generate indices using 20 different sets of variables chosen from the categories of staple foods, non-staple foods, semi-durables, housing variables, and durable goods, as well as subsets and combinations of these categories.

After generating the indices, we evaluate their performance against a benchmark of household per capita expenditure according to three criteria: rank correlation coefficients, sensitivity to identifying poor households, and accuracy of classifying households as poor or non-poor. In the countries we look at, the assessed wealth indices are moderate proxies for per capita expenditure, with rank correlation coefficients ranging between 0.45 and 0.71

[1]Senior Economist, World Bank Africa Region

within the national samples. When we measure poverty across various poverty lines, the various wealth indices are better able to identify poor households at higher levels of poverty.

Comparing the indices across the two dimensions of construction method and variable inclusion, we find that index performance is generally insensitive to construction method. This result is consistent across the various countries and the multiple criteria considered. Although the cross-country meta-analysis identifies PCA as the highest performing method, on average, the differences between construction methods are small according to the assessed criteria. When comparing indices generated using different sets of variables, we find that indices generated using a combination of variables from the five categories of staple foods, non-staple foods, semi-durables, housing variables, and durable goods tend to perform well across multiple criteria, though other variable groups sometimes dominate depending on the application.

We also perform additional analyses to assess index performance within urban and rural subpopulations and in applications using samples pooled across repeated cross-sections. In addition to assessing various construction methods and variable inclusion groups, we examine whether or not indices generated using subsample weights offer improved performance over indices generated using pooled sample weights. We find that the wealth indices continue to moderately proxy for per capita expenditure within rural and urban subsamples but to a lesser extent than in national samples. Indices constructed using national weights perform similarly to or better than indices constructed using within subsample weights. The combination of variables from the five categories continues to perform well in the subsample analyses. The wealth indices are also moderate proxies for per capita expenditure when applied to pooled cross-sectional samples, and there appears to be no difference between indices generated using weights derived from pooled cross-sectional samples and indices generated from weights derived from the initial wave and applied to subsequent waves.

The rest of the paper proceeds as follows. Section 2.1 presents a review of prior work on wealth indices generated from dichotomous variables. We discuss the analytical strategy for assessing wealth indices in Section 2.2. Sections 2.3 and 2.4 present the data and results, and Section 2.5 concludes.

2.1 Background

The growing interest in poverty monitoring requires accurate data on household wealth. Poverty analyses are traditionally done using expenditure aggregates generated from detailed consumption surveys. The primary drawback to standard expenditure aggregates are the data requirements, specifically the need for detailed consumption modules and accurate regional and inter-temporal price deflators. In addition, measures of consumption and expenditure are sensitive to recall bias and framing and face additional difficulties in economies with agricultural and informal economies which require the imputation of the value of home produced goods. Together, these issues make expenditure aggregates prohibitively expensive for the frequent data collection necessary for timely measurements of poverty. This has resulted in the development of alternative strategies for measuring wealth using data that can be inexpensively collected.

These alternate wealth measures are often generated using asset ownership and housing quality indicators and have been increasingly used by researchers in the past two decades as proxies for wealth in instruments lacking detailed expenditure modules. They have been primarily used in the literature on health determinants and inequalities with many applications using the Demographic and Health Surveys (DHS), but have also been used in other areas including studies on the environment (Caldas et al. 2007; Holmes 2003), education (Acham et al. 2012), and more general economics (Doocy et al. 2005; Gunning et al. 2000; Sender 2002; Simmons et al. 2010).[2] In addition to allowing for less costly data collection, these asset-based indices are generally thought to be subject to less measurement error than consumption-based approaches (Sahn and Stifel 2003).

Various methods currently exist to construct such wealth indices; I discuss a few strategies that have been commonly used. The simplest scheme applies equal weights to all assets (Bollen, Guilkey, and Mroz 1995; Gorbach et al. 1998; Guilkey and Jayne 1997; Havanon, Knodel, and Sittitrai 1992; Jensen 1996; Razzaque et al. 1990). Others have attempted to determine weights using policy analysis and consultation with national experts (Navajas et al. 2000) or have chosen to restrict analyses to items which have clear prices and have used asset prices as the appropriate weights (Dargent-Molina et al. 1994). Using the rationale that less commonly owned goods are generally more valuable, Lyte and Morris apply weights based on the inverse of population ownership levels (Lyte et al. 2001; Morris et al. 2000). Latent variable models, which use ownership propensities to generate estimates of the underlying latent variable (argued to be wealth), have been also been used (Ferguson et al. 2003; Montgomery et al. 2000).

The most commonly used methods are statistical methods such as factor analysis and principal components analysis (Filmer and Pritchett 2001; Vyas and Kumaranayake 2006). Aimed at reducing many variables into fewer dimensions, both procedures derive weights based on the variance structure of the underlying data. Although some rationalize the use of such methods under the assumption that wealth drives the underlying processes determining the correlation between asset ownership (Vyas and Kumaranayake 2006), the lack of corresponding economic theory is explicitly noted by most researchers (Filmer and Pritchett 2001; Vyas and Kumaranayake 2006) .[3]

Given the expanding use of these wealth indices, a fair amount of research has focused on discussing and validating their ability to capture an underlying measure of wealth. Earlier works have shown that indices generated using PCA (Filmer and Pritchett 2001), inverse frequency weights (Morris et al. 2000), and latent variable methods (Ferguson et al. 2003; Montgomery and Hewett 2005) show reasonable correspondence to per capita expenditure

[2]See Howe and Filmer for extensive lists of studies which apply such methods (Filmer and Scott 2012; Howe, Hargreaves, and Huttly 2008; Howe et al. 2009).

[3]See Howe and Filmer for a detailed discussion of the strengths and limitations of different methodological categories for an extensive cross-country empirical comparison and discussion (Filmer and Scott 2012; Howe et al. 2012).

and outcomes of interest. More recent reviews caution that asset-based measures are relatively weak proxies for consumption expenditure (Howe et al. 2009) and are more strongly associated with particular expenditure subcomponents such as non-food items (Filmer and Scott 2012).

Given the diversity of methods and the strengths and limitations of each, some work has been done to compare between methods. The evidence on the differences between methods for index construction is mixed. Researchers have compared principal components analysis, an equal weights approach, and a weighting strategy based on population ownership levels and have found similar correlations between each index and per capita consumption across a wide set of countries (Filmer and Scott 2012; Howe, Hargreaves, and Huttly 2008). Some comparisons between approaches have shown that the alternatives result in similar levels of inequality across a varied set of outcome indicators in many countries (Filmer and Pritchett 2001; Wagstaff and Watanabe 2003), while other analyses revealed important discrepancies in inequality measurements across different indices (Houweling, Kunst, and Mackenbach 2003). In spite of these mixed results, different studies have consistently found that household rankings are sensitive to method choice (Filmer and Scott 2012; Houweling, Kunst, and Mackenbach 2003; Howe, Hargreaves, and Huttly 2008).

Thus, although work has been done to discuss the validity of various wealth indices and their relation to one another, a few outstanding issues remain. First, there remains a lack of guidelines on choosing between various index construction methods. As poverty classifications depend highly on household rankings relative to specific thresholds, indices which are broadly correlated may not result in similar inferences about poverty. Since correlation coefficients obscure information on the dispersion of values across the index scale, it is valuable to examine more disaggregate patterns such as how well indices discriminate within poor, middle, and rich household groups. The tendency for some of the index methods to assign large numbers of households very low index values (Howe, Hargreaves, and Huttly 2008) raises concerns about the left end of the wealth distribution.

Furthermore, there may be substantial variation in index performance across different levels of the wealth distribution. While the existing poverty analyses focus on one or two poverty lines relevant to each particular study area, an examination of a broader range of poverty lines may reveal different results and provide insight into the optimal method and asset choices to use in each application.[4] In addition to offering guidance on index choice, a more disaggregated analysis may draw attention to potential areas which require additional examination. For instance, if the existing wealth indices consistently provide low discriminatory power within the poorest households, alternative methods or survey questions should be identified when stratification within this subgroup is particularly important.

Second, little work has been done to identify the most appropriate set of variables to include in these alternate wealth indices. Although most prior work has constructed indices using the most extensive set of durable goods ownership and housing quality variables avail-

[4]Sahn and Stifel show that country rankings based on poverty incidence change based on the choice of poverty line (Sahn and Stifel 2000).

able in the data (Filmer and Scott 2012), this inclusion criteria provides little guidance on whether or not additional items should be included in future survey instruments or omitted without reducing the power for wealth measurement. There is a slim body of empirical evidence suggesting that the number of items included in a wealth index can be potentially reduced with a minimal loss in discriminatory power (Bollen, Glanville, and Stecklov 2001; Filmer and Pritchett 2001; Filmer and Scott 2012), but further work is necessary. Additionally, data on durable goods ownership and housing quality often provides little differentiation among the poorest households which spend the majority of their expenditures on food and non-durables. Indices which include indicators for food and semi-durable expenditures may offer the potential for improved poverty measurement but have not been explored systematically to the best of our knowledge.

There has been some discussion on the appropriate criteria for variable inclusion, but the existing literature presents disjointed suggestions as opposed to a cohesive strategy for selecting variables. Potential considerations include excluding assets which may affect the outcomes of interest through non-wealth pathways (Houweling, Kunst, and Mackenbach 2003) and addressing the issue of economies of scale in household assets (Filmer and Scott 2012). Other considerations are method-specific-Sahn and Stifel omit variables with a covariance structure inconsistent with the identifying assumptions of factor analysis (Sahn and Stifel 2000), and Ferguson, et al, restrict their analyses to normal goods due to the structure of their underlying model (Ferguson et al. 2003). Some empirical strategies include using step-wise regressions to identify statistically significant variables which maximize explanatory power (Stifel and Christiaensen 2007) and selecting variables based on an algorithm which maximizes the correlation between the index and expenditures (Mark, Thomas, and Decarli 1996). As these empirical methods were each applied to specific contexts, it remains to be seen whether or not these strategies identify the same variables in different countries.

Third, the use of non-expenditure wealth indices in time series analyses presents additional methodological issues which warrant further attention. Outstanding issues include the question of how to generate weights when data is drawn from multiple time periods in light of relative price changes, unobserved improvements in asset quality, and shifting trends in preferences. These issues are exemplified in the introduction of the cellular phone and other rapid changes in the area of electronics.

Fourth, important geographical variations such as differences between urban and rural areas may require special attention when identifying the most appropriate ways to measure wealth. It is important to understand whether or not the methods applied to national samples remain valid within population subgroups and whether or not the variables which best capture wealth may differ by region.

In this work, we begin to address this research agenda through an analysis of wealth indices along the various dimensions discussed above.

2.2 Methods

To address the issues of constructing valid wealth indices, we employ an empirical approach to identify strategies which correspond most closely to the benchmark of household per capita expenditure. Specifically, we generate a set of wealth indices which vary along the dimensions of construction method and variable choice. We then benchmark these indices against per capita expenditures using a set of performance criteria, repeating the analysis across a set of countries to determine the interaction between contextual factors and index performance. This empirical assessment is meant to complement analyses by other researchers which draw from economic theory to identify expenditure proxies.

Construction methods:

The analysis focuses on two dimensions of index variation-construction method and item choice. We explore three methods of index construction using inverse frequency weighting, principal components analysis (PCA) and a latent variable model referred to as the threshold method.

The inverse frequency method generates an index based on the weighted sum of the included indicator variables. The weights for each variable are calculated using the inverse of the sample mean for each indicator. The inverse frequency measure assigns greater weight to the less common items, making the implicit assumptions that households which own or consume less common items have higher levels of underlying wealth. The index was proposed by Morris and colleagues (Morris et al. 2000) and shown to be well correlated with total monetary asset values in samples of rural households in Northern Mali and Central Malawi. Of the three indices we evaluate here, the inverse frequency method has the advantage of being the most simple and therefore most transparent.

The second method we use is PCA, a statistical procedure used to reduce multidimensional data into a single index. The method uses the underlying variance structure of the data to generate a set of orthogonal principal components and then uses these resulting components to generate a univariate index. As a statistical data reduction procedure, the economic assumptions mapping each variable to the resulting index are unclear, but researchers have posed the idea that household wealth is the underlying attribute driving the correlation between the included variables and is therefore what is being captured by the variance structure of the data. The procedure has been popularized by Filmer and Pritchett (Filmer and Pritchett 2001) and is the most commonly used method for constructing wealth indices.[5]

The third index construction procedure we assess is the threshold method. This method is based on the assumption that households are likely to purchase an item if their wealth surpasses the threshold level of wealth associated with that item. The wealth thresholds

[5]Factor analysis is used as an alternative method for statistically deriving weights (Sahn and Stifel 2000) and has been shown to generate indices that are highly correlated with PCA indices (Filmer and Pritchett 2001).

associated with each item are estimated using a latent variable model where the underlying latent variable is assumed to be household wealth. This is operationalized using a random effects probit model pooled across the included variables. Using the estimated thresholds, the conditional probability of observing a set of items can be calculated for a given level of underlying wealth. From this, the conditional probability of observing each underlying level of wealth given the observed set of asset stocks and purchases can be calculated using Bayes' rule. Wealth indices are generated using the expected value of the latent variable conditional on the observed data. The threshold method was developed by Ferguson and colleagues for the purpose of enhancing cross-country comparability and was shown to perform similarly to PCA using samples from Greece, Peru, and Pakistan (Ferguson et al. 2003). Compared to the inverse frequency method and PCA, the threshold method has the advantage that it is more closely connected to economic theory and can be rationalized under hierarchical preferences where items are purchased following a specified ordering. One drawback of the threshold method is that it is more computationally intensive.

Item selection:

Since there are no current procedural guidelines to determine the appropriate sets of variables to include in each index, we use an exploratory approach. We first choose a range of broad categories that might be considered good indicators for household wealth. These categories include staple food consumption, other food (meats, dairy, and oil) consumption, semi-durable household goods purchases, housing quality, and durable goods ownership.

Before describing the variables we include, we pause to discuss the validity of including indicators for consumption and expenditure in the wealth index constructs. The three index construction methods we employ were initially developed and are generally applied to asset ownership and housing variables. To apply these methods to indicators for consumption and expenditure, the assumption is that underlying household wealth determines whether or not households have any consumption or expenditure on these items. Because the connection to underlying theory is somewhat informal for the chosen construction methods, it is difficult to meaningfully discuss whether or not using dichotomous indictors of consumption and expenditure poses substantial problems in generating wealth indices. Again, the current work is an exploratory empirical analysis designed to complement theoretical work in determining the appropriate wealth indices.

Returning to the issue of item choice, we choose a short list of indicator variables within each category. This list is chosen to include items which occur at varying levels of frequency. Under the rationale that very common items provide discrimination among the poorest while very uncommon items provide discrimination among the wealthy, including both common and uncommon items should allow us to classify households across the wealth distribution. The specific items we include vary by country due to survey and contextual differences. As an example, Table 2.1 shows the included items for the Ugandan analysis and the population frequencies for the three waves of data.

