J Popul Econ (2005) 18:229–260
DOI 10.1007/s00148-005-0224-7

Journal of
Population Economics

Measuring inequality with asset indicators

David J. McKenzie

Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072, USA
(Fax: +1-650-725-5702; e-mail: mcken@stanford.edu)

Received: 31 October 2003/Accepted: 2 August 2004

Abstract. This paper examines whether, in the absence of information on household income or consumption, data on household infrastructure, building materials, and ownership of certain durable assets can be used to measure inequality in living standards. Principal components analysis is used to obtain a relative measure of inequality, and a bootstrap prediction method is provided for use when auxiliary surveys are available. Mexican data is used to show that the inequality methods provided do provide reasonable proxies for inequalities in living standards. An application finds that after controlling for household income and demographics, school attendance of boys in Mexico is negatively related to state-level inequality.

JEL classification: D31, C81, J10

Key words: Inequality, asset indicators, measurement

1. Introduction

The purpose of this paper is to determine whether, in the absence of information on household income or consumption, data on household infrastructure, building materials, and ownership of certain durable assets can be used to measure inequality in living standards. Data in this format are the predominant source of comparable and nationally representative survey information on fertility, health, mortality and other demographic factors in developing countries. For example, the Demographic and Health Surveys (DHS) now

I thank Hugo Ñopo, two anonymous referees, and participants at the VI Meetings of the LACEA/IDB/World Bank Network on Inequality and Poverty in Puebla, Mexico for useful comments. Alberto Diaz-Cayeros provided helpful insight into political economy of Mexico's educational funding system.

cover more than 170 surveys in 70 countries throughout the developing world.[1] Apart from a few experimental modules, these surveys do not contain income and consumption information, but do contain data on asset indicators along with detailed demographic questions. The World Health Survey (WHS) is a major World Health Organization survey of health conditions and the coverage and responsiveness of health systems.[2] This survey purposively asked about asset ownership rather than income in order to get at a comparable measure of permanent income across countries. In addition, the major Latin American migration surveys take this form: both the Mexican Migration Project (MMP) and subsequent Latin American Migration Project (LAMP) contain asset indicator information along with detailed migration questions, but not income or consumption.[3] Data on asset indicators but not income or consumption is also a feature of some more general social and political surveys in the developing and transition world. For example, the South Eastern Europe Public Agenda Survey conducted by International IDEA has detailed questions on attitudes to democracy, political participation, trust in national and international institutions and the role of the media, along with asset indicators for nine South Eastern European countries.[4]

An index consisting of the first principal component of such asset indicators has recently been shown by Filmer and Pritchett (2001) and Minujin and Bang (2002) to provide reasonable estimates of wealth *level* effects. Their approach has been used with the DHS to examine, for example, how educational attainment differs within countries with wealth (Filmer and Pritchett 1999). Such asset indicators have been used in practice to identify marginal communities and poor households in Mexico's PROGRESA program.[5] Measurement of inequality using a principal-component asset index will open up further new avenues for research given the wide range of existing developing country datasets which contain asset information, but not consumption and income. Measuring inequality in this context is shown to face additional challenges from measuring levels, which this paper seeks to address.

The first challenge is that, by definition, the first principal component asset index has mean zero, and takes negative values for some households. As a result, many standard measures of inequality are not well-defined. We instead define a relative measure of inequality, defined as the standard deviation of the first principal component in a given community of interest relative to the standard deviation in the sample as a whole. Secondly, the asset index may suffer from clumping and truncation issues, which can affect inequality measures much more than measures of levels of well-being. An index could potentially be a good proxy for wealth but provide poor measures of inequality. Graphing the probability density function of the asset index can be used to determine how likely these issues are to be a problem. Thirdly, the relationship between the asset index and non-durable consumption is likely to be monotone, but non-linear. Inequality in asset holdings will therefore differ from inequality in consumption. If the interest lies in consumption inequality, we show how auxiliary surveys which have data on consumption and assets, but not on the variables of interest (such as health or education) can be used to predict consumption inequality in the main survey containing the variables of interest and asset indicators, but not consumption or income. We outline three different approaches which could be used to predict non-durable consumption inequality in these circumstances, and find most support for a bootstrap prediction method which takes account of heteroskedasticity in the prediction error and results in the most accurate predictions amongst the methods studied.

It is argued that there are a number of theoretical questions of interest in which wealth inequality is more important than consumption or income inequality, so an asset-based inequality measure may be preferred in empirical tests of these theories. Moreover, the use of asset indices avoids many of the problems of recall bias, seasonality and mismeasurement that can occur with income and consumption based measures of inequality. There are therefore both theoretical and practical reasons why an asset indicator approach to inequality should be of interest.

To determine how well the various methods work and to validate our approach, we use data from 31 Mexican states and the Federal District from Mexico's national household income and expenditure survey, which has information on income, consumption, and asset indicators. Our relative inequality measure based on asset indicators is found to have a high, positive, and significant correlation with inequality in non-durable consumption. Inequality measured with an asset index based on 30 indicators has as high a correlation as food inequality does with non-durable consumption inequality. The auxiliary survey approach results in even higher correlations with non-durable consumption inequality, and suggests the methods of this paper work well in proxying inequality in living standards.

To illustrate our method in practice, we examine the effect of inequality across states in Mexico on school attendance of 14–18 year olds. Conditional on household income and demographic characteristics, inequality might affect school attendance through a political economy channel of state expenditure on education, or through its effect on the returns to schooling. Both the asset index based measures of inequality and inequality in non-durable consumption give that boys in states with higher inequality have a lower probability of attending schools, but that there is no significant effect on the attendance of girls. The fact that we obtain similar conclusions from both our asset indicator method and from using non-durable consumption provides further evidence that the approach here provides reasonable proxies of inequality in living standards.

The remainder of the paper is structured as follows. Section 2 outlines the principal components method and construction of a relative inequality measure based on the first principal component. Advantages and disadvantages of this approach to inequality measurement are also discussed. Section 3 then provides several methods for predicting inequality in non-durable consumption from asset indicators given an auxiliary survey. Section 4 summarizes the Mexican data used and shows that the asset index is able to distinguish differences in levels of well-being across states. Section 5 examines the performance of the various inequality measures, while Sect. 6 provides some robustness checks. The application to school attendance and inequality is provided in Sect. 7 while Sect. 8 concludes.

2. Estimating inequality from asset indicators without auxiliary surveys

Let $\mathbf{x} = (x_1, x_2, \ldots, x_P)'$ be a vector of asset indicators. Most of these will be dummy variables for ownership of specific assets (e.g., owning a fridge or a television) or for characteristics of the housing structure (e.g., walls made of brick, dirt floor), however non-binary variables such as the number of rooms in the house could also be included. The question of interest is how to construct a measure of inequality of living standards from $\mathbf{x}$. Traditionally most measures of inequality have been uni-dimensional, and for practical purposes we will often want a single measure of inequality for each village, state or other region. Therefore it is desirable to first reduce the dimensionality of the data in order to calculate an informative index of asset wealth, and then construct an inequality measure based on this asset index. We would expect ownership of different assets to be highly correlated across households, so that a single summary measure should account for a reasonable amount of the cross-household variation in living standards.