For the food categories, we generate a dichotomous variable for each item to indicate whether or not households had consumed the item, through purchase or self-production, within the reference time period. The reference time periods for food consumption were either one week or one month recalls depending on the country instrument. Staple food variables include a list of 10 to 13 grains and tubers such as rice, maize, and cassava. Other foods (referred to as the meat category) include fresh milk, eggs, cooking oil, and the two most common meats.

Within the category of semi-durables, we choose 5 items and again create indicators for household purchases within the relevant reference period. These items include personal goods such as clothing and shoes and household goods such as soap, toothpaste, toilet paper, and cooking fuels such as kerosene or charcoal. The relevant reference periods differ by item; annual recall periods are commonly used for clothing while monthly or weekly recall periods are used for household supplies. Again, the included items differ depending on each country's survey instrument.

Within the category of housing, we use 5 to 7 variables on housing quality and ownership. To capture housing quality, we use indicators for housing materials, sanitation and density. In the case of Uganda, quality roof materials are defined as iron sheets, quality floor materials are defined as concrete, and baked bricks and concrete are classified as quality wall materials. Quality materials are similarly defined in the other countries. Sanitation is captured by an indicator for toilets with piping. An indicator for low density housing is also included and defined as more than 0.5 rooms per person. When available, housing and land ownership are also included.

The durables category includes indicators for durable goods ownership, which often include furniture, electronics, kitchen appliances, and transportation assets. The durable goods indicators range widely across surveys, and we include between 9 and 12 items depending on the survey. We include the 12 most commonly owned durables in surveys with a more expansive set of durables (Malawi and Tanzania). Furthermore, we restrict our analysis to assets that are primarily not used for livelihood activities in order to limit any biases associated with specific productive sectors. For instance, we do not include livestock as the wealth indices would be biased against non-farming households, since methods do not currently exist to generate comparable indices constructed using different variables.

To determine whether or not smaller groups of variables are sufficient for capturing household wealth, we also examine strict subsets of the larger categories of staple foods and durable goods. We employ two strategies for choosing the variable subsets. The first strategy restricts the analysis to the five most common indicators, under the rationale that the most common indicators provide the most information about the poorest households. The second strategy uses the correlation structure of the variables to eliminate items which are highly correlated with one another under the rationale that highly correlated items add little additional discriminatory power. Specifically, we calculate the average correlation between each item and other items in the category and eliminate the item with the highest average correlation, repeating until we have identified the five items with the lowest average correlation.

In addition to generating indices using indicators grouped by variable category, we also assess indices generated using combinations of these categories. Within each sample, we test 20 alternate variable groupings, which are detailed in Table 2.2. These variable groupings include category-specific variables, combinations of common durables and each of the category-specific variables, and a combination of variables from all categories.

Performance criteria:

We assess each index according to a set of performance criteria related to the potential applications associated with alternative wealth measures. We focus on two main goals - classifying relative household rankings and identifying poor households. Following the literature, we use annual per capita household expenditure as the empirical benchmark for household wealth. Under partial consumption smoothing, per capita expenditure is generally considered less sensitive to temporal shocks than household income but more sensitive to shocks than durable goods ownership. We use the Oxford (or OECD) equivalent scale for calculating household size.[6]

We use three measures to compare the various indices-rank correlation coefficients, true positive rates for poverty, and accuracy rates for poverty. The rank correlation coefficients are used to assess the capacity of each index to broadly rank order households. The latter two measures focus on the use of such indices for the specific purpose of measuring poverty and identifying poor households. Households are classified as poor or non-poor using expenditure and the various indices. The true positive rate is defined as the fraction of poor households (as measured using expenditure) that are correctly identified as poor using the wealth indices. Because this measure fails to take false positive errors into account, we also use a measure of accuracy defined as the sum of the true positive and true negative rates. The true negative rate is defined as the fraction of non-poor households that are correctly identified as non-poor using the wealth indices. We examine the poverty measures using a range of poverty lines to understand how index performance may vary across low, moderate, and high poverty settings. Since the indices we examine generate ordinal measures of wealth, we use relative poverty lines to classify households.

Additional analyses:

In addition to assessing index performance across each country in a given time period, we are also interested in index validity within population subgroups and in repeated cross-sectional analysis.

We focus on the urban and rural population subgroups and compare two strategies. We first determine whether or not indices generated using national samples create measures

[6]This scale assigns a weight of 1 to the first adult, 0.7 to subsequent adults, and 0.5 to each child (defined as a household member under the age of 15). http://www.oecd.org/eco/growth/OECD-Note-EquivalenceScales.pdf

which are valid within urban populations. We then generate indices using the urban subsample to see whether or not these result in improved performance in the urban population. We repeat the analysis for rural populations.

We follow a similar process to evaluate index validity for the analysis of repeated cross-sections, comparing two alternate construction strategies. The first strategy derives weights from a pooled dataset constructed from all time periods. The second strategy derives weights from the initial wave of the sample and applies those weights to the later waves of data.

In addition, we repeat the analysis of indices which differ by construction method and variable inclusion groups to determine whether or not indices which perform well in analyses of national, single wave samples continue to perform well in urban and rural areas and in pooled cross-sectional samples.

Meta-analysis:

To determine index sensitivity to contextual factors, we repeat our analysis for a set of countries and run a set of basic meta-analyses to highlight broad patterns of performance. From these meta-analyses, we can generate a set of policy recommendations for practitioners interested in utilizing such wealth indices.

2.3 Data

We perform our analysis using data from the Living Standards Measurement Surveys (LSMS), which are collected in conjunction with the World Bank to provide accurate measures of household wellbeing across a variety of contexts. The survey instruments include modules on consumption and expenditure of extensive lists of items as well as modules on housing quality and durable goods ownership. As such, the LSMS datasets allow us to compare the indices of interest to well-constructed aggregate expenditure measures. We use the expenditure aggregates generated by the World Bank, which account for spatial variation in prices.

We use data from five countries in Sub-Saharan Africa-Uganda, Rwanda, Tanzania, Ghana, and Malawi. With the exception of Ghana, which is classified as lower-middle income, the countries are grouped in the World Bank low-income category with per capita gross national incomes below 1,035 international dollars (2013). We include three rounds of data from Uganda in 2005/2006, 2009/2010, and 2010/2011, three rounds of data from Rwanda in 2000/2001, 2005/2006, and 2010/2011, two rounds of data from Tanzania in 2008/2009 and 2010/2011, three rounds of data from Ghana in 1991/1992, 1998/1999, and 2005/2006, and two rounds of data from Malawi in 2004/2005 and 2010/2011. Sample sizes range from approximately 3200 in Tanzania to approximately 14300 in the last wave of the Rwandan data. Although there are panel portions of data from the Tanzanian and Ugandan samples, we do not exploit this aspect of the data for our analysis.

2.4 Results

We first present results for one country, Uganda, and then the meta-analysis using the results from the five country samples. Table 2.1 shows the summary statistics for variables included in the Ugandan analysis. The percent of households reporting a positive indicator for each variable are broadly comparable across time. Some exceptions include kitchen appliances, jewelry, and cellular phones; the first two variables decreased substantially while cellular phone ownership increased largely over time.

The subsets of variables used to construct each index are shown in Table 2.2. The five staple foods identified by the low correlation procedure described above are maize and cassava, which are commonly consumed and millet, potato, and sorghum, which are less commonly consumed. Similarly, the five durable goods identified by the low correlation procedure (furniture, kitchen appliances, bicycles, generators, and solar panels) capture a range of ownership levels.

Table 2.3 presents the first set of empirical results for Uganda, the rank correlation coefficients between the indices and per capita expenditure for the three waves of the Ugandan sample. Within the first wave, the correlation between the various indices and per capita expenditure varies widely, ranging from a minimum of -0.25 to a maximum of 0.71.

Looking within each row, we find that, in general, the three methods of construction generate indices with broadly similar levels of correlation to per capita expenditure. For instance, the rank correlation coefficients for the indices constructed using the five meat and other food variables are 0.43, 0.45 and 0.42 using the inverse frequency method, PCA, and the threshold method, respectively. Larger variation exists between construction methods within certain variable groups. For instance, the rank correlation coefficients for the indices constructed using the full list of staples are 0.28, 0.47 and 0.36 using the inverse frequency method, PCA, and the threshold method, respectively.

The construction method with the highest correlation coefficients varies by variable group and sample. Although the PCA index is associated with the highest correlation coefficients for most of the variable groups, it is sometimes outperformed by the inverse frequency index or the threshold index. Furthermore, within variable groups, the construction method with the highest performance in one wave may not necessarily generate the highest correlation coefficients in another wave.

Comparing across variable groups, the index constructed using the set of 26 variables across all categories has the highest correlation with per capita expenditure, with a coefficient ranging between 0.66 and 0.71, depending on the construction method. High correlation coefficients are also associated with indices generated with fewer variables; the highest performing variable groups are the housing variables, semi-durables, the full set of durable goods, and the combination of common durables and housing variables. In the context of Uganda, indices generated using indicators for staple food consumption perform the most poorly and are sometimes negatively correlated with per capita expenditure.

Comparing between indices constructed from variables in the same category, we find that indices constructed from a strict subset of variables can potentially perform similarly to indices constructed from the full list. Within the durables category, the inverse frequency index including all 12 durables has a correlation coefficient of 0.56; the same index using the 5 most common durables has only a slightly lower correlation coefficient of 0.48. The index associated with the 5 least correlated durables performs less well, with a correlation coefficient of 0.33. Within the staple foods category, the inverse frequency index including all 11 staple foods has a correlation coefficient of 0.28, while the same index including the 5 common staple foods performs slightly more poorly with a correlation coefficient of 0.22. Again, the index associated with the 5 least correlated staple foods performs less well, with a correlation coefficient of -0.06. Similar patterns are seen using the second and third waves of the Ugandan LSMS.

Table 2.4 presents the sensitivity statistics (true positive rates) for the various indices for the first wave of the Ugandan data. Similar patterns are seen for the other two waves. Again, the true positive rates are measured as the fraction of the true poor (as measured by consumption) that are classified as poor by the index. Poverty is determined using relative poverty lines and is shown for varying levels of relative poverty. The true positive rates for the various indices are increasing in the level of poverty considered. This pattern is partially generated by the use of relative poverty lines since the true positive rates are necessarily non-zero at all poverty lines above 50%. For example, under the 60% poverty line, an index which incorrectly classifies the wealthiest 40% of the sample as poor will necessarily correctly classify 20% of the sample as poor. The pattern still holds for poverty lines below 50%, though, indicating that the different indices are better able to identify poor households at higher levels of poverty.

Comparing across construction methods within each row, we find that the true positive rates are similar for the inverse frequency and threshold indices for all variable groups. There is higher variation in the true positive rates using the PCA index. At the lower poverty lines, the true positive rates associated with PCA exceed those for the corresponding inverse frequency and threshold indices for the indices constructed using certain food categories. The true positive rates associated with PCA are much lower than the other construction methods for the index constructed using semi-durable variables. At higher poverty lines, the true positive rates are generally similar across all three construction methods.

Comparing across variable groups, we find that the indices generated using the sets of semi-durable indicators and housing variables perform the best in terms of identifying poor households. At the lowest poverty lines, the short list of low correlation durables has the highest true poverty rates, but this index is outperformed by the semi-durable indicators and housing variables at higher poverty lines. As with performance measured using rank correlation coefficients, the indices generated using food indicators perform the most poorly.

Table 2.5 presents the accuracy rates for the various indices for the first wave of the Ugandan data. Again, we find similar patterns in the other two waves. We define accuracy as the fraction of the sample correctly classified as either poor or non-poor using per capita expenditure as the benchmark. Compared to the correlation and true positive rates, there is less variation in accuracy rates across both dimensions of construction method and item choice.

The indices constructed using the set of variables from all five categories have the highest accuracy rates across the various poverty lines, though the accuracy rates are also high for the indices generated using the combination of common durables and housing and the combination of common durables and semi-durables. Accuracy rates are highest for the lowest and highest poverty lines and lower using the middle poverty lines. This pattern is a result of the use of relative poverty lines which automatically generate variation in the potential levels of accuracy at different poverty lines. For instance, at the 10% poverty line, the lowest possible level of accuracy is 0.80, which occurs under the scenario that all of the households classified as poor are in fact non-poor, resulting in an incorrect classification for 20% of the population.[7] In contrast, the lowest level of accuracy that is possible at the 50% relative poverty line is 0 if the wealth indices classify all poor households as non-poor and all non-poor households as poor.

Table 2.6 shows the rank correlation coefficients for the additional analyses by geographical location and within the pooled cross-sectional samples. The first two columns show the analysis for the first wave of data in rural areas, the middle two columns show the analysis for the first wave of data in urban areas, and the right two columns show the analysis for the pooled cross-sectional sample. The correlation coefficients for the threshold indices are not shown for the sake of space but are broadly comparable. Comparing across construction methods, we find that the indices generated using PCA tend to outperform those constructed using inverse frequency weights.

Within rural areas, indices generated using weights derived from the rural sample perform similarly to indices generated using weights derived from the national sample, though the inverse frequency indices generated using rural weights slightly outperform the corresponding indices generated using national weights. The correlation coefficients within rural areas are relatively similar to the levels in the national sample. Though the indices constructed using food variables perform relatively better than the corresponding indices in the national sample, the highest performing variable groups are similar to those identified in the national sample. These high performing variable groups are the large set of 26 variables, the combination of common durables and housing variables, the combination of common durables and semi-durables, and the combination of common durables and common foods.

Within urban areas, indices generated using weights derived from the urban sample also perform similarly to indices generated using weights derived from the national sample. In this case, the inverse frequency indices generated using national weights often outperform the corresponding indices generated using urban weights. The correlation levels in urban areas are generally lower than those seen in both the national sample and the rural sample. Within urban areas, the highest performing indices are generated using the subset of housing variables, the combination of common durables and housing variables, and the large set of 26 variables across all categories.

[7]Because the wealth indices provide somewhat lumpy classifications, they may classify more than 10% of households as poor using a 10% relative poverty line. For this reason, the accuracy rates fall below 0.80 at the 10% poverty line for a few of the indices in Table 2.5.

The right most panel of Table 2.6 shows the corresponding results using data pooled from all three rounds. The indices in the columns labeled wave 1 weights are generated using weights derived from the first wave which are then applied to the second and third waves. The indices in the columns labeled pooled weights are generated using weights derived from the pooled cross-sectional sample. The correlation coefficients tend to be slightly lower than those seen within each wave when generated using inverse frequency weights; the indices generated using PCA perform similarly in the pooled cross-section analysis when compared to the wave-specific samples. Variable groups which perform well within each wave also perform well in the pooled analysis, with the highest correlation coefficient associated with the index constructed from the set of 26 variables.

To see whether or not the results we find in Uganda are consistent across different contexts, we turn to the five country meta-analysis of index performance. Table 2.7 shows the results from a regression of index performance on different dimensions of index construction, where each data point is a country/wave/index observation. The omitted groups are the full list of staple variables and the inverse frequency weighting method.

The first column of estimates presents results using rank correlation coefficients as the performance criteria. Comparing between construction methods, we find that the best performance is associated with indices generated using PCA. Though the difference is statistically significant, the difference in magnitude is relatively small. The threshold method performs similarly to the inverse frequency method. After controlling for the list of variables included, there is no relationship between the total number of variables included and index performance. Across the set of five countries, there is a negative but insignificant relationship between country-level per capita GDP and index performance. Comparing between variable categories (Table 2.7), we find that, on average, the indices generated using variables from all five categories (durables, staples, meats, semi-durables, and housing) have the highest correlation coefficients. Other higher performing categories include the full list of durables, the combination of common durables and semi-durables, and the combination of common durables and housing variables.

The second column of coefficients shows the corresponding estimates using the true positive rate as the performance criteria. In this regression, each data point is a country/wave/index/poverty line observation where the poverty lines are relative poverty lines at 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. Again, we find that the PCA construction method outperforms the inverse frequency and threshold methods, but the differences in magnitude are small. We find no relationship between the total number of variables included and index performance, but we do find that the wealth indices tend to perform more poorly in country-years with higher per capita GDP when true positive rates are used as the performance criteria. Comparing between variable categories (Table 2.7) we find that, on average, the indices generated using meats and other foods and indices generated using semi-durables variables are associated with the highest true positive rates. Other higher performing categories include the list of low correlation durables and the combination of variables from all five categories.

The right panel of Table 2.7 shows the regression coefficients using accuracy (the sum of the true positive and true negative rates) as the performance criteria. The first column of the right panel shows the estimates for the meta-analysis pooled across poverty lines, while the second and third columns restrict the data to the 10% and 50% poverty lines, respectively. The results for the construction methods is similar to those found using the other performance criteria, though the difference between PCA and the inverse frequency method is no longer significant when assessing accuracy at the lowest poverty line. Again, we find no relationship between the total number of variables included and index performance. The relationship between per capita GDP and index performance varies depending on the poverty line considered, but the magnitude is small in all cases.

Taking all poverty lines into consideration, the indices generated using the combination of variables from all five categories are associated with the highest accuracy rates. Other higher performing categories include the combination of common durables and semi-durables and the combination of common durables and housing variables. When we restrict the analysis to the lowest poverty line, we find very little variation in accuracy by variable group, with the exception of indices constructed using meats and indices constructed using low correlation staples, which perform significantly worse. The highest performing variable groups are the combination of common durables and common staple foods and the combination of common durables, common staples, and meats, but the differences are not statistically significant. At the 50% relative poverty line, the highest performing variable groups are semi-durables, the full set of durables, and the combination of common durables and semi-durables.

Table 2.8 presents the results of the meta-analysis on the urban and rural subsamples as well as the pooled cross-sectional samples. The rank correlation coefficient is the independent variable. Each column repeats the analysis for a different sample. Looking across samples, we find that the dominant construction method is PCA, though the differences between methods are not significant in the urban sample.

The left panel of Table 2.8 shows the results from the urban/rural analysis. The first column assesses index performance pooling the results from the national sample and the rural and urban samples. We find that, compared to the national sample analyses, the indices perform more poorly within rural areas. Performance is further diminished when restricting the sample to urban areas. The second column presents the results for rural areas. We find that the indices generated using national weights are not statistically different than indices constructed using weights generated from the rural subsample. As with the national sample, the highest performing index is generated using variables from all five categories; other indices which perform well are the combination of common durables, common staples foods, and meats. The third column presents results of the urban sample analysis. In urban areas, indices generated using national weights are better correlated with per capita expenditure than indices generated using weights generated from urban areas. In urban areas, the highest performing index is generated using variables from all five categories; other indices which perform well are the full set of durables, and the combination of common durables, low correlation staples foods, and meats.

The right panel of Table 2.8 shows the results from the pooled cross-sectional analysis. The first column of the right panel compares index performance in the single wave analyses to index performance in pooled cross-sectional analyses. Each observation is a country-index-time combination, where the time dimension is either a single wave or a pooled cross-section. We find that indices applied in single country pooled cross-sectional samples perform similarly to those applied in single country-wave samples. Furthermore, there is no difference between indices constructing from the pooled sample and those constructed by applying weights from the first wave to data from later waves. The last column presents the results of the meta-analysis restricted to pooled cross-sectional samples. The highest performing index is generated using the set of semi-durables; other high performing indices are generated using the combination of common durables and meats and the combination of common durables and semi-durables.

2.5 Discussion

In summary, we have assessed the ability of various wealth indices to proxy for per capita household expenditure using an empirical analysis of five African countries across three performance criteria. Comparing indices along the dimensions of construction method, we find that PCA generates indices which correspond most closely to per capita expenditure, though there is little variation in performance across construction methods. Comparing indices across variable inclusion groups, there is some variation in the variable groups which best proxy for expenditure depending on the criteria considered. In general, we find that indices generated using a combination of variables from the five chosen categories tend to perform well across multiple performance criteria.

In our additional analyses, we find that the wealth indices continue to moderately proxy for per capita expenditure within rural and urban subsamples but to a lesser extent than in national samples. Indices constructed using national weights perform similarly to or better than indices constructed using within subsample weights. Indices generated using the combination of variables from the five categories perform well in both urban and rural subsamples. The wealth indices are also moderate proxies for per capita expenditure when applied to pooled cross-sectional samples, and there appears to be no difference between indices generated using weights derived from pooled cross-sectional samples and indices generated from weights derived from an initial wave.

Although this analysis addresses important dimensions of wealth measurement, there are many limitations which require further work. First, the three chosen performance criteria are a small subset of many potential measures, each capturing different potential applications of these wealth indices. One directly related application which we have not explored is the ability of such wealth indices to accurately measure changes in poverty headcounts over time. Second, the performance criteria considered are summary statistics which may mask important discrepancies generated by the wealth indices. For instance, though the different wealth indices are moderately accurate in classifying poor and non-poor households, over 10% of each sample are incorrectly classified. Without a more thorough examination of these classification errors, it is unclear whether or not the non-expenditure indices are systematically biased against certain types of households. Third, though we consider multiple countries, they represent a relatively narrow range of development levels, generally ranking among the poorest in the world. Thus, we are unable to predict how such wealth measurement strategies will perform in lower-middle and middle income countries.

Nevertheless, this work represents a step in improving the understanding of non-expenditure measures of household wellbeing. With additional research on the properties of such measures, these indices can potentially be used to monitor poverty both inexpensively and accurately.

2.6 Figures and tables

Table 2.1: Summary statistics for included indicator variables - Uganda

	Indicator mean		
Item	Wave 1	Wave 2	Wave 3
Staple foods			
Consumed beans in past week	0.74	0.75	0.76
Consumed maize in past week	0.63	0.62	0.53
Consumed cassava in past week	0.53	0.61	0.41
Consumed sweet potato in past week	0.50	0.57	0.12
Consumed matooke in past week	0.46	0.34	0.46
Consumed ground nuts in past week	0.42	0.39	0.37
Consumed rice in past week	0.29	0.25	0.24
Consumed bread in past week	0.23	0.22	0.20
Consumed millet in past week	0.17	0.15	0.19
Consumed potato in past week	0.12	0.13	0.11
Consumed sorghum in past week	0.11	0.13	0.13
Meats and other foods			
Consumed cooking oil in past week	0.58	0.57	0.60
Consumed fresh milk in past week	0.35	0.35	0.34
Consumed beef in past week	0.34	0.32	0.33
Consumed fish in past week	0.22	0.17	0.16
Consumed eggs in past week	0.14	0.13	0.11
Housing variables			
Owns house	0.75	0.83	0.82
Roof material - iron sheets	0.63	0.69	0.63
Low density housing ($>$ 0.5 rooms per person)	0.42	0.36	0.34
Wall material - baked bricks or concrete	0.40	0.42	0.40
Floor material - concrete	0.33	0.29	0.27
House has toilet with pipes	0.04	0.07	0.05
Semi-durables			
Purchased soap (for washing or bathing) in past month	0.98	0.92	0.90
Purchased clothing in past year	0.88	0.85	0.78
Purchased shoes in past year	0.70	0.67	0.60
Purchased toothpaste in past month	0.50	0.47	0.39
Purchased charcoal in past month	0.26	0.24	0.23
Durables			
Owns at least one piece of furniture	0.99	0.90	0.90
Owns at least one television/radios	0.66	0.68	0.67
Owns at least one kitchen appliance	0.39	0.20	0.15
Owns at least one piece of jewelry	0.37	0.20	0.16
Owns at least one bicycle	0.36	0.40	0.40
Owns at least one cellular phone	0.22	0.51	0.56
Owns at least one motorcycle	0.02	0.07	0.06
Owns at least one power generator	0.01	0.02	0.01
Owns at least one solar panel	0.01	0.01	0.02

Table 2.2: Variations on item choice - index definitions (Ugandan sample)

Variable group	Num of var	Included variables
Staples (full)	11	Beans, maize, cassava, sweet potato, matooke, ground nuts, rice, bread, millet, potato, sorghum
Staples (list 1, most common)	5	Beans, maize, cassava, sweet potato, matooke
Staples (list 2, low correlation)	5	Maize, cassava, millet, potato, sorghum
Meats and other foods	5	Cooking oil, milk, beef, fish, egg
Foods (full)	16	Beans, maize, cassava, sweet potato, matooke, ground nuts, rice, bread, millet, potato, sorghum Cooking oil, milk, beef, fish, egg
Foods (staples 1 and meats)	10	Beans, maize, cassava, sweet potato, matooke Cooking oil, milk, beef, fish, egg
Foods (staples 2 and meats)	10	Maize, cassava, millet, potato, sorghum Cooking oil, milk, beef, fish, egg
Semi-durables	5	Soap, clothes, shoes, toothpaste, charcoal
Housing	6	Flooring, roofing, toilet, house, density, walls
Durables (full)	9	Furniture, tv/radio, kitchen appliance, jewelry, bike, cell phone, motorcycle, generator, solar panel
Durables (short list 1)	5	Furniture, tv/radio, kitchen appliance, jewelry, bike
Durables (short list 2)	5	Furniture, kitchen appliance, bike, generator, solar panel
Durables 1 and staples 1	10	Furniture, tv/radio, kitchen appliance, jewelry, bike Beans, maize, cassava, sweet potato, matooke
Durables 1 and staples 2	10	Furniture, tv/radio, kitchen appliance, jewelry, bike Maize, cassava, millet, potato, sorghum
Durables 1 and meats	10	Furniture, tv/radio, kitchen appliance, jewelry, bike Cooking oil, milk, beef, fish, egg
Durables 1 and food (short 1)	15	Furniture, tv/radio, kitchen appliance, jewelry, bike Beans, maize, cassava, sweet potato, matooke Cooking oil, milk, beef, fish, egg
Durables 1 and food (short 2)	15	Furniture, tv/radio, kitchen appliance, jewelry, bike Maize, cassava, millet, potato, sorghum Cooking oil, milk, beef, fish, egg
Durables 1 and semi-durables	10	Furniture, tv/radio, kitchen appliance, jewelry, bike Soap, clothes, shoes, toothpaste, charcoal
Durables 1 and housing	11	Furniture, tv/radio, kitchen appliance, jewelry, bike Flooring, roofing, toilet, house, density, walls
All (durables 1, staples 1, meats, semi-durables, housing)	26	Furniture, tv/radio, kitchen appliance, jewelry, bike Beans, maize, cassava, sweet potato, matooke Cooking oil, milk, beef, fish, egg Soap, clothes, shoes, toothpaste, charcoal Flooring, roofing, toilet, house, density, walls

Table 2.3: Rank correlations with per capita expenditure - Uganda

	Spearman correlation coefficients								
	Wave 1			Wave 2			Wave 3		
	Construction method: IW=inverse frequency weights, PCA=principal components analysis, TH=threshold method								
Variable group	IW	PCA	TH	IW	PCA	TH	IW	PCA	TH
Staples (full)	0.28	0.47	0.36	0.37	0.45	0.38	0.35	0.47	0.41
Staples (list 1, most common)	0.22	0.22	0.19	0.15	0.15	0.16	0.20	0.30	0.22
Staples (list 2, low correlation)	-0.06	-0.25	0.01	0.09	-0.19	0.11	0.07	-0.20	0.08
Meats and other foods	0.43	0.45	0.42	0.43	0.45	0.44	0.42	0.44	0.43
Foods (full)	0.41	0.53	0.45	0.46	0.53	0.48	0.46	0.55	0.51
Foods (staples 1 and meats)	0.43	0.45	0.41	0.42	0.42	0.40	0.41	0.49	0.46
Foods (staples 2 and meats)	0.26	0.45	0.33	0.34	0.45	0.38	0.33	0.46	0.38
Semi-durables	0.56	0.51	0.51	0.58	0.52	0.51	0.58	0.52	0.53
Housing	0.60	0.61	0.54	0.50	0.50	0.45	0.49	0.50	0.43
Durables (full)	0.51	0.56	0.53	0.47	0.50	0.48	0.45	0.51	0.48
Durables (short list 1)	0.38	0.48	0.42	0.42	0.42	0.41	0.35	0.40	0.41
Durables (short list 2)	0.18	0.33	0.33	0.32	0.21	0.32	0.30	0.21	0.30
Durables 1 and staples 1	0.39	0.45	0.38	0.40	0.34	0.35	0.38	0.45	0.41
Durables 1 and staples 2	0.18	0.49	0.33	0.33	0.45	0.37	0.29	0.44	0.36
Durables 1 and meats	0.51	0.56	0.52	0.52	0.53	0.52	0.48	0.53	0.51
Durables 1 and food (short 1)	0.49	0.54	0.48	0.50	0.49	0.47	0.48	0.55	0.51
Durables 1 and food (short 2)	0.37	0.56	0.45	0.45	0.54	0.48	0.42	0.55	0.48
Durables 1 and semi-durables	0.57	0.59	0.55	0.57	0.57	0.53	0.57	0.60	0.57
Durables 1 and housing	0.62	0.67	0.59	0.55	0.55	0.52	0.52	0.56	0.50
All (durables 1, staples 1, meats, semi-durables, housing)	0.66	0.71	0.63	0.65	0.66	0.59	0.65	0.69	0.65

Table 2.4: Sensitivity rates (poor/true poor) by index - Uganda (wave 1)

	Sensitivity = Poor / True poor								
	Construction method: IW=inverse frequency weights, PCA=principal components analysis, TH=threshold method								
	10% relative poverty line			30% relative poverty line			50% relative poverty line		
Variable group	IW	PCA	TH	IW	PCA	TH	IW	PCA	TH
Staples (full)	0.20	0.29	0.20	0.44	0.52	0.46	0.60	0.67	0.63
Staples (list 1, most common)	0.19	0.19	0.17	0.43	0.42	0.47	0.56	0.60	0.58
Staples (list 2, low correlation)	0.09	0.21	0.09	0.28	0.24	0.44	0.50	0.41	0.51
Meats and other foods	0.46	0.46	0.46	0.63	0.63	0.63	0.67	0.67	0.67
Foods (full)	0.21	0.30	0.22	0.47	0.53	0.50	0.66	0.71	0.68
Foods (staples 1 and meats)	0.26	0.24	0.25	0.50	0.49	0.47	0.67	0.68	0.66
Foods (staples 2 and meats)	0.14	0.24	0.14	0.46	0.49	0.44	0.59	0.67	0.63
semi-durables	0.56	0.21	0.59	0.80	0.49	0.86	0.74	0.73	0.73
Housing	0.57	0.53	0.59	0.60	0.58	0.62	0.74	0.73	0.65
Durables (full)	0.42	0.42	0.41	0.52	0.62	0.55	0.71	0.72	0.71
Durables (short list 1)	0.41	0.43	0.42	0.46	0.53	0.46	0.64	0.69	0.69
Durables (short list 2)	0.62	0.63	0.61	0.52	0.53	0.52	0.48	0.66	0.65
Durables 1 and staples 1	0.30	0.31	0.30	0.47	0.51	0.46	0.64	0.67	0.64
Durables 1 and staples 2	0.17	0.35	0.22	0.40	0.52	0.46	0.55	0.68	0.60
Durables 1 and meats	0.28	0.39	0.30	0.52	0.56	0.53	0.69	0.71	0.69
Durables 1 and food (short 1)	0.28	0.31	0.29	0.53	0.54	0.54	0.69	0.71	0.68
Durables 1 and food (short 2)	0.20	0.34	0.24	0.46	0.55	0.51	0.64	0.71	0.67
Durables 1 and semi-durables	0.43	0.35	0.34	0.55	0.58	0.55	0.73	0.74	0.72
Durables 1 and housing	0.34	0.34	0.34	0.62	0.61	0.60	0.74	0.76	0.73
All (durables 1, staples 1, meats, semi-durables, housing)	0.43	0.42	0.41	0.61	0.64	0.59	0.75	0.78	0.73

Notes: Poverty is measured using per capita expenditure. Results for additional poverty lines are not shown due to space restrictions.