One approach is to simply add up the number of assets that the household owns, weighting equally each asset. While reasonably common in the literature (see e.g., Montgomery et al. 2000), it is not clear that one should count owning a television, owning a bicycle, and having a flush toilet equally. Furthermore, this approach makes it more difficult to include measures of the quality of the materials used in the house when there are more than two options. In some circumstances it may be possible to aggregate assets by valuing them at market prices, preferably those in a well-functioning second-hand market. Conceptually this has several advantages, including providing a more easily interpretable measure of asset wealth, and allowing for easier comparisons across time and space.[6] However, well-designed collection of the prices of these assets is more time-consuming than simply asking about ownership, and many of the main demographic surveys do not include this information. Moreover, such a method is more difficult to extend to the inclusion of housing infrastructure, as it is not clear what the price of having a flush toilet or a wood rather than a dirt floor should be.[7]

An alternate approach is then to use the first component from principal components analysis as a proxy for levels of well-being. This approach has been used for over a decade to proxy the level of socioeconomic status in the child development and nutrition literature (e.g., Schroder et al. 1992; Pollitt et al. 1993) and has been recently evaluated and promoted by Filmer and Pritchett (2001) and Minujin and Bang (2002), who find that the resulting index provides a reasonable measure of wealth *level* effects in Indonesia, Pakistan, Nepal, India and Argentina. Filmer and Pritchett (2001) note that their choice of principal-components is a pragmatic one to a data constraint problem, with the underlying assumption being that household long-run wealth (or standard of living) explains the maximum variance in the asset variables. Formally, given asset vector $\mathbf{x}$, the first principal component of the observations, y, is the linear combination

$$y = a_1\left(\frac{x_1 - \bar{x}_1}{s_1}\right) + a_2\left(\frac{x_2 - \bar{x}_2}{s_2}\right) + \dots + a_p\left(\frac{x_p - \bar{x}_p}{s_p}\right) \tag{1}$$

whose sample variance is greatest among all such linear combinations, subject to the restriction $a'a = 1$, where $a = (a_1, a_2, \dots, a_p)'$ is the vector of coefficients, and $\bar{x}_k$ and s_k are the mean and standard deviation of variable x_k.[8] The principal component score for household i with vector of assets $\mathbf{x}_i$ is then $y_i = a'\widetilde{\mathbf{x}}_i$ where $\widetilde{\mathbf{x}}_i$ is the vector of standardized variables, $\frac{x_k - \bar{x}_k}{s_k}$. This transformed variable has zero mean and variance λ, where λ is the largest eigenvalue of the correlation matrix of the $\mathbf{x}$. If variable x_k is a dummy variable, then a_k/s_k gives the effect of a change from 0 to 1 in x_k on the principal component index y.

The first principal component therefore gives an index providing maximum discrimination between households, with the assets which vary most across households being given larger weight. An asset which all households own will thus be given zero weight in the first principal component, as it explains none of the variation across households (and likewise for an asset which no household owns). From the perspective of measuring relative inequality, there is then perhaps even more reason to favour principal components than for the case of measuring wealth levels, since principal components explicitly puts more weight on asset variables that are more unequally distributed across households.[9] Like Filmer and Pritchett (2001), we do not adjust the asset index for household size, since the benefits of infrastructure, the quality of housing materials, and many durable assets are available at the household level, and therefore focus on inequality across households.

For each household i we can then calculate the first principal component score, y_i, which we take to be a proxy for household wealth or standard of living. Note however that y_i can take negative values and has zero mean across all households. Many measures of inequality, such as the Gini coefficient and Atkinson, Theil and Generalized entropy indices, divide by the mean and hence will not be defined for the sample as a whole. The latter three measures also do not apply with data which takes negative values. Instead then, for community c, we propose using as our measure of inequality:

$$I_c = \frac{\sigma_c}{\sqrt{\lambda}} \tag{2}$$

where σ_c is the sample standard deviation of the y_i across households in community c, and λ is both the eigenvalue corresponding to the first principal component, and also the variance of y_i over the whole sample.[10] That is, our measure is a relative measure of inequality, defined as the standard deviation of the first principal component in a given community of interest relative to the standard deviation in the sample as a whole. I_c will be greater that one if community c displays more inequality within it than does the sample population as a whole. I_c satisfies many of the commonly accepted desired properties[11] of an inequality measure:

Lemma 1. *I_c satisfies the following version of the standard four axioms required of an inequality measure:*

1. *Anonymity – I_c is invariant to permutations of the asset index among households in the same community.*
2. *Scale independence – multiplying the asset index for each household in the sample by the same non-zero constant does not change any of the I_c's.*
3. *Population independence – replicating the whole sample of asset indices an integral number of times leaves I_c unchanged.*
4. *Pigou-Dalton transfer property – consider a positive transfer of δ from household i to household j in the same community c, where $y_i > y_j$ and $y_i - \delta > y_j + \delta$. Holding the asset indices for all other households constant, this transfer causes I_c to fall when there is more than one community in the whole sample (i.e., relative inequality in community c falls).*

Proof. See Appendix 1.

This measure of inequality can then be used for comparing inequality across communities or in the same community over time. It does not, however, provide an absolute measure of inequality for a single community. Since the question of interest is often to compare the determinants or effects of inequality across regions or over time, this does not seem to be that much of a limitation. To make comparisons over time, one should pool the data for all years in order to calculate the first principal component, since it is likely that the weights on the asset variables which most explain variation among households in the current cross-section may differ from the weights which explained the cross-sectional variation 10 years before. For example, if phones were owned by one-third of households a decade ago, but now most households own a phone (for reasons of rising incomes or falling technology prices), then principal components for the sample 10 years ago would put more weight on phone ownership than would principal components in the sample today. Using the pooled sample for both periods will however put weight on variables which explain variation over time as well as across households. Similarly, if one wishes to compare inequality in asset ownership across countries, then the household surveys from all countries should be pooled in order that the same weighting method can be used for each country, and that principal components can put weight on variables which explain variation across countries as well as households within countries.

2.1. Advantages and disadvantages of using asset indicators to measure inequality

We follow Filmer and Pritchett (2001) in viewing both the asset index and non-durable consumption as proxying unobserved long-run economic status or living standards, and hence inequality measures of both as proxying for inequalities in long-run living standards. Consumption is viewed by many as "the best measure of the economic component of living standards" (Deaton and Grosh 2000, p. 95), and is thus the preferred unit of analysis for study of poverty and inequality in developing countries. Nevertheless, there are a number of theoretical reasons why wealth inequality, which the asset-index measure may better capture, may be of more importance for development purposes than income or consumption inequality. In the models of Banerjee and Newman (1993) and Galor and Zeira (1993), it is the wealth distribution which determines investment in physical and human capital in economies with indivisible investments and credit market imperfections. Bardhan (1999) survey a variety of other mechanisms through which wealth inequality can affect economic performance. Empirically, cross-country studies which examine the relationship between initial inequality and subsequent growth have found a stronger effect of land and human capital inequality, than of income inequality, suggesting that it is asset inequality which matters more (see, e.g., Birdsall and Londoño 1997).

In addition to these possible theoretical reasons for wishing to use wealth inequality, there are a number of practical advantages of using asset indicators rather than income and consumption to measure inequality. The main advantage is a measurement one. There is likely to be much less recall bias or mismeasurement in questions such as whether the household owns a television, than in recalling consumption expenditure over the past week for each expenditure item. Measurement of income for self-employed and agricultural workers is notoriously difficult due to seasonality and accounting issues, while even consumption faces difficulty in the imputation of rental values and in accounting for the service flow from durables.[12] As Deaton (2003) notes, even if measurement error has little effect on the measurement of mean income, it will inflate the measured variance and measured income inequality. Information on the quality of housing and ownership of particular assets is much less seasonal and does not face these same measurement problems. The time taken to collect information on asset indicators is also likely to be much shorter than for detailed consumption or income collection, allowing surveys to potentially collect more information on other issues of interest.

The main challenge to using asset indicators to measure inequality in living standards is in ensuring that there are a sufficiently broad class of asset indicators collected as to allow for differentiation of living standards across all households. Minujin and Bang (2002) describe this as the fundamental condition for using this type of index to measure levels of living standards. Two possible problems that might arise in practice are clumping and truncation. If an insufficient number of asset indicators are used, then households will be clumped together in a small number of groups (in the extreme, with only one asset indicator, there is a group of owners and a group who don't own this asset), which limits the amount of useful information about inequality that can be inferred from the asset index. The second potential issue is that of truncation of the asset index distribution, which can arise if there are not indicators which allow one to tell between the poor and the very poor, or between the rich and the upper middle class. The amount of concern one should have about truncation depends on the reason why inequality is of interest: for example in theories in which inequality plays a role due to credit constraints and lumpy investment costs (e.g., Banerjee and Newman 1993), inequality among the rich is of less interest. Graphing a histogram and the probability density function of the asset index will enable one to see if clumping or truncation appear to be concerns in practice. The solution is to incorporate more indicators into the principal components analysis.

3. Using auxiliary surveys

While there are measurement and theoretical reasons why the inequality measure obtained directly from asset indicators can be expected to provide a good proxy of inequalities in living standards, it may also be desirable to consider inequalities in non-durable consumption. Suppose we are interested in examining the relationship between consumption inequality and specific variables of interest z, and the survey which contains information on z contains asset indicators $\mathbf{x}$, but no information on income and consumption. For example, the Demographic and Health Surveys contain detailed information on fertility, child nutrition status and utilization of basic health and education services which are not usually available in standard income and expenditure surveys, along with asset indicators. Denote the sample of households in this main survey by S_m. Often it is the case that a separate income and expenditure survey exists, such as the Living Standards Measurement Study (LSMS) surveys of the World Bank, which contains both information on non-durable consumption, *ndc*, and on $\mathbf{x}$, but no information on z. The sample of households in this auxiliary survey is denoted S_a. We then consider several methods for estimating the relationship between inequality in non-durable consumption and z, given these two samples S_m and S_a. The basic approach in all cases is to use the auxiliary survey to predict non-durable consumption from the information on assets, and then use this relationship to predict non-durable consumption for the households in the main survey. The methods considered differ in how this approach is implemented.

The first approach, which we will call the *direct prediction* method, is to use the auxiliary survey to regress non-durable consumption on the asset indicators and on demographic and other controls, w, which are available in both surveys. This method is analogous to that used by Skinner (1987) in attempting to estimate *levels* of total consumption in the PSID based on data from the CEX, although instead of asset indicators, he uses sub-components of expenditure, such as food, rent and utilities. Since the relationship between the level of non-durable consumption and asset ownership is likely to be non-linear[13], we use the natural log of non-durable consumption and carry out the following regression across households $i \in S_a$

$$\ln ndc_i = \beta' \mathbf{x}_i + \gamma' w_i + \varepsilon_i. \tag{3}$$

Using the fitted coefficients $\widehat{\beta}$ and $\widehat{\gamma}$, the predicted non-durable consumption for household $j \in S_m$ in the main survey is then:

$$\widehat{ndc}_j = \exp\left(\widehat{\beta}' \mathbf{x}_j + \widehat{\gamma}' w_j\right). \tag{4}$$

This approach is simple to employ, and has the advantage of weighting the various indicators by coefficients which directly translate asset ownership into consumption. However, since the residual terms ε_i are ignored, inequality measures based on the $\widehat{ndc}_j$ will tend to understate true inequality in current consumption.[14] Moreover, if there is heteroskedasticity in the ε_i (possibly due to non-linearities in the relationship between $\ln ndc$ and x), it may not even be the case that the ranking of communities by inequality according to $\widehat{ndc}_j$ is consistent for the ranking of communities by inequality according to true non-durable consumption. Non-linearities are likely to be the norm both due to an expected non-linear relationship between dollar wealth and non-durable consumption, and also due possibly to asset indicators being less able to distinguish between the wealth levels of the richest households than they do amongst poorer and middle-class households.

Blundell et al. (2003) face a related problem in trying to impute total non-durable consumption from food consumption. Their suggestion is to regress food on consumption, and then invert this relationship. We will refer to this as the *inversion* method. Applying this method to the situation at hand, we could use the first principal component of the asset indicators, y, and regress this on $\ln ndc$ and w for the households $i \in S_a$

$$y_i = \eta \ln ndc_i + \theta' w_i + v_i. \tag{5}$$

Assuming that $\widehat{\eta} \neq 0$, the imputed non-durable consumption for household $j \in S_m$ is then:

$$\widehat{ndc}_j = \exp\left(\frac{y_j - \widehat{\theta}' w_j}{\widehat{\eta}}\right) \tag{6}$$

As Blundell et al. (2003) show, such a method will overestimate the true variance of non-durable consumption, and hence will tend to overestimate levels of inequality. Moreover, if the v_i are heteroskedastic, again this method may not give consistent rankings of inequality.

To allow for heteroskedasticity in the error structure and to obtain measured levels of inequality closer to the truth, we consider a modification of the direct prediction method, which we will refer to as the *bootstrap prediction* method. The basic idea is to use the empirical distribution of the residuals obtained in fitting (3) to S_a and then draw from this distribution in obtaining the predicted non-durable consumption for households in S_m. To allow for heteroskedasticity, we suggest drawing from the distribution of residuals for households with similar levels of assets. The steps involved are:

1. Carry out the regression in (3) using either all the asset indicators $\mathbf{x}$, or the first principal component y as a regressor, and obtain the residuals from this regression, denoted $\widehat{\varepsilon}_i$.
2. Divide the sample in S_a into G groups according to quantiles of the distribution of the first principal component y in the auxiliary data. This then gives a separate set of residuals for each group G.
3. Divide the sample in S_m into the same G groups, using the same cut-off values for y as in the auxiliary sample.
4. For each household j in group g in S_m, draw an $\widetilde{\varepsilon}_j$ (with replacement) from the empirical distribution of residuals for households in group g in S_a, and use this to obtain the predicted value of non-durable consumption:

$$\widehat{ndc}_j = \exp\left(\widehat{\beta}' \mathbf{x}_j + \widehat{\gamma}' w_j + \widetilde{\varepsilon}_j\right) \tag{7}$$

5. Calculate the Gini coefficient (or other inequality measure of interest) for each community in S_m based on the predicted non-durable consumption from Step 4.
6. Repeat Steps 4 and 5 R times and take the mean Gini over all replications.