Table 2.5: Accuracy rates (true poor + true non-poor) by index - Uganda (wave 1)

	Accuracy=(True Poor + True Non-Poor)/Total								
	Construction method: IW=inverse frequency weights, PCA=principal components analysis, TH=threshold method								
	10% relative poverty line			30% relative poverty line			50% relative poverty line		
Variable group	IW	PCA	TH	IW	PCA	TH	IW	PCA	TH
Staples (full)	0.83	0.86	0.84	0.65	0.71	0.67	0.60	0.67	0.63
Staples (list 1, most common)	0.84	0.83	0.83	0.65	0.64	0.64	0.56	0.60	0.58
Staples (list 2, low correlation)	0.80	0.65	0.80	0.55	0.54	0.53	0.50	0.41	0.51
Meats and other foods	0.75	0.75	0.75	0.68	0.68	0.68	0.67	0.67	0.67
Foods (full)	0.84	0.86	0.84	0.68	0.72	0.70	0.66	0.71	0.68
Foods (staples 1 and meats)	0.84	0.85	0.85	0.69	0.69	0.68	0.67	0.68	0.66
Foods (staples 2 and meats)	0.82	0.83	0.82	0.65	0.69	0.66	0.59	0.67	0.63
Nonfood	0.82	0.84	0.79	0.70	0.69	0.67	0.74	0.73	0.73
Housing	0.78	0.80	0.77	0.72	0.74	0.70	0.74	0.73	0.65
Durables (full)	0.81	0.81	0.81	0.68	0.69	0.70	0.71	0.72	0.71
Durables (short list 1)	0.80	0.80	0.80	0.67	0.71	0.67	0.64	0.69	0.69
Durables (short list 2)	0.64	0.64	0.64	0.63	0.63	0.63	0.48	0.66	0.65
Durables 1 and staples 1	0.86	0.86	0.85	0.68	0.70	0.68	0.64	0.67	0.64
Durables 1 and staples 2	0.83	0.87	0.83	0.64	0.71	0.67	0.55	0.68	0.60
Durables 1 and meats	0.83	0.85	0.83	0.71	0.73	0.72	0.69	0.71	0.69
Durables 1 and food (short 1)	0.86	0.86	0.86	0.72	0.72	0.72	0.69	0.71	0.68
Durables 1 and food (short 2)	0.84	0.87	0.85	0.67	0.73	0.70	0.64	0.71	0.67
Durables 1 and nonfood	0.84	0.86	0.87	0.73	0.75	0.72	0.73	0.74	0.72
Durables 1 and housing	0.86	0.85	0.86	0.75	0.76	0.74	0.74	0.76	0.73
All (durables 1, staples 1, meats, nonfood housing)	0.89	0.88	0.88	0.77	0.78	0.75	0.75	0.78	0.73

Notes: Poverty is measured using per capita expenditure. Results for additional poverty lines are not shown due to space restrictions.

Table 2.6: Rank correlations with per capita expenditure by urban location and for pooled cross-sectional sample

	Spearman correlation coefficients											
Wave	Wave 1				Wave 1				Wave 1, 2, and 3 (pooled)			
Sample	Rural				Urban				National			
Weights	Rural		National		Urban		National		Wave 1		Pooled	
Variable group	IW	PCA	IW	PCA	IW	PCA	IW	PCA	IW	PCA	IW	PCA
Staples (full)	0.34	0.47	0.27	0.48	0.05	0.21	0.21	0.28	0.21	0.47	0.23	0.47
Staples (list 1, common)	0.36	0.36	0.35	0.34	-0.06	-0.09	0.02	0.00	0.11	0.24	0.11	0.23
Staples (list 2, low correlation)	0.01	-0.16	-0.01	-0.15	-0.08	-0.15	-0.04	-0.20	-0.02	-0.20	0.00	-0.21
Meats and other foods	0.40	0.42	0.40	0.42	0.31	0.33	0.32	0.33	0.33	0.44	0.32	0.44
Foods (full)	0.44	0.53	0.39	0.54	0.14	0.32	0.31	0.35	0.29	0.53	0.30	0.53
Foods (staples 1 and meats)	0.46	0.48	0.46	0.48	0.22	0.23	0.28	0.27	0.29	0.46	0.29	0.46
Foods (staples 2 and meats)	0.30	0.44	0.26	0.45	0.11	0.28	0.24	0.31	0.19	0.46	0.21	0.45
Semi-durables	0.48	0.44	0.47	0.43	0.26	0.28	0.25	0.28	0.38	0.52	0.39	0.52
Housing	0.46	0.47	0.46	0.47	0.52	0.51	0.59	0.52	0.39	0.54	0.39	0.54
Durables (full)	0.46	0.47	0.44	0.48	0.32	0.52	0.49	0.53	0.33	0.52	0.36	0.51
Durables (short list 1)	0.41	0.42	0.38	0.44	0.21	0.40	0.29	0.39	0.27	0.44	0.34	0.45
Durables (short list 2)	0.30	0.24	0.24	0.30	0.13	0.12	0.14	0.28	0.14	0.32	0.24	0.20
Durables 1, staples 1	0.47	0.47	0.45	0.49	0.05	0.15	0.20	0.26	0.23	0.43	0.29	0.43
Durables 1, staples 2	0.26	0.45	0.20	0.48	0.03	0.37	0.18	0.37	0.12	0.46	0.20	0.47
Durables 1, meats	0.49	0.51	0.48	0.53	0.32	0.44	0.40	0.44	0.36	0.54	0.39	0.54
Durables 1, food (short 1)	0.52	0.54	0.51	0.55	0.23	0.33	0.34	0.36	0.32	0.53	0.35	0.53
Durables 1, food (short 2)	0.41	0.52	0.37	0.54	0.13	0.40	0.32	0.41	0.24	0.54	0.30	0.54
Durables 1, semi-durables	0.52	0.52	0.50	0.53	0.25	0.43	0.36	0.42	0.39	0.59	0.42	0.59
Durables 1, housing	0.53	0.55	0.52	0.55	0.44	0.61	0.59	0.61	0.40	0.59	0.43	0.59
All (dur. 1, staples 1, meats, semi-durables, housing)	0.62	0.64	0.61	0.64	0.41	0.48	0.52	0.54	0.41	0.68	0.43	0.68

Notes: Construction method: IW=inverse frequency weights, PCA=principal components analysis. Results for threshold method excluded for space.

Table 2.7: Index performance across multiple countries and various performance criteria

	Performance criteria				
	Correlation coefficient	True positive rate	Accuracy (all poverty lines)	Accuracy (10% poverty line)	Accuracy (50% poverty line)
	b/se	b/se	b/se	b/se	b/se
Staples (short list 1, most common)	-0.150	-0.003	-0.014	0.000	-0.015
	(0.069)	(0.012)	(0.008)	(0.028)	(0.024)
Staples (short list 2, low correlation)	-0.269	0.021	-0.070	-0.142	-0.066
	(0.069)	(0.012)	(0.008)	(0.028)	(0.024)
Meats and other foods	0.064	0.129	0.006	-0.097	0.050
	(0.069)	(0.012)	(0.008)	(0.028)	(0.024)
Foods (full)	0.119	0.020	0.014	0.004	0.010
	(0.058)	(0.010)	(0.007)	(0.023)	(0.020)
Foods (staples 1 and meats)	0.086	0.022	0.022	0.011	0.031
	(0.035)	(0.006)	(0.004)	(0.014)	(0.012)
Foods (staples 2 and meats)	0.060	0.024	0.009	-0.013	0.019
	(0.035)	(0.006)	(0.004)	(0.014)	(0.012)
Semi-durables	0.115	0.127	0.037	-0.043	0.079
	(0.069)	(0.012)	(0.008)	(0.028)	(0.024)
Housing	0.063	0.044	0.021	-0.021	0.035
	(0.059)	(0.010)	(0.007)	(0.024)	(0.021)
Durables (full)	0.201	0.066	0.039	-0.029	0.066
	(0.032)	(0.006)	(0.004)	(0.013)	(0.011)
Durables (short list 1)	0.014	0.062	0.013	-0.048	0.038
	(0.069)	(0.012)	(0.008)	(0.028)	(0.024)
Durables (short list 2)	-0.046	0.107	-0.020	-0.100	-0.001
	(0.069)	(0.012)	(0.008)	(0.028)	(0.024)
Durables 1 and staples 1	0.058	0.013	0.014	0.021	0.023
	(0.035)	(0.006)	(0.004)	(0.014)	(0.012)
Durables 1 and staples 2	0.020	0.004	0.004	0.006	0.005
	(0.035)	(0.006)	(0.004)	(0.014)	(0.012)
Durables 1 and meats	0.173	0.056	0.040	0.005	0.060
	(0.035)	(0.006)	(0.004)	(0.014)	(0.012)

Continued on next page

Table 2.7 – continued from previous page

	Performance criteria				
	Correlation coefficient	True positive rate	Accuracy (all poverty lines)	Accuracy (10% poverty line)	Accuracy (50% poverty line)
	b/se	b/se	b/se	b/se	b/se
Durables 1 and food (short 1)	0.197	0.040	0.032	0.017	0.040
	(0.045)	(0.008)	(0.005)	(0.018)	(0.016)
Durables 1 and food (short 2)	0.178	0.034	0.027	0.012	0.034
	(0.045)	(0.008)	(0.005)	(0.018)	(0.016)
Durables 1 and semi-durables	0.196	0.056	0.048	0.014	0.072
	(0.035)	(0.006)	(0.004)	(0.014)	(0.012)
Durables 1 and housing	0.195	0.041	0.042	0.016	0.058
	(0.032)	(0.006)	(0.004)	(0.013)	(0.011)
All (durables 1, staples 1, meats, semi-durables, housing)	0.404	0.084	0.050	0.006	0.048
	(0.137)	(0.024)	(0.016)	(0.055)	(0.048)
PCA	0.034	0.011	0.010	0.006	0.013
	(0.012)	(0.002)	(0.001)	(0.005)	(0.004)
Threshold method	-0.002	0.001	-0.001	0.005	0.001
	(0.012)	(0.002)	(0.001)	(0.005)	(0.004)
Total number of variables	-0.007	-0.000	0.001	0.002	0.003
	(0.009)	(0.002)	(0.001)	(0.004)	(0.003)
Log per capita GDP	-0.016	-0.014	-0.002	0.030	-0.010
	(0.022)	(0.004)	(0.003)	(0.009)	(0.008)
Poverty line		0.101	-0.120		
		(0.002)	(0.001)		
Poverty line squared		-0.003	0.013		
		(0.000)	(0.000)		
Constant	0.473	0.278	0.892	0.595	0.636
	(0.188)	(0.033)	(0.022)	(0.076)	(0.066)

Notes: Sample consists of country - wave - index observations. The omitted groups are the full list of staples and the inverse frequency weights construction method.