In choosing the number of groups G, the trade-off is that increasing G allows for more heteroskedasticity, but lowers the sample size available for each group, leading to noisier empirical distributions. As in pseudo-panel estimation, a sensible rule of thumb may be to try and have cell sizes of at least 100–200 observations. In particular, it is unlikely that one would want to define groups based on specific cells of the asset indicators themselves (e.g., a group of people owning a television, stove and iron, with a dirt floor, flush toilet and brick walls) since the number of cells grows rapidly with the number of indicators.[15] The bootstrap replication process allows for averaging out the bootstrap sampling error.

4. Data and levels

In order to examine how our asset indicator-based measures of inequality compare to standard consumption and income measures of inequality, we use Mexico's national income and expenditure survey, the *Encuesta Nacional de Ingresos y Gastos de los Hogares* (ENIGH) for the third quarter of 1998. The survey is conducted by and is available from the Instituto Nacional de Estadística, Geografía e Informática (INEGI). This survey contains extremely detailed information on income and expenditure for each household, as well as a set of questions on housing characteristics, household infrastructure and yes-no questions on household ownership of six types of vehicles and twenty-five types of assets.[16] Complete asset indicator data is available for 10,777 households. Non-durable consumption is computed by subtracting out expenditures on furniture and household appliances, leisure, entertainment and therapeutic equipment and vehicles. Income is net income from all sources excluding income from capital transactions such as the sale of a house or vehicle. Both income and non-durable consumption are measured for the past quarter.

The asset indicators can be divided into characteristics of the housing structure (number of rooms, type of floor, roof and walls, ownership status), household access to utilities and infrastructure (electricity, telephone, sewerage, etc.) and ownership of durable assets (vehicles, electronic equipment, whiteware, heating and air-conditioning). Table 1 provides the scoring factors of the first principal component when a separate index is made for each of these three categories, and then when all 30 indicators are used together. Each of the three specialized indices has a correlation of 0.84 or more with the overall index made up of all indicators. The first component accounts for 27% of the total variation across the 30 indicators. The scoring factors are positive for all infrastructure categories and for all assets except owning a bicycle, while poor quality housing materials such as a dirt floor, adobe walls or a plastic or cardboard roof receive negative coefficients. These coefficients therefore suggest the first principal component is indeed providing a measure of wealth. The scoring coefficients have been rescaled by the standard deviations of their respective variables, so that the represent the effect of a change from 0 to 1 in the dummy variables of interest. Having piped water, electricity and owning a computer raise the overall asset index by the largest amount, while having a dirt floor or roof made of poor quality materials has the largest negative effect.

The fifth and sixth columns of Table 1 provide the overall mean and standard deviation for each one of the 30 indicators. We see there is a wide range of average levels of ownership, from goods which very few households have, such as a computer, dryer and air-conditioning, to goods which most households have, such as a stove and television. The last three columns of Table 1 presents means for each of the indicators for households classified into terciles of the overall asset index. As in Filmer and Pritchett (2001), we find the overall index to be internally coherent in the sense that mean asset ownership differs markedly across different groups. Only 0.2% of the poorest tercile of households own a phone, while 69.7% of the richest tercile do; 5.8% of poor households have piped water in their home compared to 98.8% of rich households; 34.5% of the poorest tercile of households have a dirt floor, while only 0.2% of the richest tercile do; all durable goods except for bicycles show increasing ownership with tercile.

Table 1. Principal components and summary statistics for asset indicators

Variables	Scoring factor for 1st principal component (divided by Std. Dev.)				Summary statistics		Means by tercile of all indicator asset index		
	Housing only	Utilities only	Durables only	All indicators	Mean	Std. dev.	Lowest	Middle	Upper
Housing characteristics									
Number of rooms	0.219			0.154	2.809	1.509	1.773	2.612	3.987
Home owner	−0.122			−0.030	0.768	0.422	0.814	0.718	0.774
House has brick walls	0.953			0.445	0.713	0.452	0.371	0.828	0.940
House has adobe walls	−0.830			−0.313	0.130	0.336	0.254	0.113	0.024
House has a brick roof	0.924			0.438	0.512	0.500	0.122	0.561	0.853
House has plastic, cardboard or absestos roof	−0.934			−0.446	0.363	0.481	0.738	0.301	0.050
House has a dirt floor	−0.955			−0.567	0.120	0.325	0.345	0.014	0.002
House has a wood floor	0.737			0.485	0.295	0.456	0.025	0.157	0.703
Utilities and access to infrastructure									
Piped water in house		2.077		1.094	0.549	0.248	0.058	0.601	0.988
Toilet facilities in the house		1.105		0.539	0.683	0.465	0.212	0.845	0.992
Household has rubbish collection		1.013		0.487	0.617	0.486	0.184	0.710	0.954
Household has electric lighting		1.255		0.645	0.957	0.204	0.871	0.999	1.000
Household has a telephone		0.912		0.530	0.267	0.442	0.002	0.101	0.697
Durable asset ownership									
Owns a car			0.723	0.462	0.205	0.404	0.020	0.108	0.491
Owns a van			0.368	0.184	0.126	0.332	0.062	0.135	0.186

Table 1. (continued)

Variables	Scoring factor for 1st principal component (divided by Std. Dev.)				Summary statistics		Means by tercile of all indicator asset index		
	Housing only	Utilities only	Durables only	All indicators	Mean	Std. dev.	Lowest	Middle	Upper
Owns a van			0.368	0.184	0.126	0.332	0.062	0.135	0.186
Owns a motocycle			0.262	0.127	0.014	0.119	0.008	0.016	0.019
Owns a bicycle			−0.072	−0.075	0.234	0.423	0.276	0.254	0.170
Owns a radio			0.212	0.137	0.328	0.470	0.237	0.338	0.417
Owns a television			0.734	0.544	0.848	0.359	0.594	0.965	0.996
Owns a video			0.694	0.462	0.295	0.456	0.028	0.203	0.660
Owns a computer			1.017	0.608	0.047	0.213	0.000	0.003	0.142
Owns a fan			0.451	0.280	0.490	0.500	0.248	0.546	0.683
Owns a sewing machine			0.394	0.238	0.291	0.454	0.146	0.280	0.455
Owns a stove			0.745	0.568	0.848	0.359	0.579	0.982	0.996
Owns a fridge			0.723	0.518	0.663	0.473	0.224	0.790	0.988
Owns a washing machine			0.689	0.466	0.452	0.498	0.073	0.443	0.851
Owns a dryer			0.992	0.592	0.063	0.242	0.000	0.006	0.185
Owns a microwave			0.884	0.539	0.130	0.336	0.001	0.032	0.361
House has airconditioning			0.651	0.371	0.070	0.256	0.008	0.046	0.160
Owns a heater			0.691	0.517	0.353	0.478	0.013	0.235	0.823
Housing indicator asset index					0.000	1.769	−1.784	0.194	1.591
Utilities asset index					0.000	0.892	−1.004	0.161	0.844
Durables asset index					0.000	2.068	−1.997	−0.124	2.192
All indicator asset index:					0.000	2.839	−3.259	0.105	3.155
Three-month household non-durable consumption					9056	7657	4534	7750	15048
Three-month household food expenditure					3335	2224	2187	3265	4620

Three-month total household income					10335	10151	4645	8569	18042
Eigenvalue associated with first component	3.13	2.67	4.28	8.06					
Share of variance associated with first component	0.39	0.53	0.25	0.27					
Number of variables used	8	5	17	30					

Notes: All variables are dummies apart from number of rooms. Non-durable consumption, food expenditure and income means and standard deviations are reported after trimming top and bottom 1% of observations.