Table 2.8: Index performance across multiple countries by location and repeated cross-sections, correlations

	Sample				
	National, urban, rural	Rural	Urban	Single waves, pooled cross-sections	Pooled cross-sections
	b/se	b/se	b/se	b/se	b/se
Staples (short list 1, most common)	-0.180	-0.072	-0.304	-0.088	-0.001
	(0.033)	(0.039)	(0.058)	(0.051)	(0.079)
Staples (short list 2, low correlation)	-0.278	-0.266	-0.293	-0.211	-0.128
	(0.033)	(0.039)	(0.058)	(0.051)	(0.079)
Meats and other foods	0.026	0.010	0.022	0.127	0.215
	(0.033)	(0.039)	(0.058)	(0.051)	(0.079)
Foods (full)	0.141	0.080	0.213	0.070	0.000
	(0.027)	(0.033)	(0.048)	(0.043)	(0.066)
Foods (staples 1 and meats)	0.060	0.056	0.052	0.100	0.121
	(0.017)	(0.020)	(0.030)	(0.026)	(0.039)
Foods (staples 2 and meats)	0.050	-0.002	0.098	0.074	0.095
	(0.017)	(0.020)	(0.030)	(0.026)	(0.039)
Semi-durables	0.042	0.027	0.022	0.157	0.217
	(0.033)	(0.039)	(0.058)	(0.051)	(0.079)
Housing	0.038	-0.044	0.106	0.102	0.156
	(0.028)	(0.034)	(0.050)	(0.044)	(0.067)
Durables (full)	0.177	0.044	0.297	0.196	0.190
	(0.015)	(0.019)	(0.027)	(0.024)	(0.036)
Durables (short list 1)	-0.049	-0.061	-0.067	0.065	0.139
	(0.033)	(0.039)	(0.058)	(0.051)	(0.079)
Durables (short list 2)	-0.078	-0.093	-0.079	0.013	0.096
	(0.033)	(0.039)	(0.058)	(0.051)	(0.079)
Durables 1 and staples 1	0.008	0.033	-0.041	0.062	0.070
	(0.017)	(0.020)	(0.030)	(0.026)	(0.039)
Durables 1 and staples 2	-0.005	-0.059	0.037	0.022	0.026
	(0.017)	(0.020)	(0.030)	(0.026)	(0.039)
Durables 1 and meats	0.143	0.078	0.194	0.181	0.193
	(0.017)	(0.020)	(0.030)	(0.026)	(0.039)
Durables 1 and food (short 1)	0.179	0.119	0.230	0.156	0.102
	(0.022)	(0.026)	(0.038)	(0.034)	(0.051)
Durables 1 and food (short 2)	0.175	0.072	0.276	0.137	0.082
	(0.022)	(0.026)	(0.038)	(0.034)	(0.051)
Durables 1 and semi-durables	0.140	0.068	0.183	0.195	0.195
	(0.017)	(0.020)	(0.030)	(0.026)	(0.039)
Durables 1 and housing	0.168	0.052	0.270	0.186	0.174

Continued on next page

Table 2.8 – continued from previous page

	Sample				
	National, urban, rural	Rural	Urban	Single waves, pooled cross-sections	Pooled cross-sections
	b/se	b/se	b/se	b/se	b/se
	(0.015)	(0.019)	(0.027)	(0.024)	(0.036)
All (durables 1, staples 1, meats, semi-durables, housing)	0.430	0.226	0.648	0.250	0.036
	(0.065)	(0.079)	(0.116)	(0.103)	(0.157)
PCA	0.027	0.033	0.017	0.042	0.053
	(0.006)	(0.007)	(0.010)	(0.009)	(0.014)
Threshold method	0.002	0.016	-0.010	0.008	0.020
	(0.006)	(0.007)	(0.010)	(0.009)	(0.014)
Total number of variables	-0.012	-0.002	-0.024	0.002	0.016
	(0.004)	(0.005)	(0.008)	(0.007)	(0.011)
Log per capita GDP	-0.027	-0.004	-0.056		
	(0.011)	(0.013)	(0.019)		
Rural areas	-0.037				
	(0.007)				
Urban areas	-0.075				
	(0.007)				
National weights		-0.009	0.050		
		(0.006)	(0.009)		
Pooled cross-section sample				0.001	
				(0.008)	
Weights from pooled cross-section					0.002
					(0.011)
Constant	0.621	0.374	0.809	0.252	0.097
	(0.089)	(0.108)	(0.158)	(0.082)	(0.126)

Notes: Sample consists of subsample - wave - weighting - index observations. The omitted groups are the full list of staples and the inverse frequency weights construction method. The omitted weighting groups for the rural and urban regressions are weights constructed from within the rural and urban samples, respectively. The omitted weighting groups for the pooled cross-section analysis are weights generated from the first wave and applied to later waves.

Chapter 3

Pay for Performance

What's in the Black Box of Pay for Performance Programs?
Health facility inputs and institutional deliveries in the Rwandan national program

Joint work with Sebastian Bauhoff[1] and Tisamarie Sherry[2]

Economic well-being, as discussed in the first two chapters, is a primary concern of development economists, but it is not the only concern. Welfare is multidimensional, and one important dimension that remains critically low in developing countries is health. Improving access to high-quality health care services at sustainable cost remains an important priority for low and middle-income countries. One intervention that has been gaining popularity in recent years is incentive-based financing, also referred to as paying for performance. Under pay-for-performance (P4P) programs, health care providers or facilities receive financial rewards based on contracted services or outcomes.

Evaluations to date show mixed impacts on health care quality and outcomes (Basinga et al. 2011; Gertler and Vermeersch 2012; Miller and Babiarz 2013; Olken, Onishi, and Wong 2012; Sherry, Bauhoff, and Mohanan 2014), but less is known about how performance incentives affect provider behavior and the production of health care services. Evidence on the health care production function and provider responses to P4P is critically important for several reasons. First, many aspects of the production of health care services remain poorly understood, particularly in resource-constrained settings. Second, while P4P is often viewed as an approach to achieve targeted improvements in certain rewarded services, it is not evident whether P4P also supports broader improvements in health systems beyond the explicitly rewarded services. Third, and more specifically, there is limited evidence on how facilities and providers allocate effort and resources in response to targeted payment

[1] Associate Economist, RAND Corporation
[2] Harvard Medical School

incentives. Understanding such responses can provide guidance for designing future P4P programs or tweaking existing programs to maximize their effect.

The goal of this paper is to glimpse inside the "black box" of how providers respond to financial incentives under P4P to improve the output and quality of health care services. We examine Rwanda's national P4P program, which was introduced through a randomized rollout from 2006 to 2008. Specifically, we examine how Rwandan providers increased institutional delivery rates in response to explicit P4P incentives for this service by modifying delivery-specific inputs and activities. Moreover, we examine whether the P4P program was associated with changes in general health care inputs that may serve to strengthen the overall health system, in addition to increasing institutional deliveries.

We examine these questions using a facility survey fielded after the randomized rollout, and evaluate differences between treatment and control facilities in physical resources, human resources, and facility management and operations. We first replicate the finding from previous studies that P4P had a positive and significant impact on institutional delivery rates. Turning to inputs, we find positive program effects for the daily presence of staff per capita providing maternity-related services and a general facility management indicator. We do not find statistically significant or substantively meaningful effects on a range of other inputs commonly targeted to improve institutional delivery rates, e.g., inputs that are included in the World Health Organization (WHO) guidelines. Additionally, we assess the linkages between inputs and outcomes and find that the observed changes in staffing and management explain a relatively small fraction of the increase in institutional deliveries.

3.1 Background

Expansion of P4P Programs in Middle- and Low-Income Countries

P4P programs have enjoyed strong support from donors interested in strengthening health systems in middle- and low-income countries: at least 26 incentive-based health programs have been implemented in these settings in the past 15 years (Miller and Babiarz 2013). Very few programs have rewarded providers for improved health outcomes (Miller and Babiarz 2013; Singh 2011)- rather, the majority have contracted on health care inputs and service provision. P4P schemes target specific health care services for improvement, but instead of prescribing the steps that providers should follow to achieve these improvements, it allows them flexibility and broad latitude in developing their own approaches. P4P therefore has the potential to increase economic efficiency through at least two routes: first, by avoiding the information asymmetries that may lead policymakers to recommend quality improvement strategies ill-suited to local conditions; and second, by recognizing that provider heterogeneity may preclude a single "one size fits all" approach (Miller and Babiarz 2013).

There is a small but growing body of evidence evaluating the impacts of P4P programs on health service provision in resource-limited settings. In Indonesia, bonus payments to villages for their performance on maternal and child health care quality indicators improved measures of preventive health care and decreased child malnutrition (Olken, Onishi, and Wong 2012). A hospital-based P4P program in the Philippines, in which performance bonuses were offered based on clinical competence and patient volume either to physicians directly, or at the health system-level to hospitals and physician groups, improved provider knowledge and clinical skills (Peabody et al. 2011). Rwanda's national P4P program had mixed effects on health care quality - more generously rewarded services such as institutional deliveries and the supply of contraception increased, as did several unrewarded prenatal care services. The program had no significant impact on less generously rewarded services, but there were also no negative spillovers to unrewarded services (Basinga et al. 2011; Sherry, Bauhoff, and Mohanan 2014). Similarly, a recent experimental study of P4P in the Democratic Republic of the Congo (DRC) found that financial incentives increased the availability of rewarded health services, with no evidence of negative spillovers to unrewarded services (Huillery and Seban 2013).

The mechanisms by which providers facing a P4P incentive scheme achieve quality improvements, however, remains poorly understood. Several empirical studies have explicitly tested for mediating factors. Olken, Onishi, and Wong (2012) find that health care quality improvements in Indonesia's P4P program were mediated in part by more efficient health care spending, and increased labor supply of providers. Huillery and Seban (2013) find that in the DRC's P4P program, quality improvements were mediated by increased provider effort (e.g. preventive health sessions and community outreach) and decreased absenteeism.

In the case of Rwanda's national P4P program, Gertler and Vermeersch (2012) show that P4P decreased the gap between provider knowledge and the practice of appropriate clinical interventions, positing that this might be one pathway through which financial incentives achieved quality improvements. The other mechanisms through which Rwanda's P4P program improved targeted health services remain unknown.

Rwanda's P4P Program

As one of the earliest large-scale P4P programs in a developing country, Rwanda's national program offers an interesting case study for identifying the mechanisms by which P4P may result in quality improvements. The program rewards facilities with varying unit payments based on their performance on a set of 24 health service indicators in the domains of maternal health, child health, family planning and HIV/AIDS. The rewarded services and the associated unit payments are shown in Table 3.1. Bonus payments are adjusted according to overall facility quality using a quality multiplier, which is constructed by measuring facility performance across several domains, including general administration, financial management, hygiene and sanitation, laboratory services and pharmacy management. Within each of these domains, performance assessments also take into account the availability of key inputs and adherence to clinical protocols. The factors and weights associated with the P4P program's quality multiplier are shown in Table 3.2. Bonus payments are disbursed at the facility level and used at the discretion of each facility. Basinga and colleagues report that prior to the program, approximately 50%, 20%, and 30% of the budget was allocated to medical personnel, medical supplies, and non-medical items, respectively (Basinga et al. 2011). On average in 2008, facilities allocated approximately 77% of their bonus payments to increasing provider salaries and the remainder to the general facility budget (Basinga et al. 2011).

The P4P program was instituted by the Rwandan Ministry of Health (MOH) through a phased, quasi-randomized rollout. In 2006, the country's 19 districts without existing P4P pilot programs were divided into 12 treatment and 7 control districts by block-randomization - P4P was launched in the 12 treatment districts in 2006 and expanded to the 7 control districts in 2008. To isolate the effect of incentive payments from increasing resources, control districts received budget increases equal to the average bonus payments paid out to treatment facilities from the period of 2006 to 2008. The overall magnitude of payments disbursed through the program was large, with bonus payments resulting in average expenditure increases of 22% above 2006 levels (Basinga et al. 2011).

The largest measured impact of Rwanda's P4P program on a rewarded service is associated with institutional deliveries, which have high per unit payments ($4.59) relative to most other rewarded services. Both Basinga et al. (2011) and Sherry, Bauhoff, and Mohanan (2014) find large and statistically significant impacts of P4P on the number of pregnant women delivering in a health care facility, roughly a 27-30% increase over baseline. Since this is the most robust positive finding from Rwanda's P4P program, in this analysis we seek to identify the factors mediating the increase in institutional deliveries.

3.2 Data and Methods

To assess the input differences associated with the Rwandan P4P program, we use data from the 2007 Rwandan Demographic and Health Services Service Provision Assessment Survey (SPA) (National Institute of Statistics (NIC) [Rwanda], Ministry of Health (MOH) [Rwanda], and Macro International Inc. 2008). The SPA instruments are designed to monitor health care systems in developing countries by assessing the availability of services, readiness of facilities to provide these services, adherence to quality protocols and satisfaction with the service provision environment. Key topics include infrastructure, management systems, and provision of services related to maternal and child health, family planning, HIV/AIDS and STIs, and communicable and non-communicable diseases. The 2007 SPA represents a census of all public health facilities, a census of all private facilities with five or more staff, and a sample of private facilities with three or more staff (National Institute of Statistics (NIC) [Rwanda], Ministry of Health (MOH) [Rwanda], and Macro International Inc. 2008). [3]

[3] There is an earlier round of the Rwandan SPA from 2001 with a sample of 222 facilities. We do not use the earlier round in this analysis due to the restricted sample size and confidentiality concerns that preclude facility matching across rounds.

To identify program effects, we leverage the randomized rollout of the program and the post treatment data collection. Assuming the randomization of districts was successful, the differences between treatment and control facilities in the post-period should represent causal effects of the P4P program on health care inputs and provider behavior. Data collection for the DHS SPA occurred from June through October 2007, roughly a year after the program was initially introduced in treatment areas but before it was introduced in control districts. Because of the timing, we can identify medium run input differences between treatment and control areas, but we are unable to observe any longer term differences in investments that do not arise within the first year of program implementation.

Of the 538 facilities included in the full dataset, we exclude facilities drawn from 11 non-random districts in which P4P was piloted prior to 2006 (194 facilities). We also exclude all hospitals (42 facilities), because hospitals were subject to a different incentive scheme under the national P4P program. Furthermore, hospitals could not be classified into treatment categories due to confidentiality concerns.[4] Our analysis sample consists of 201 treatment facilities and 101 control facilities. The sample contains only polyclinics and health posts which serve smaller catchment areas and tend to spend less per capita than hospitals.

Table 3.3 provides basic summary statistics for the analyzed sample, which includes facilities from all five provinces (though only one facility from Kigali City is included in the analyzed sample). 65% of the sampled facilities are publicly managed, while the remaining facilities are classified as private facilities, non-government organizations, or community facilities. The majority of facilities in the Rwandan health care system are polyclinics, which are expected to provide a full range of basic services, i.e. pediatric care, treatment of sexually transmitted infections (STIs), family planning (FP) and antenatal care (ANC). Health posts exist in areas that are far from the main polyclinics and generally offer fewer services. Those services are limited to curative outpatient care, a subset of diagnostic tests, child immunization and growth monitoring, health education, and ANC and FP services (National Institute of Statistics (NIC) [Rwanda], Ministry of Health (MOH) [Rwanda], and Macro International Inc. 2008). The vast majority of the facilities have general outpatient care clinics, while only 5.1% of the facilities have inpatient medical clinics. Polyclinics and health posts have average catchment populations of 21,000 and 17,000 respectively, and there is large variation in per capita spending within each facility type. All public and private facilities, including health posts, were subject to the P4P scheme.

Using this sample, we focus our interest on the most largely and robustly impacted service-institutional deliveries. Because the P4P program allows providers to innovate freely, there is a large set of health care delivery inputs that providers may modify in response to payment incentives. We focus on two categories of inputs classified by their associated services-delivery-specific inputs and general health care delivery inputs. We look first at delivery-specific inputs because economic theory predicts that facilities should allocate re-

[4]Although facility location is excluded from the publicly available dataset due to confidentiality concerns, we obtained facility treatment identifiers from SPA facilitators. Treatment identifiers were not provided for hospitals due to the concerns that the existence of few hospitals within each district could lead to their direct identification.

sources towards inputs that produce the highest marginal increases in incentivized outcomes. In this case, these inputs are presumably those inputs with direct applications to delivery services. We also examine a set of more general inputs as indicators for facility quality. In the SPA report, the MOH of Rwanda identified health care quality as a priority issue and described the P4P program as one of three primary "quality strategies" addressing this issue (National Institute of Statistics (NIC) [Rwanda], Ministry of Health (MOH) [Rwanda], and Macro International Inc. 2008).[5] By providing incentives for a large set of services, the P4P program indirectly encourages investment in general facility quality since the various incentivized services share certain common inputs. General facility quality is also directly incentivized through the P4P quality multiplier, which scales bonus payments for unit services according to performance on domains such as administration and cleanliness.