Scoring factor is the weight assigned to each variable (normalized by subtracting its mean) in the linear combination of variables comprising the 1st principal component

Omitted type of walls includes walls made of wood, cardboard or cement, Omitted type of roofing materials include concrete and tile.

Omitted floor type is cement. Data are unweighted.

Source: Author's calculations from ENIGH 1998.

Table 2. Mean levels of indicators, non-durable consumption and income by Mexican state

State	State means							State means in fitted non-durable consumption according to							
	Non-durable consumption	Income	Food expenditure	Housing index	Utilities index	Durables index	Overall index	Prediction method	Inversion method	Bootstrap on overall index			Bootstrap on all 30 assets		
										10 groups	25 groups	100 groups	10 groups	25 groups	100 groups
Baja California Norte	15243	18313	4553	−0.48	0.53	1.91	1.30	11509	19387	11218	11270	11206	12846	12880	12798
Federal District	14486	18591	4605	1.25	1.33	1.33	2.26	11118	20410	12807	12768	12832	12471	12423	12476
Nuevo Leon	13016	16644	3162	0.92	0.95	1.29	1.85	10591	22290	11888	11776	11877	11886	11766	11842
Baja California Sur	11997	19728	3553	−0.20	0.04	0.41	0.19	8729	12820	8880	9043	8920	9803	10028	9863
State of Mexico	11494	12599	3797	0.73	0.67	0.36	0.97	9596	16268	10922	10884	10877	10789	10738	10743
Jalisco	11067	13771	4000	0.90	0.71	0.56	1.26	9437	20346	10693	10690	10684	10634	10597	10591
Coahuila	10794	12427	3476	−0.03	0.13	0.88	0.65	9072	17464	10097	10066	10053	10262	10166	10186
Durango	9872	10974	3165	−0.84	−0.30	0.23	−0.36	7470	10425	8264	8297	8345	8536	8528	8574
Tamaulipas	9864	14034	2699	−0.34	−0.32	0.19	−0.27	8152	12625	8694	8968	8774	9262	9503	9296
Guanajuato	9709	10782	3686	0.40	0.07	−0.01	0.26	8475	16087	9689	9754	9707	9588	9612	9580
Colima	9517	9808	3530	−0.20	−0.04	−0.43	−0.48	7355	10317	8242	8201	8274	8368	8315	8382
Sinaloa	9358	11237	3324	0.51	−0.29	0.09	0.10	8050	14548	9337	9252	9157	9126	9000	8948
Morelos	9054	9722	3382	−0.22	−0.02	−0.16	−0.22	7536	10127	8590	8638	8648	8494	8544	8543
Aguascalientes	8964	11052	3056	0.31	0.88	0.55	1.05	9393	18762	10748	10822	10663	10583	10712	10540
Sonora	8631	12960	2848	−0.39	0.24	0.61	0.29	8468	13604	9438	9297	9214	9638	9485	9387
Quintana Roo	8605	10019	3261	−0.36	−0.39	−0.73	−0.95	6833	9003	7898	7893	7867	7832	7789	7773
Michoacan	8583	9924	3351	0.21	0.03	−0.32	−0.06	7948	13406	8998	9143	9085	8939	9080	8984
Veracruz	8477	10020	3144	−0.03	−0.10	−0.29	−0.28	7482	10553	8395	8268	8538	8497	8355	8585
Queretaro	8310	8324	2909	−0.11	−0.81	−0.80	−1.04	6866	9270	7840	8019	7756	7782	7978	7712
Nayarit	8200	8914	3529	0.00	−0.32	−0.07	−0.24	6933	10639	7802	7925	7983	7773	7918	7967
Zacatecas	8060	8475	2697	−0.72	−0.66	−0.04	−0.68	7343	10234	8138	8022	8268	8416	8274	8487
Chihuahua	7980	12046	2884	−0.90	0.54	1.19	0.66	8766	14697	9473	9407	9386	9829	9720	9723

Guerrero	7954	8046	2973	−0.85	−1.03	−0.86	−1.57	6044	6350	6933	6896	6920	6971	6942	6946
Hidalgo	7895	7862	2578	−0.18	−0.69	−1.01	−1.15	6683	8059	7932	7786	7785	7695	7585	7605
Puebla	7168	7361	2760	−0.19	−0.78	−1.20	−1.34	6309	8551	7436	7408	7401	7193	7228	7155
Tlaxcala	7003	7349	2665	0.06	−0.08	−0.72	−0.48	6584	8491	8018	8017	7930	7526	7524	7442
San Luis Potosi	6603	7586	2534	−0.48	−0.91	−0.61	−1.14	6803	9836	7764	7780	7785	7669	7673	7646
Campeche	6523	7833	2770	−1.31	−1.23	−1.43	−2.31	5114	4061	5891	5858	5796	6033	6007	5957
Chiapas	6400	6879	2360	−1.28	−0.99	−1.63	−2.33	5857	6011	6798	6777	6673	6818	6779	6680
Tabasco	6211	7434	2740	−0.65	−0.50	−1.01	−1.31	6491	6563	7509	7492	7580	7466	7454	7517
Oaxaca	6045	5945	2500	−1.31	−1.23	−1.71	−2.50	4708	4560	5528	5651	5491	5421	5565	5393
Yucatan	5891	6486	2818	−1.03	−1.25	−1.27	−2.07	5531	6137	6623	6390	6319	6548	6301	6261
Correlation with NDC	1.000	0.922	0.849	0.633	0.782	0.848	0.850	0.911	0.833	0.860	0.869	0.867	0.911	0.916	0.916
Rank-order Correlation	1.000	0.919	0.809	0.598	0.798	0.843	0.847	0.894	0.853	0.865	0.875	0.869	0.896	0.902	0.894

Notes: NDC = Non-durable Consumption. $R = 20$ replications used for Bootstrap Method. Spearman rank-correlation test p-values are less than 0.01 for all columns.

Top and bottom 1% of national observations trimmed when calculating non-durable consumption, income and food means.

Data are unweighted.

Source: Author's calculations from ENIGH 1998.