Within the two categories of delivery-specific and general health facility inputs, we analyze three groups of input types-physical resources, human resources, and management and operational inputs. Input variables are drawn primarily from facility administrator interviews, direct observation of relevant equipment, and administrative reports. We use the guidelines in the Service Availability and Readiness Assessment (SARA) developed by the WHO and the P4P multiplier to help identify important indicators for delivery and facility readiness and quality. These guidelines provide lists of equipment, supplies, and processes required for basic service provision (Health Contractual Approach Unit 2008; World Health Organization 2012). The relevant inputs from these guidelines are provided in Table 3.4.

When possible, we combine sets of related inputs into single indices. Since P4P programs allow facilities to respond idiosyncratically, facilities may adjust differently in accordance with their specific needs. For instance, one facility may need an exam table for delivery while another facility may need an obstetrical stethoscope. In aggregate, then, we may not see systematic differences for specific inputs. By combining related inputs into indices, we look at more general categories that give us a sense of broader investment patterns. These systematic responses may exist even though there is heterogeneity across facilities, since many of the underlying problems facing facilities are the same. For example, the shortage of doctors and nurses throughout the country (National Institute of Statistics (NIC) [Rwanda], Ministry of Health (MOH) [Rwanda], and Macro International Inc. 2008) may lead to similar human resource response strategies.

We weight the components using equal weights unless otherwise indicated in the P4P quality multiplier guidelines.[6] The index weights are presented in Table 3.4. As a note, the correspondence between the variables available in the SPA data and the stated components of the P4P quality multiplier is moderate. For example, one checklist item in the P4P quality multiplier is a monthly inventory of equipment and supplies for each service. This variable is not available in the SPA data, but we use reported routine equipment maintenance as a substitute component. The resulting indices are therefore only proxies for the incentivized

[5]The other strategies include the quality assurance and mutual health insurance programs.

[6]The P4P quality multiplier guidelines prescribe unequal weights for the components of the management indicator and equal weights for all other relevant indices.

components of facility quality.

To determine program impacts, we use linear regressions of the inputs of interest on an indicator for treatment, controlling for fixed facility characteristics. These controls include the log of the catchment population size, log per capita spending, and indicators for public facilities, health posts, provinces, clinic availability, and funding types. Although randomization was done at the district level, we cannot cluster within districts due to confidentiality concerns which exclude district identifiers from the data.

Since we assess a broad set of indicators, we are concerned with issues associated multiple inference. Specifically, when a large number of dependent variables are tested, significant coefficients may arise when the true treatment effects are zero. To address this concern, we adjust our raw p values using a method to control the false discovery rate (FDR) proposed by Benjamini, Krieger, and Yekutieli (2006). FDR control is less conservative than methods such as the Bonferroni correction and can be used in exploratory analyses when the cost of false discoveries is relatively low (Anderson 2008). The procedure for determining statistical significance at level q for M tests is given below. To calculated adjusted p values following this algorithm, we employ the procedure described in Anderson (2008).

1. Sort hypotheses in order of decreasing significance such that $p_1 < p_1 < p_2 < ... < p_M$, and let subscript r denote the rank.

2. Let $q' = q/(1+q)$ and c be the number of hypothesis such that $p_r < q'$.

3. Beginning with p_M, check if $p_r < (q'r)/(M-c)$. If it does, take p_M and all smaller p values as significant. If it does not, step down to the next highest p value and repeat.

In addition to identifying the effect of payment incentives on a variety of health care delivery inputs, we are interested in understanding the linkages between inputs and service provision. Due to the existence of numerous unobserved variables and potential endogeneity, it is impossible to determine the causal links between clinic inputs and service provision. Nevertheless, given the policy importance of identifying successful mechanisms for increasing institutional deliveries, we calculate the mediation effects associated with input changes and discuss these results in light of the likely existence of biases. We follow the method developed by Imai et al. (2011). Compared to the traditional linear structural equation modeling framework popularized by Baron and Kenny (1986), Imai and colleagues provide an estimation strategy with fewer parametric assumptions and additional methods for sensitivity analyses. This method has been employed in the political participation and medical literatures to study mediation effects (Karpowitz, Mendelberg, and Shaker 2012; Varese, Barkus, and Bentall 2012).

The method calculates the average causal mediation effects (ACME) of intermediate indicators on final outcomes under the assumption that there are no unmeasured factors that affect both the mediator and the outcome. This is done by fitting two models, a model of the mediator as a function of treatment and covariates and a model of the outcome as a function of the mediator, treatment, and covariates. We then predict two potential outcomes under treatment. The first is the predicted outcome under treatment when the mediator is also predicted under treatment. The second is the predicted outcome under treatment when the mediator is predicted using control conditions. The causal mediation effects are computed as the difference between the two predicted outcomes. Average causal mediation effects are computed by repeating the prediction step under different values of the model parameters drawn from the parameter distribution estimated in the initial step.

We present the basic equations below, where M_i, T_i, Y_i,and X_i represent the value of the mediator, treatment, outcomes, and covariates for each observation, respectively, and θ_M and θ_Y represent model parameters.

1. Fit parametric models $f_{\theta_M}(M_i|T_i, X_i)$ and $f_{\theta_Y}(Y_i|T_i, M_i, X_i)$.
2. Simulate model parameters by sampling J copies of θ_M and θ_Y.
3. Calculate ACME.
    a) Simulate potential mediator values by drawing K potential mediator values from $f^j_{\theta_M}(M_i|T_i, X_i)$ denoting each as $M_i^{(jk)}(t)$, for each potential mediator value.
    b) Simulate potential outcomes by drawing one potential outcome $Y_i^{(jk)}(t, M_i^{(jk)}(t'))$ for each $f^j_{\theta_Y}(Y_i|T_i, M_i, X_i)$.
    c) Compute the average causal mediation effect for treatment t as

$$\bar{\delta}^j(t) = \frac{1}{nL}\sum_{i=1}^{n}\sum_{k=1}^{K}\left\{Y_i^{(jk)}\left(t, M_i^{(jk)}(1)\right) - Y_i^{(jk)}\left(t, M_i^{(jk)}(0)\right)\right\}$$

We apply this algorithm to the inputs with positive and statistically significant coefficients on the treatment indicator.[7] For example, when analyzing the effect of management on institutional delivery rates, we regress the management index on treatment and covariates and then regress institutional delivery rates on the management index, treatment, and covariates. We then calculate the predicted institutional delivery rate in treatment clinics conditional on the management level expected under treatment. We also calculate the predicted institutional delivery rate in treatment clinics conditional on the management level expected in control areas. We compute the average causal mediation effects by taking the difference between the two predicted institutional delivery rates.

3.3 Results

Before turning to program impacts, we assess the balance between treatment and control facilities on observable characteristics. To assess potential biases, we would like to ensure that treatment and control facilities are balanced on pre-treatment characteristics. Since this

[7] We use the medeff Stata package provided by Imai and colleagues.

data is not available to us, we instead perform balance tests on post-treatment characteristics that are unlikely to have been affected in the first year of the program. Although we can only assess balance across fixed facility characteristics, previous studies have shown that the characteristics of the catchment population as well as the outcomes of interest are also well-balanced in the pre-intervention period.[8] Table 3.5 indicates that facility characteristics that are unlikely to have been affected in the first year of the program are generally comparable across treatment and control clinics with a few exceptions. Among treatment districts, there are more health post facilities, facilities with equity funds for the poor, and facilities from the Northern province, while there are less facilities from the Southern province. We account for these differences by including indicators for health posts, funding types, provinces, and other fixed facility characteristics as controls in the regression analysis.

We first replicate the program impacts on institutional delivery rates using the facility data to address two potential concerns associated with using the SPA data. First, the SPA data was collected only a year after program implementation, whereas the household datasets used in previous studies were collected between 1.5 to 2 years after program implementation (Basinga et al. 2011; Sherry, Bauhoff, and Mohanan 2014). The SPA data therefore observes facilities after a relatively short program exposure period, which could be problematic for identifying impacts if facilities responded slowly to program incentives. Second, due to the exclusion of hospitals, the SPA represents an unrepresentative facility sample. This could raise concerns if program effects were concentrated in hospital facilities.

To identify the program impacts on institutional deliveries, we construct a measure of facility delivery rates by using monthly administrative records for a period of six months and normalizing by the number of births expected in the facility's catchment population.[9] We report the results in Table 3.6. We find that institutional delivery rates in treatment facilities are 31% higher than rates in control facilities (a 12.4 percentage point increase). This difference is statistically significant and consistent with analyses conducted using the reported annual number of deliveries and facility estimates of delivery coverage (results not shown). The magnitude corresponds closely with the estimates from household surveys, allaying potential concerns associated with the SPA data.

We now turn to the program impacts on facility inputs and look first at delivery-specific physical resources, human resources and management indicators. We present one set of regression results for maternity beds per thousand pregnant women in Table 3.7. We run various specifications, with different subsets of controls in each specification. The regression results show an insignificant treatment effect on maternity beds across the specifications. Health posts have fewer maternity beds, while private clinics and clinics with ANC and HIV clinics have more maternity beds. The number of maternity beds per pregnant woman is lower in facilities serving larger catchment areas and higher in facilities with higher per

[8] Prenatal care visits are one exception. Basinga and colleagues find that treatment areas in the pre-intervention period had more women with 4 or more prenatal care visits, but Sherry and colleagues do not find a difference (Basinga et al. 2011; Sherry, Bauhoff, and Mohanan 2014).

[9] We follow the procedure described in WHO's service availability indicators to calculate the expected number of births from the catchment population (World Health Organization 2012).

capita spending. Facilities in the Southern Province have more maternity beds per pregnant woman compared to the Northern Province. The covariate patterns for other indicators are mixed (results not shown), but private clinics tend to do better on most indicators.

For the remaining indicators, we present coefficients on the treatment indicator from regressions corresponding to regression specification (3) in Table 3.7, controlling for log catchment population, log per capita spending and indicators for health posts, provinces, private clinics, clinic availability and funding types. Table 3.8 presents the treatment effects for the analyzed inputs. The indicator levels are low to moderate. On average, control clinics have only 42% of the recommended delivery room inputs and medications, 62% of the recommended sanitation supplies, and 63% of the recommended general facility equipment. Performance on the management and operations indicators is slightly higher, with averages of 80%, 76%, and 67% in control facilities for the delivery statistics monitoring indicator, the clinic cleanliness indicator, and the management indicator, respectively.

To determine the statistical significance of the coefficients on the treatment indicators, we refer to the FDR-adjusted p values and use a threshold of $p \leq 0.10$ since we have relatively low power due to the small sample size. The coefficients are of mixed signs and do not appear to show consistent patterns within each category. Overall, the majority of the indicators that we analyze are statistically insignificant. The two exceptions are the number of staff per capita providing maternity-related services who are present today and the management indicator. In both cases, the coefficients on the P4P program are positive, with adjusted p values of 0.10. When we separately analyze the different components of the management indicator (results not shown), we find that the positive effect is driven primarily by higher use of quality assurance reports in treatment clinics.

Table 3.9 shows the ACMEs following the algorithm by Imai and colleagues described above (Imai et al. 2011). We perform the analysis for the two indicators associated with positive and significant treatment effects-the number of staff per capita providing maternity-related services present today and the management indicator. The estimated relationship between institutional delivery rates and the number of maternity-related staff present each day is negative, resulting in a negative ACME. We find that the management indicator accounts for a small fraction, 1.5% of the overall impact on institutional delivery rates. This estimate has wide confidence intervals, indicating a high level of uncertainty in the linkages between the management indicator and the corresponding institutional delivery rates.[10]

3.4 Discussion

In summary, we find that Rwanda's P4P incentives for institutional deliveries are associated with a positive and significant impact on institutional delivery rates, consistent with the results reported by Basinga, Sherry, and colleagues (Basinga et al. 2011; Sherry, Bauhoff, and Mohanan 2014). Yet, we find little evidence that treated clinics respond to payment

[10] Imai and colleagues also provide an algorithm for a sensitivity analysis of the ACME (Imai et al. 2011). We do not present the analysis since the ACMEs are small and not statistically significant.

incentives by systematically investing in inputs, a number of which are recommended under the WHO guidelines as important measures for basic facility and delivery-specific readiness.

We identify two inputs with positive and significant results-clinic management and staff per capita providing maternity-related services present today. The effect on clinic management is consistent with the program incentives provided through the P4P quality multiplier. At the same time, other indices with P4P quality multiplier components are not associated with positive and statistically significant coefficients, indicating that management may be of particular importance in achieving improvements in service provision. The effect on maternity-related staff present each day is consistent with the idea that P4P programs work by targeting worker productivity and absenteeism (Meessen, Kashala, and Musango 2007) and consistent with reports that the majority of incentive payments were distributed as increased salaries (Basinga et al. 2011). Together, the coefficients on these two indicators suggest the potential importance of management and provider presence, mechanisms which were highlighted in other studies on the impact of P4P programs on provider labor supply and effort (Huillery and Seban 2013; Olken, Onishi, and Wong 2012).

Although we find two inputs that are associated with positive responses to payment incentives, the mediation analysis shows that these input differences explain a small fraction of the difference in institutional delivery rates between treated and control clinics. It is possible, then, that there are other important mechanisms underlying differences in institutional delivery rates that could not been identified in this data.

We now discuss several policy implications and then turn to the limitations of our study. The first implication is that management and provider presence are potentially important for increasing service provision rates. Again, this is consistent with the problems of low provider effort and absenteeism affecting health systems in developing countries (Chaudhury et al. 2006) and with the use of P4P programs as potential solutions. The results suggest that P4P implementation in other health delivery systems could be complemented by trainings on management strategies and strategies for encouraging provider presence. It is important to note, though, that modifying management or provider absenteeism without payment incentives will not necessarily result in the same increases in outcomes if the key feature of P4P programs is the conditionality.

Second, we find that providers do not appear to systematically respond to incentives by investing in additional physical and human resources. This implies that these inputs may not be binding constraints in Rwanda despite the low to moderate overall levels, and overall levels of physical and human resources appear less important for achieving increases in institutional delivery rates. We note, though, that these inputs may be important constraints in other contexts or for other health services. Furthermore, the lack of differences in the set of analyzed indicators implies that P4P programs may not be very effective at improving the strength of the general health care delivery system as measured by input levels. Provider responses to incentives do not appear to extend to inputs which are not directly incentivized or are not closely associated with the contracted service. In contexts where input constraints are identified as priority issues, direct incentives or other policy levers may be necessary to attain higher input levels. For example, additional efforts may be needed to address the overall shortage of doctors and nurses in Rwanda.