In order to show how well the various indicators do at proxying for levels of well-being, Table 2 presents means for each of Mexico's 31 states and the Federal District for each of the asset indices considered in Table 1, and compares these to state mean non-durable consumption (NDC), income and food expenditure. The sample size for each state averages 337 households, with at least 223 households in each state. The ENIGH surveys are designed to be representative at the National level, and at the urban and rural sublevels, so we do not use the sampling weights since they are not designed for state-level comparisons. Overall one sees that states with higher consumption and income (such as the Federal District and Nuevo Leon) also have higher values of the asset indices, while the poorest states of Chiapas, Oaxaca and Yucatan have negative mean asset indices. The bottom of the Table presents the correlation and rank-order correlation of each measure with non-durable consumption. The housing index has a rank-order correlation of 0.598 with NDC, while the utilities index, durables index and overall index all do better. The overall index has slightly higher correlation with NDC than does food expenditure, suggesting that it does provide an appropriate proxy for well-being. The remaining columns of Table 2 present predicted non-durable consumption using the methods developed for use with auxiliary surveys. The inversion method seems to perform worst, although the correlation in levels is still 0.833 with actual non-durable consumption. The prediction method and the bootstrap prediction method using all 30 indicators as regressors in equation (3) both give correlations of at least 0.9 with actual non-durable consumption. Overall Table 2 then confirms that the result of Filmer and Pritchett (2001) and Minujin and Bang (2002) that a principal-component based method does a good job in proxying for *levels* of well-being is seen to extend to the Mexican data.

5. Results for inequality

The methodological discussion in Sect. 2.1 pointed out that even if the asset index succeeds in proxying levels of well-being, it may do a poor job in proxying for inequality if the index suffers from clumping or truncation. Figure 1 plots histograms and kernel-density estimates of the distribution of each of the four asset indices in order to visually determine whether these are likely to be concerns.[17] Both the utilities and housing indices show evidence of clumping, suggesting that the 5 or 8 indicators which make up these indices are not able to fully distinguish households. The utilities index also shows evidence of truncation at the top, with a lot of households having all or most of the infrastructure. The histogram is useful for identifying the households which are located at each clump.[18] Three large spikes are seen in the histogram for utilities. The first, at −1.251, consists of households with electricity, but none of the other utilities. The second, at 0.657, consists of households with piped water, toilet facilities, rubbish collection and electricity, but without telephones, while the third major spike, at 0.979, consists of households with all five utilities. As 23.7% of households lie in this last category, the utilities index will clearly not be helpful for determining inequality amongst the rich. The housing characteristics index shows more categories, with the largest spike at 1.087 covering only 6% of households,

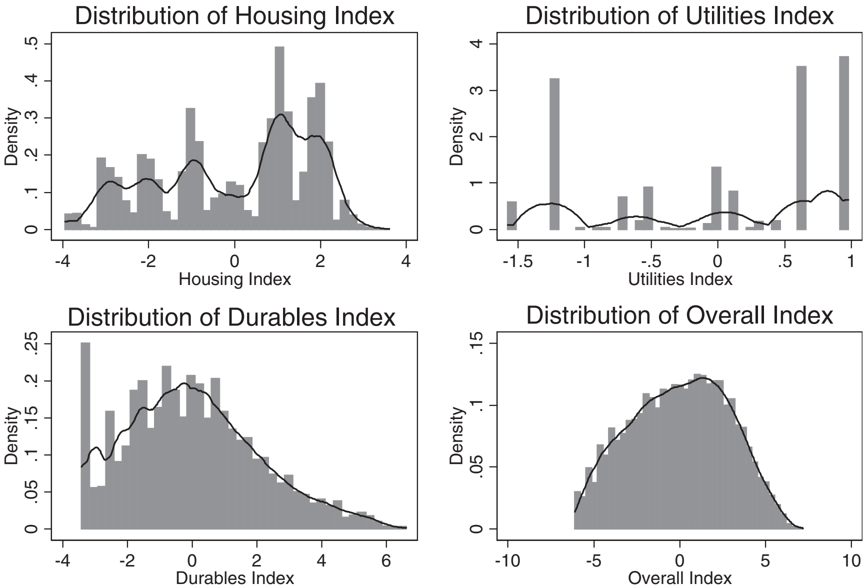

Fig. 1. Histogram and Kernel densities of the national distribution of different indices

comprising of those with three rooms, who own their homes, have brick walls and a brick roof, but a cement floor.

In contrast, the durables index is much smoother, with little evidence of clumping, but some truncation at the bottom. The spike at the left tail consists of about 5% of households which have no durable assets, grouped together with a few households which just have a bicycle. This truncation may make the durables index less useful for determining inequality amongst the poor. The overall index, which uses 30 indicators, does not show any clumping and truncation seems much less of an issue for this index. No single value of the overall index accounts for more than 0.6% of all households.

Table 3 then presents the calculated gini coefficients for non-durable consumption, income, and food expenditure for each state in the sample, along with the relative inequality measure I_c for each of the four asset indices. The three states with the highest non-durable consumption and income ginis, Chiapas, Hidalgo and Puebla, all have I_c above one for the overall asset index, indicating that they have more asset inequality within them than within Mexico as a whole. The Federal District has the second-lowest NDC gini, and lowest overall asset index I_c, identifying this as a more equal state. The foot of Table 3 shows the correlation and rank-order correlation of each measure with state-level NDC ginis. Income and consumption inequality are seen to be highly correlated across states, while the asset indicator based measures have lower correlations. The durable asset index performs particularly poorly in this regard, have essentially zero correlation across states between relative inequality in durables and NDC inequality. This is likely a result of the truncation seen in Fig. 1, which prevents this measure from adequately capturing inequality amongst the poor. To examine this possibility, we trim the top and bottom 5% of observations from the distribution of each asset index, and then compare the I_c measure for this truncated distribution to the NDC gini after

Table 3. Inequality measures by Mexican state

State	State gini coefficients			State relative inequality (I_c)			
	Nondurable consumption	Income	Food expenditure	Housing index	Utilities index	Durables index	Overall index
Baja California Norte	0.329	0.375	0.277	0.848	0.857	0.973	0.867
Federal District	0.359	0.422	0.282	0.609	0.508	0.927	0.703
State of Mexico	0.364	0.403	0.313	0.776	0.818	0.944	0.858
Baja California Sur	0.376	0.385	0.385	0.884	0.988	0.912	0.897
Tlaxcala	0.376	0.417	0.308	0.931	0.893	0.748	0.791
Durango	0.380	0.429	0.370	1.115	0.976	0.886	0.930
Jalisco	0.384	0.422	0.324	0.776	0.772	0.882	0.798
Coahuila	0.391	0.427	0.365	1.027	0.985	0.953	0.971
Guanajuato	0.393	0.434	0.332	0.911	1.007	0.924	0.949
Sonora	0.394	0.437	0.409	1.049	0.909	0.904	0.901
Aguascalientes	0.398	0.443	0.353	0.767	0.639	0.914	0.766
Nayarit	0.401	0.457	0.332	0.922	0.903	0.749	0.831
Nuevo Leon	0.405	0.420	0.382	0.809	0.673	0.900	0.763
Colima	0.405	0.481	0.361	0.971	0.866	0.861	0.911
Tabasco	0.405	0.464	0.328	0.866	0.932	0.879	0.909
Morelos	0.406	0.433	0.356	1.118	0.895	0.851	0.929
Chihuahua	0.407	0.432	0.376	1.048	0.861	0.922	0.869
Guerrero	0.411	0.483	0.377	1.005	0.860	0.803	0.836
Queretaro	0.412	0.467	0.365	0.890	0.941	0.909	0.898
Zacatecas	0.412	0.472	0.406	1.121	0.922	0.835	0.910
Sinaloa	0.415	0.455	0.364	0.985	1.089	0.957	1.002
Michoacan	0.416	0.463	0.343	0.950	0.989	0.904	0.940
Veracruz	0.416	0.475	0.335	1.049	0.976	0.938	1.010
Campeche	0.426	0.490	0.352	0.729	0.848	0.725	0.793
San Luis Potosi	0.431	0.530	0.373	1.068	1.057	0.974	1.040
Quintana Roo	0.432	0.478	0.343	1.011	0.963	0.951	1.014
Yucatan	0.435	0.483	0.320	0.743	0.912	0.845	0.883
Tamaulipas	0.441	0.495	0.402	0.990	1.120	1.029	1.027
Oaxaca	0.465	0.484	0.429	1.004	0.865	0.775	0.873
Puebla	0.474	0.507	0.404	1.011	1.004	0.930	1.014
Hidalgo	0.483	0.527	0.430	0.952	1.004	0.962	1.001
Chiapas	0.499	0.551	0.453	1.032	1.034	0.899	1.014
Correlation with NDC gini	1.000	0.898	0.747	0.343	0.448	−0.013	0.539
Rank-order correlation	1.000	0.905	0.536	0.325	0.496	0.042	0.566
Spearman p-value	0.000	0.000	0.002	0.069	0.004	0.821	0.001
Correlation after trimming	1.000	0.853	0.800	0.329	0.188	0.406	0.517
Rank-order correlation	1.000	0.878	0.714	0.214	0.267	0.443	0.540
Spearman p-value	0.000	0.000	0.000	0.239	0.140	0.011	0.001
Correlation with food gini	0.747	0.585	1.000	0.566	0.418	0.049	0.454
Rank-order correlation	0.536	0.481	1.000	0.563	0.396	0.045	0.400
Spearman p-value	0.002	0.005	0.000	0.001	0.025	0.808	0.023

Notes: NDC = Non-durable consumption.
Spearman rank-correlation tests null of independence. Top and bottom 1% of national observations trimmed when calculating non-durable consumption, income and food means.
Correlation after trimming drops top and bottom 5% of observations when calculating Ic measures and when calculating ginis.
Data are unweighted.
Source: Author's calculations from ENIGH 1998.

also trimming the top and bottom 5% of the NDC distribution. The correlation of the durables index is now much stronger, showing it does a better job at measuring inequality for the truncated distribution. Conversely, the utilities index now performs worse, which is perhaps not surprising given that Fig. 1 shows that much of the distribution is in the tails.

A Spearman rank-order correlation tests rejects the null of independence between the raw measures of relative inequality and the NDC gini for all but the durables index, while the truncated durables index also shows a strong relationship with truncated consumption inequality. The relationship is strongest for the overall index. The I_c for the overall index has a rank-order correlation of 0.566 with the NDC gini, which is higher than the 0.536 rank-order correlation of the food gini with the NDC gini. The overall index also does almost as well as income in terms of its correlation with food inequality.

Thus while the relationship between relative inequality measured by the asset indicators and inequality in NDC is not as strong as the relationship in levels, Table 3 does show significant correlations. Relative inequality in the overall asset index does appear to be a reasonable proxy for both inequality in non-durable consumption and inequality in food expenditure. As discussed in Sect. 2.1, there are a variety of measurement issues that can affect the measurement of income and NDC inequality, so the I_c index may be preferred in some circumstances as a measure of inequality in overall levels of wellbeing. Furthermore, conceptually the asset index is closer to measuring wealth inequality, which we would expect to be positively, but not perfectly, correlated with consumption and income inequality. Hence which measure is to be preferred will depend on the particular concept of inequality which is of interest.

If inequality in non-durable consumption is the prime interest, then the auxiliary survey approach can be used to predict inequality in non-durable consumption as outlined in Sect. 3. Figure 2 plots the predicted Gini's by state for the Prediction method, Inversion method, and for the Bootstrap prediction method with 20 groups, using the overall asset index and using each of the 30 indicators as regressors in equation (3). As expected, the prediction

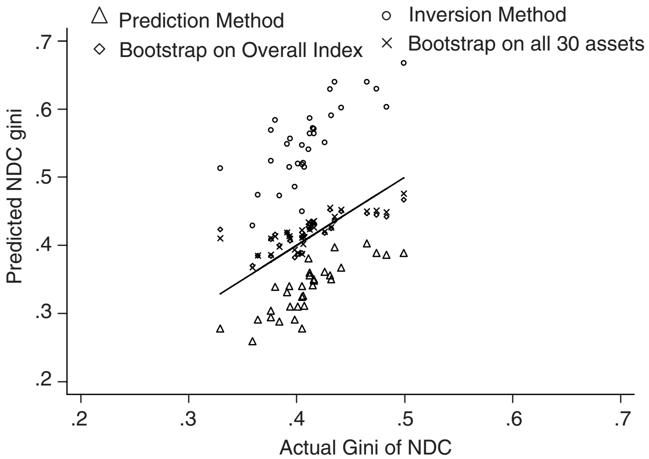

Fig. 2. Actual and predicted NDC ginis by state

method tends to underestimate the true gini, with all observations lying below the 45 degree line, while the inversion method overestimates the true gini. Both bootstrap methods do much better in terms of fitting the level of the gini, with similar results obtained from using the overall index or each of the 30 indicators as regressors.

Table 4 presents the full set of predicted NDC gini's by state. The correlations between predicted NDC inequality and actual NDC inequality are larger than those in Table 3 between I_c and NDC inequality. This is not surprising given that now we are using the asset indicators to predict NDC, rather than using them directly as proxies for standards of living. The inversion method seems to do most poorly, with the lowest correlation between predicted and actual inequality, and implausibly high predicted gini coefficients. The prediction method undermeasures the levels of inequality, but does a better job in the rankings of states by inequality. The bootstrap prediction method works slightly better when all asset indicators are used as regressors in Eq. (3) than when just the overall asset index from principal components is used. The rank-order correlations between predicted inequality using the bootstrap method on all 30 indicators and actual inequality are around 0.85. In terms of the number of groups used in the bootstrap, one finds reasonably similar results from 10, 25 and 100 groups, although slightly poorer results from using only 2 groups.[19] Overall, the bootstrap prediction method appears to work well at predicting both the levels and rankings of NDC inequality, and is to be preferred to the prediction and inversion methods.