We now discuss the limitations of our analysis. First, our approach can only detect behavior that is systematic across providers. Fundamentally P4P allows providers to choose their own approach, so it is possible that providers had idiosyncratic responses to the P4P. The lack of many systematic results may point to substantial facility heterogeneity. Second, we can only observe a subset of health care delivery inputs in the SPA facility surveys. It is possible that there were systematic provider responses along other important dimensions such as provider effort or client recruitment through community meetings. Moreover, without detailed budget information, we are unable to explicitly trace the distribution of payment bonuses. These data limitations highlight potentially valuable additions to the SPA survey instruments, which are used in multiple countries for health care delivery monitoring. Third, the large number of insignificant results could be due to low sample power. Although we cannot rule this out, we do not see consistent patterns in the direction of treatment effects by category, so it does not appear to be the case that input responses are systematic. Finally, it is possible that we find few differences between treatment and control clinics because control facilities are forward-looking and adjust inputs in anticipation of receiving the P4P program. Though this may be the case, control clinics have not adjusted in such a way to achieve institutional delivery rates as high as those in treatment facilities, indicating that some differences between control and treatment clinics likely remain.

Despite the limitations of our analysis, this study adds to the existing literature on P4P programs in developing countries by identifying the input responses associated with payment incentives distributed at the facility level. Much of the “black box” behind provider responses remains unknown, though, warranting further work on the mechanisms tied to improved health outputs and outcomes.

In conclusion, this dissertation provides some theoretical, empirical, and policy-relevant insights into the challenges of accurately measuring poverty and improving the quality of health care delivery systems. This work brings extensive data analyses and a more complete incorporation of economic theory to address practical measurement issues, and it provides an exploration into the mechanisms underlying a welfare improvement program in one country. The steps taken here represent small but significant contributions to the knowledge on welfare improvement in developing countries and call for continued work on broadening this evidence base.

3.5 Tables

Table 3.1: P4P program description: Rewarded services and unit payments

Service	Payment (US $)	Service	Payment (US $)
Primary Care			
Curative care visits	0.18	First-time family planning visits	1.83
Emergency referrals during curative treatment	1.83	1-month contraceptive resupply	0.18
Maternal Health			
First prenatal care visits	0.09	Deliveries in the facility	4.59
Women who completed 4 prenatal care visits	0.37	Emergency transfers to hospital for obstetric care during delivery	4.59
Women who received appropriate tetanus vaccination during prenatal care	0.46	At-risk pregnancies referred to hospital for delivery during prenatal care	1.83
Women who received 2nd dose of malaria prophylaxis during prenatal care	0.46		
Child Health			
Child (0-59 months) preventive care visits	0.18	Children who completed vaccination on time	1.83
Malnourished children referred for treatment during preventive care visit	1.83		
HIV/AIDS			
Voluntary counseling and testing	0.89	New adult clients put on ARVs	4.58
Prevention of mother-to-child-transmission (PMTCT): partner tested	4.58	HIV+ clients treated with cotrimoxazole each month	0.44
PMTCT: exposed children tested	8.93	New pediatric clients put on ARVs	6.70
PMTCT: women under treatment with ARVs during labor	4.58	HIV+ women who use modern method of family planning	2.68
HIV+ clients tested for CD4 count	4.58	HIV+ clients tested for TB	2.68

Notes: Payments reported for each unit.

Table 3.2: P4P program description: Quality index multiplier

Service Category	Quality Index Weight	Share of Weight: Structural Factors	Share of Weight: Processes
General Administration	0.052	1.00	0.00
Cleanliness	0.028	1.00	0.00
Curative Care	0.170	0.23	0.77
Delivery	0.130	0.40	0.60
Prenatal care	0.126	0.12	0.88
Family Planning	0.114	0.22	0.78
Immunization	0.070	0.40	0.60
Preventive Care	0.052	0.15	0.85
HIV Services	0.090	1.00	0.00
TB Services	0.028	0.28	0.72
Laboratory Services	0.030	1.00	0.00
Pharmacy Management	0.060	1.00	0.00
Financial Management	0.050	1.00	0.00
Total	1.000		

Table 3.3: Sample summary statistics

	Polyclinic	Health post	Total
Number of facilities			
Management type			
Public	185	12	197
Private/NGO/community	71	34	105
Province			
Northern	60	9	69
Southern	50	3	53
Eastern	78	12	90
Western	67	22	89
Kigali City	1	0	1
Clinic availability			
Has general outpatient clinic	249	46	295
Has inpatient medical clinic	16	0	16
Group means and standard deviations			
Catchment population	21,088	16,856	20,738
	[11,042]	[31,674]	[13,887]
Spending/catchment population	2,696	10,210	3,327
	[17,235]	[29,308]	[18,583]

Data from hospitals excluded from remaining analysis due confidentiality concerns.
Standard deviations in square brackets.

Table 3.4: Indicator components from P4P multiplier and WHO guidelines

Component	Source	Weight	Component	Source	Weight
Delivery room inputs indicator					
Exam table	P4P	0.06	Sterile gloves	P4P	0.06
Exam light	P4P	0.06	Umbilical cord clamp	P4P	0.06
Infant scale	P4P	0.06	Skin disinfectant	WHO	0.06
Sterilized instruments	P4P	0.06	Injectable diazepam	WHO	0.06
Neonatal aspirator	P4P	0.06	IV solution with infusion set	WHO	0.06
Obstetrical stethoscope	P4P	0.06	Injectable antibiotic	WHO	0.06
Suture thread	P4P	0.06	Injectable magnesium sulphate	WHO	0.06
Ophthalmic ointment	P4P	0.06	Injectable uterotonic	WHO	0.06
Local anesthesia	P4P	0.06			
Sanitation supplies indicator					
Soap availability (averaged across clinics)	P4P	0.50	Disinfectant availability (averaged across clinics)	P4P	0.50
Basic equipment indicator					
Any source of electricity	Authors	0.17	Functioning incinerator	P4P	0.17
Water dispensers available (averaged across clinics)	P4P	0.17	Functioning sterilizing equipment	P4P	0.17
Clean water source available	P4P	0.17	Beds per 10,000 people (N/25, max 1)	WHO	0.17
Delivery statistics monitoring indicator					
Evidence of delivery statistics monitoring	Authors	0.33	Meetings to discuss adverse deliveries	Authors	0.33
Meetings to discuss delivery statistics	Authors	0.33			
Management indicator					
Services available full time	P4P	0.11	Covered waiting areas	P4P	0.17
Minutes for monthly management meetings	P4P	0.17	Minutes for monthly community meetings	P4P	0.17
Medical systems reports observed	P4P	0.17	Routine equipment maintenance	P4P	0.06
Quality assurance reports	P4P	0.17			

Notes: P4P source refers to components of P4P quality multiplier. WHO guidelines taken from World Health Organization (2012) Only components available in SPA data are listed.

Table 3.5: Facility characteristics: balance tests

	Control mean	Difference	P-value
General			
Catchment population	20492	376	.830
Health post	.07	.12	.004
Private/NGO/community	.33	.03	.589
Spending per capita in catchment area, 2006	1558	2681	.281
Province			
Eastern	.32	-.03	.612
Kigali City	.01	-.01	.158
Northern	.13	.15	.003
Southern	.23	-.08	.091
Western	.32	-.03	.55
Clinic availability			
Has ANC clinic	.18	-.01	.759
Has general outpatient clinic	.96	.02	.179
Has VCT/HIV/special diagnoses (incl HIV) clinic	.21	-.03	.546
Has inpatient medical clinic	.05	.01	.848
Has inpatient/outpatient TB clinic	.11	-.05	.087
Funding sources			
Employer	.06	.05	.157
Equity fund for poor	.15	.09	.083
Insurance	.56	-.09	.155
User fees only	.12	.07	.147
Other	.06	-.01	.581
Government risk pool	.7	-.09	.14

Table 3.6: Treatment effects: Institutionalized deliveries across datasets

	Mean (control)	β	se	p-val	N	Data
Sherry, Bauhoff, Mohanan	0.30	0.098	0.036		6144	Individual
Basinga, et al.	0.36	0.074		0.033	2108	Individual
SPA facility data	0.40	0.124	0.037	0.001	302	Facility

Number of pregnancies calculated for SPA data from catchment population using WHO service availability indicator guidelines. Mean control reported for pretreatment year in Sherry and Basinga.

Table 3.7: Regression results: Maternity beds per 1000 pregnant women

	(1)	(2)	(3)	(4)
Treatment	0.329	0.248	0.874	0.443
	(1.088)	(1.114)	(1.169)	(1.347)
Health post		-5.324*	-6.080*	-6.621**
		(2.853)	(3.115)	(3.229)
Private/NGO/community (x100)		3.401***	3.858***	2.860
		(1.251)	(1.338)	(2.041)
Log catchment population		-5.032***	-5.586***	-5.664***
		(1.157)	(1.168)	(1.176)
Log per capita spending		1.007**	0.781	0.790
		(0.494)	(0.497)	(0.497)
Southern province		2.671*	4.247**	4.220**
		(1.577)	(1.915)	(1.918)
Eastern province		0.370	0.374	0.391
		(1.439)	(1.527)	(1.530)
Western province		-1.760	-1.845	-1.758
		(1.505)	(1.526)	(1.535)
Has general outpatient clinic			2.194	1.901
			(3.263)	(3.299)
Has ANC clinic			1.310	1.277
			(1.418)	(1.421)
Has inpatient/outpatient TB clinic			0.301	0.353
			(1.758)	(1.763)
Has VCT/HIV/special diagnoses (incl HIV) clinic			4.048***	4.100***
			(1.340)	(1.345)
Has inpatient medical clinic			-4.722**	-4.992**
			(2.369)	(2.409)
Funding: equity fund for poor			-1.621	-1.566
			(1.425)	(1.430)
Funding: employer			0.972	0.937
			(1.769)	(1.773)
Funding: insurance			0.305	0.272
			(1.150)	(1.153)
Funding: gov't risk pool (various)			-0.861	-0.793
			(1.368)	(1.375)
Treatment X Private				1.652
				(2.551)
Constant	10.96***	53.05***	56.85***	58.07***
	(0.879)	(12.73)	(12.97)	(13.12)
Observations	242	210	210	210

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

Table 3.8: Treatment effects

	Mean (cont.)	β	se	p-val	FDR adj. p-val	N
Physical Resources						
Delivery-specific						
Maternity beds/1,000 pregnant women	10.96	0.87	1.17	0.46	0.41	210
Emergency transport for obstetric emergencies	0.13	-0.08	0.04	0.03	0.11	225
Delivery equipment and medication indicator†	0.42	0.04	0.04	0.24	0.26	207
General						
Sanitation supplies indicator†	0.62	-0.01	0.03	0.81	0.53	240
Basic equipment indicator†	0.63	-0.05	0.03	0.16	0.20	186
Human Resources						
Delivery-specific						
No. staff providing maternity-related services present today/10,000 people	0.97	0.34	0.13	0.01	0.10	240
Has community health worker, delivery	0.78	0.07	0.05	0.17	0.20	240
No. midwives/10,000 people	0.03	-0.01	0.01	0.69	0.47	240
General						
No. staff/10,000 people	9.95	-1.31	0.62	0.04	0.11	240
No. staff present today/10,000 people	5.74	0.70	0.61	0.26	0.26	240
Management and Operations						
Delivery-specific						
ANC services: days per month provided	7.37	0.54	0.58	0.35	0.34	234
24 hour delivery coverage provided at facility	1.00	0.00				210
Delivery stats monitoring indicator†	0.80	-0.08	0.04	0.04	0.11	209
General						
Cleanliness of clinics (averaged across clinics)	0.76	0.04	0.02	0.05	0.12	240
Management indicator†	0.67	0.06	0.02	0.01	0.10	240

† Indicators described in Table 3.4. Regressions controls include: log catchment population, log per capita spending,and indicators for managing authority, province, health post, clinic types, and funding types.

Table 3.9: Mediating increases in institutional deliveries

	Average Causal Mediation Effects†	90% Confidence Interval		Fraction of Total Effect Mediated
No. staff providing ANC/FP/delivery/10,000 people (to-day)	-0.004	-0.016	0.004	-0.035
Management indicator	0.002	-0.009	0.013	0.015

† Total treatment effect on institutional deliveries = 0.114

Bibliography

Acham, Hedwig et al. (2012). “Breakfast, midday meals and academic achievement in rural primary schools in Uganda: implications for education and school health policy”. In: *Food & nutrition research* 56.

Alderman, Harold (2002). “Do local officials know something we don’t? Decentralization of targeted transfers in Albania”. In: *Journal of Public Economics* 83.3, pp. 375–404.

Anderson, Michael L (2008). “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects”. In: *Journal of the American statistical Association* 103.484.

Baron, Reuben M and David A Kenny (1986). “The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations”. In: *Journal of Personality and Social Psychology* 51.6, pp. 1173–82.

Basinga, Paulin et al. (2011). “Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation”. In: *The Lancet* 377.9775, pp. 1421–1428.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli (2006). “Adaptive linear step-up procedures that control the false discovery rate”. In: *Biometrika* 93.3, pp. 491–507.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). “Automobile prices in market equilibrium”. In: *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Blundell, Richard and Thomas MaCurdy (1999). “Labor supply: A review of alternative approaches”. In: *Handbook of labor economics* 3, pp. 1559–1695.

Bollen, Kenneth A, Jennifer L Glanville, and Guy Stecklov (2001). “Socioeconomic status and class in studies of fertility and health in developing countries”. In: *Annual Review of Sociology*, pp. 153–185.

Bollen, Kenneth A, David K Guilkey, and Thomas A Mroz (1995). “Binary outcomes and endogenous explanatory variables: tests and solutions with an application to the demand for contraceptive use in Tunisia”. In: *Demography* 32.1, pp. 111–131.

Bonus, Holger (1973). “Quasi-Engel curves, diffusion, and the ownership of major consumer durables”. In: *Journal of Political Economy* 81.3, pp. 655–77.

Caldas, Marcellus et al. (2007). “Theorizing land cover and land use change: The peasant economy of Amazonian deforestation”. In: *Annals of the Association of American Geographers* 97.1, pp. 86–110.

Caldés, Natàlia, David Coady, and John A Maluccio (2006). “The cost of poverty alleviation transfer programs: a comparative analysis of three programs in Latin America”. In: *World Development* 34.5, pp. 818–837.

Carranza, Juan Esteban (2007). “Estimation of demand for differentiated durable goods”. In: *V manuscript, University of Wisconsin'Madison.*

Chaudhury, Nazmul et al. (2006). “Missing in action: teacher and health worker absence in developing countries”. In: *The Journal of Economic Perspectives* 20.1, pp. 91–116.

Dargent-Molina, Patricia et al. (1994). “Association between maternal education and infant diarrhea in different household and community environments of Cebu, Philippines”. In: *Social Science & Medicine* 38.2, pp. 343–350.

Deaton, Angus and John Muellbauer (1988). *Economics and consumer behavior, 1980.*

Doocy, Shannon et al. (2005). “Outcomes of an Ethiopian microfinance program and management actions to improve services”. In: *Journal of Microfinance/ESR Review* 7.1, pp. 79–95.

Ferguson, BD et al. (2003). “Estimating permanent income using indicator variables”. In: *Health systems performance assessment: debates, methods and empiricism. Geneva: World Health Organization*, pp. 747–760.

Filmer, Deon and Lant H Pritchett (2001). “Estimating wealth effects without expenditure DataOr tears: An application to educational enrollments in states of india*”. In: *Demography* 38.1, pp. 115–132.

Filmer, Deon and Kinnon Scott (2012). “Assessing asset indices”. In: *Demography* 49.1, pp. 359–392.

Gertler, Paul and Christel Vermeersch (2012). “Using Performance Incentives to Improve Health Outcomes”. In: *World Bank Policy Research Working Paper, Impact Evaluation Series* 6100.60.

Goldberg, Pinelopi Koujianou (1995). “Product differentiation and oligopoly in international markets: The case of the US automobile industry”. In: *Econometrica: Journal of the Econometric Society*, pp. 891–951.

Gorbach, Pamina M et al. (1998). “Contraception and abortion in two Vietnamese communes.” In: *American journal of public health* 88.4, pp. 660–663.

Guilkey, David K and Susan Jayne (1997). “Fertility transition in Zimbabwe: Determinants of contraceptive use and method choice”. In: *Population Studies* 51.2, pp. 173–189.

Gunning, Jan Willem et al. (2000). “Revisiting forever gained: income dynamics in the resettlement areas of Zimbabwe, 1983–96”. In: *The Journal of Development Studies* 36.6, pp. 131–154.

Hanemann, W Michael (1984). “Discrete/continuous models of consumer demand”. In: *Econometrica: Journal of the Econometric Society*, pp. 541–561.

Havanon, Napaporn, John Knodel, and Werasit Sittitrai (1992). “The impact of family size on wealth accumulation in rural Thailand”. In: *Population Studies* 46.1, pp. 37–51.

Health Contractual Approach Unit, Ministry of (2008). *Quarterly Quality Assessment Grid for Health Centers.* Tech. rep.

Holmes, Christopher M (2003). "Assessing the perceived utility of wood resources in a protected area of Western Tanzania". In: *Biological Conservation* 111.2, pp. 179–189.

Houweling, Tanja AJ, Anton E Kunst, and Johan P Mackenbach (2003). "Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter?" In: *International journal for equity in health* 2.1, p. 8.

Howe, Laura D, James R Hargreaves, and Sharon RA Huttly (2008). "Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries". In: *Emerging Themes in Epidemiology* 5, p. 3.

Howe, Laura D et al. (2009). "Is the wealth index a proxy for consumption expenditure? A systematic review". In: *Journal of Epidemiology and Community Health* 63.11, pp. 871–877.

Howe, Laura D et al. (2012). "Measuring socio-economic position for epidemiological studies in low-and middle-income countries: a methods of measurement in epidemiology paper". In: *International journal of epidemiology* 41.3, pp. 871–886.

Huillery, Elise and Juliette Seban (2013). "Performance-Based Financing for Health: Experimental Evidence from the Democratic Republic of Congo". In: *Working paper*.

Imai, Kosuke et al. (2011). "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies". In: *American Political Science Review* 105.4, pp. 765–789.

Jensen, Eric R (1996). "The fertility impact of alternative family planning distribution channels in Indonesia". In: *Demography* 33.2, pp. 153–165.

Karpowitz, Christopher F, Tali Mendelberg, and Lee Shaker (2012). "Gender inequality in deliberative participation". In: *American Political Science Review* 106.03, pp. 533–547.

Lyte, Richard et al. (2001). "Persistent and consistent poverty in the 1994 and 1995 waves of the European Community Household Panel Survey". In: *Review of income and wealth* 47.4, pp. 427–449.

Mark, Steven D, Donald G Thomas, and Adriano Decarli (1996). "Measurement of exposure to nutrients: an approach to the selection of informative foods". In: *American journal of epidemiology* 143.5, pp. 514–521.

Meessen, Bruno, Jean-Pierre I Kashala, and Laurent Musango (2007). "Output-based payment to boost staff productivity in public health centres: contracting in Kabutare district, Rwanda". In: *Bulletin of the World Health Organization* 85.2, pp. 108–115.

Miller, Grant and Kimberly Singer Babiarz (2013). *Pay-for-performance incentives in low- and middle-income country health programs*. Tech. rep. National Bureau of Economic Research.

Montgomery, Mark R and Paul C Hewett (2005). "Urban poverty and health in developing countries: Household and neighborhood effects". In: *Demography* 42.3, pp. 397–425.

Montgomery, Mark R et al. (2000). "Measuring living standards with proxy variables". In: *Demography* 37.2, pp. 155–174.

Morris, Saul S et al. (2000). "Validity of rapid estimates of household wealth and income for health surveys in rural Africa". In: *Journal of epidemiology and community health* 54.5, pp. 381–387.

National Institute of Statistics (NIC) [Rwanda], Ministry of Health (MOH) [Rwanda], and Macro International Inc. (2008). *Rwanda Service Provision Assessment Survey 2007*. Tech. rep. NIS, MOH, Macro International Inc.

Navajas, Sergio et al. (2000). "Microcredit and the Poorest of the Poor: Theory and Evidence from Bolivia". In: *World development* 28.2, pp. 333–346.

Olken, Benjamin A, Junko Onishi, and Susan Wong (2012). *Should Aid Reward Performance? Evidence from a field experiment on health and education in Indonesia*. Tech. rep. National Bureau of Economic Research.

Peabody, John et al. (2011). "Financial incentives and measurement improved physicians quality of care in the Philippines". In: *Health Affairs* 30.4, pp. 773–781.

Razzaque, Abdur et al. (1990). "Sustained effects of the 1974–5 famine on infant and child mortality in a rural area of Bangladesh". In: *Population Studies* 44.1, pp. 145–154.

Sahn, David E and David Stifel (2003). "Exploring alternative measures of welfare in the absence of expenditure data". In: *Review of income and wealth* 49.4, pp. 463–489.

Sahn, David E and David C Stifel (2000). "Poverty comparisons over time and across countries in Africa". In: *World development* 28.12, pp. 2123–2155.

Schady, Norbert and Maria Caridad Araujo (2008). "Cash transfers, conditions, and school enrollment in Ecuador". In: *Economía* 8.2, pp. 43–70.

Sender, John (2002). "Womens struggle to escape rural poverty in South Africa". In: *Journal of Agrarian Change* 2.1, pp. 1–49.

Sherry, Tisamarie B, Sebastian Bauhoff, and Manoj Mohanan (2014). "Paying for Performance in Health Care: Results from the Randomized Roll-Out of Rwanda's National Program". In: *Working Paper*.

Simmons, Cynthia et al. (2010). "Doing it for themselves: Direct action land reform in the Brazilian Amazon". In: *World Development* 38.3, pp. 429–444.

Singh, Prakarsh (2011). "Performance Pay and Information: Reducing Child Malnutrition in Urban Slums". In: *MPRA Working paper*.

Stifel, David and Luc Christiaensen (2007). "Tracking poverty over time in the absence of comparable consumption data". In: *The World Bank Economic Review* 21.2, pp. 317–341.

Varese, F, E Barkus, and RP Bentall (2012). "Dissociation mediates the relationship between childhood trauma and hallucination-proneness". In: *Psychological medicine* 42.5, p. 1025.

Vyas, Seema and Lilani Kumaranayake (2006). "Constructing socio-economic status indices: how to use principal components analysis". In: *Health policy and planning* 21.6, pp. 459–468.

Wagstaff, Adam and Naoko Watanabe (2003). "What difference does the choice of SES make in health inequality measurement?" In: *Health economics* 12.10, pp. 885–890.

World Health Organization (2012). *Measuring Service Availability and Readiness: A health facility assessment methodology for monitoring health system strengthening*. Tech. rep. World Health Organization.

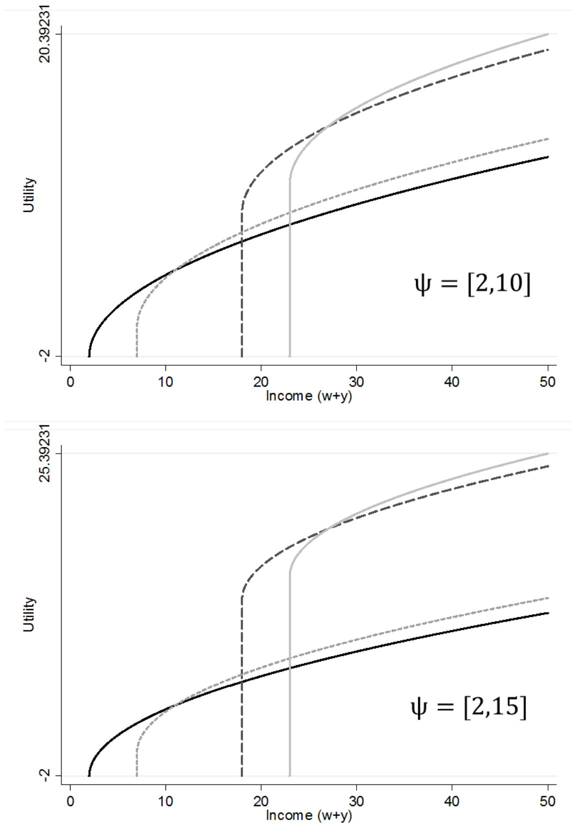
Appendix A

Model estimation details

Defining constraints

Different ψ vectors may result in the same log likelihood measures, resulting in a solution set as opposed to a point estimate when these parameters best fit the data. Such a case is illustrated in Figure A.1; in this case, the utility parameter associated with the more expensive good is so high ($\psi_2 = 10$) that the bundle is optimal at exactly the income at which it is affordable. Increasing the parameter to $\psi_2 = 15$ does not change the income intervals over which each bundle is optimal; the log likelihood measures associated with these two different parameter vectors are therefore identical.

Figure A.1: Alternate parameters with identical log likelihoods



To guarantee unique point estimate solutions to the problem, I define a set of upper bounds on the marginal utility parameters ψ_j. These maximum values are defined as the minimum values at which the relevant bundle dominates all other bundles in the relevant comparison set. To formalize this, I define the relevant bundle for calculating the maximum ψ_j as $x_{j,only}$, the bundle with only good j. I sort the discrete goods by lowest to highest price, and define the relevant comparison set as the set of all bundles which do not include good j or any goods more expensive than good j. As an example, when $j = 3$, there are four bundles in the relevant comparison set defined by all the potential combinations of the two least expensive goods.

$$x_{j,only} \equiv [0, 0, \underbrace{1}_{j}, ..., 0]$$

Set $X_{j-} \equiv$ all bundles where $x_n = 0 \; \forall j \leq n \leq J$ For example, when $j = 3$,

$$X_{j-} = \begin{Bmatrix} (0,0,0,...,0) \\ (1,0,0,...,0) \\ (0,1,0,...,0) \\ (1,1,0,...,0) \end{Bmatrix}$$

The value at which bundle $x_{j,only}$ dominates all the bundles in set X_{j-} depends on whether the bundles in X_{j-} are more or less expensive than bundle $x_{j,only}$. If all the bundles in the relevant comparison set are less expensive, $\psi_{j,max}$ is given by the value which causes bundle $x_{j,only}$ to be optimal at the income level at which the bundle becomes affordable. If some of the bundles in the relevant comparison set are more expensive, $\psi_{j,max}$ is given by the value which causes bundle $x_{j,only}$ to just dominate the relevant bundles when income is infinite. The relevant upper bound for each good in a time series estimation is given by the maximum of the upper bounds defined within each round.

If $\max\left(p_1 x_1^k + ... + p_J x_J^k | x^k \in X_{j-}\right) \leq p_j$,

$$\psi_{j,max} = \max\left(\left[\frac{(p_j - (p_1 x_1^k + ... + p_J x_J^k))^{1-\gamma} - 1}{1-\gamma} + (\psi_1 x_1^k + ... + \psi_J x_J^k) \right] - \underbrace{\frac{-1}{1-\gamma}}_{min\ utility} \middle| x^k \in X_{j-} \right)$$

If $\max\left(p_1 x_1^k + ... + p_J x_J^k | x^k \in X_{j-}\right) > p_j$

$$\begin{aligned} \psi_{j,max} = \max\Bigg(& \left[\frac{(y_\infty - (p_1 x_1^k + ... + p_J x_J^k))^{1-\gamma} - 1}{1-\gamma} + (\psi_1 x_1^k + ... + \psi_J x_J^k) \right] \\ & - \left[\frac{(y_\infty - p_j)^{1-\gamma} - 1}{1-\gamma} \right] \Bigg| x^k \in X_{j-} \Bigg) \end{aligned}$$

Estimation assumptions

To estimate the model, I assume an underlying theoretical distribution for income and normalize prices under that income distribution. Specifically, I assume that income is distributed lognormally in all time periods with $\mu = 0$ and $\sigma = 1$ and use the mean of annual per capita expenditure to rescale the prices accordingly. Although this normalization requires external information on summary statistics on income or expenditure in the population of interest, crude approximations will suffice. Alternatively, an arbitrary price normalization setting the minimum price across surveys to 0.05 results in robust marginal utility orderings. Under the arbitrary price normalization, I set prices to guarantee that the proportion of households owning each asset is not limited by affordability. If the price normalization was set such that the minimum price was too high, any model parameters would be unable to rationalize data where ownership exceeds the fraction of households able to afford the goods. For instance, if the minimum price was set to be at the tenth percentile of the income distribution, the model would always predict that at least 10% of households own no assets, regardless of any variation in model parameters. On the other hand, if prices are set low, such that more households can theoretically afford each good than observed in the data, the marginal utility parameters can adjust accordingly.

I operationalize the maximum likelihood estimation in Matlab by directly computing the indirect utility functions associated with every available bundle for each potential parameter set. The indirect utility functions are defined over 1000 points spanning the underlying income distribution; at this accuracy, I cannot detect bundles which are optimal for less than 0.1% of the population. I operationalize the search with a combination of a user-defined grid search and the fmincon function. Because of the large set (2^J) of potential bundles, the estimation is computationally intensive and estimation time increases greatly with the inclusion of additional discrete goods.

With a relatively small set of J utility parameters explaining an exponentially larger set of potential bundles, there are cases where there is no set of ψ_j parameters that can rationalize observing all of the bundles appearing in the data. This occurs when the indirect utility function associated with a given bundle is dominated by the indirect utility function of another bundle or set of bundles across the entire income distribution. The log likelihood of observing such a bundle when the model predicts that it is never optimal is negative infinity. To operationalize the maximum likelihood search, I replace the log likelihood of bundles with zero probability with negative one million.[1]

Because of the discrete choice nature of the problem, the log likelihood function is not smooth, as a small increase in a marginal utility parameter may increase the log likelihood of observing a bundle from negative infinity to a finite number, resulting in a discrete decrease in the associated likelihood of another bundle. As such, I run the maximum likelihood search using a set of different starting points which span the parameter space.

The marginal utility parameters estimated from the algorithmn generate results in utils. I rescale the marginal utility parameters into dollars using the unconditional indirect utility function associated with the estimated parameters. Specifically, utils are rescaled into dollars by dividing by the mean marginal utility per dollar (averaged across the distribution).

[1]This issue is currently being addressed by estimating a model which incorporates a household level random effect.