With an average of 337 households per state in the sample, the sampling error in calculating the actual non-durable consumption gini for each state is non-negligible. Figure 3 plots pointwise 95 percentage confidence intervals for the actual NDC gini, which are calculated as bias-corrected percentiles via the bootstrap.[20] The resulting confidence intervals are approximately the actual value$\pm$0.03, with smaller intervals for states with larger samples, such as the Federal District. Figure 3 also plots the predicted NDC gini obtained from using the bootstrap prediction method with all 30 indicators and 25 groups. The predicted NDC ginis almost all lie within the 95% confidence interval, with the main exception being the state of Baja California Norte, which has a significantly higher predicted NDC gini according to asset levels in the state than the actual NDC gini actually is. Figure 3 therefore further shows that the bootstrap prediction method succeeds in obtaining accurate measures of NDC inequality.

6. Robustness

We have found that the relative inequality measure, I_c, has a significant correlation with the gini of non-durable consumption across states. The gini coefficient is more sensitive than some other inequality measures to inequality (or measurement error) at the top of the NDC distribution (see Deaton 2003). Table 5 examines how robust the relationship between I_c and inequality in NDC is to alternative measures of inequality. We consider the Atkinson measures for $\varepsilon = 0.5, 1$ and 2; the variance of logarithms; the coefficient of variation; the 90-10 ratio; and Generalized entropy measures for $\alpha = -1, 0, 1$ and 2.[21] All measures show significant correlations with I_c, with slightly higher correlations found for measures which are more sensitive to inequality

Table 4. Inequality measures for auxiliary methods by Mexican state

State	Actual State ginis for fitted non-durable consumption according to								
	NDC gini	Prediction method	Inversion method	Bootstrap on overall index			Bootstrap on all 30 assets		
				10 groups	25 groups	100 groups	10 groups	25 groups	100 groups
Baja California Norte	0.329	0.278	0.513	0.415	0.423	0.418	0.405	0.410	0.404
Federal District	0.359	0.259	0.429	0.370	0.370	0.372	0.369	0.367	0.369
State of Mexico	0.364	0.291	0.474	0.386	0.385	0.383	0.388	0.385	0.383
Baja California Sur	0.376	0.294	0.569	0.407	0.409	0.403	0.405	0.410	0.404
Tlaxcala	0.376	0.304	0.524	0.391	0.384	0.380	0.394	0.386	0.381
Durango	0.380	0.339	0.584	0.410	0.416	0.409	0.408	0.413	0.406
Jalisco	0.384	0.288	0.473	0.400	0.400	0.395	0.397	0.398	0.392
Coahuila	0.391	0.331	0.549	0.417	0.419	0.423	0.420	0.419	0.423
Guanajuato	0.393	0.340	0.515	0.410	0.413	0.409	0.413	0.413	0.411
Sonora	0.394	0.310	0.557	0.413	0.407	0.404	0.416	0.409	0.404
Aguascalientes	0.398	0.291	0.486	0.380	0.382	0.384	0.390	0.393	0.393
Nayarit	0.401	0.310	0.520	0.385	0.387	0.390	0.380	0.389	0.392
Nuevo Leon	0.405	0.278	0.450	0.384	0.389	0.387	0.382	0.387	0.381
Colima	0.405	0.340	0.519	0.412	0.412	0.413	0.413	0.411	0.410
Tabasco	0.405	0.324	0.547	0.419	0.419	0.426	0.426	0.422	0.428
Morelos	0.406	0.326	0.521	0.415	0.413	0.409	0.401	0.402	0.399
Chihuahua	0.407	0.311	0.515	0.413	0.415	0.412	0.411	0.412	0.409
Guerrero	0.411	0.381	0.541	0.428	0.427	0.418	0.435	0.434	0.426
Queretaro	0.412	0.356	0.564	0.412	0.424	0.417	0.416	0.427	0.417
Zacatecas	0.412	0.360	0.587	0.439	0.431	0.434	0.435	0.423	0.424
Sinaloa	0.415	0.341	0.572	0.435	0.433	0.431	0.440	0.434	0.435
Michoacan	0.416	0.350	0.564	0.418	0.429	0.427	0.427	0.435	0.430
Veracruz	0.416	0.348	0.571	0.433	0.433	0.433	0.429	0.426	0.427
Campeche	0.426	0.361	0.551	0.418	0.418	0.410	0.422	0.420	0.413
San Luis Potosi	0.431	0.356	0.629	0.446	0.452	0.445	0.449	0.455	0.446
Quintana Roo	0.432	0.350	0.591	0.428	0.425	0.425	0.436	0.426	0.425
Yucatan	0.435	0.397	0.640	0.450	0.436	0.438	0.457	0.442	0.445
Tamaulipas	0.441	0.367	0.602	0.436	0.450	0.441	0.436	0.452	0.441
Oaxaca	0.465	0.403	0.640	0.442	0.447	0.441	0.441	0.450	0.441
Puebla	0.474	0.389	0.630	0.443	0.445	0.447	0.446	0.451	0.449
Hidalgo	0.483	0.386	0.603	0.449	0.442	0.446	0.456	0.449	0.454
Chiapas	0.499	0.389	0.668	0.462	0.467	0.466	0.478	0.476	0.475
Correlation with NDC gini	1.000	0.830	0.749	0.753	0.738	0.767	0.796	0.809	0.825
Rank-order Correlation	1.000	0.880	0.775	0.835	0.830	0.831	0.853	0.860	0.853
Spearman P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Correlation with food gini	0.747	0.591	0.615	0.586	0.602	0.591	0.594	0.643	0.615
Rank-order correlation	0.536	0.510	0.556	0.461	0.488	0.489	0.498	0.518	0.480
Spearman p-value	0.002	0.003	0.001	0.008	0.005	0.005	0.004	0.002	0.005

Notes: NDC = Non-durable Consumption. R = 20 replications used for bootstrap method. Spearman rank-correlation tests null of independence. Top and bottom 1% of national observations trimmed when calculating and predicting non-durable consumption.
Data are unweighted.
Source: Author's calculations from ENIGH 1998.

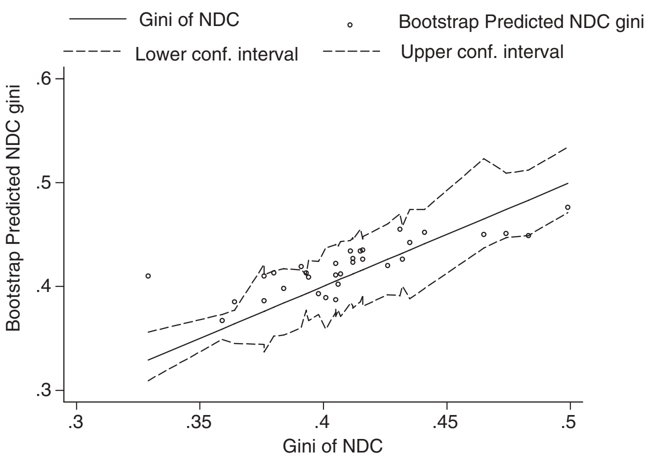

Fig. 3. Do the Bootstrap prediction method NDC gini coefficients lie within 95% confidence intervals of the true NDC gini?

Table 5. Robustness to alternative inequality measures

Measure of inequality in non-durable consumption	Relationship with overall state relative inequality measure (I_c)		
	Correlation	Rank-order correlation	Spearman p-value
Gini	0.539	0.566	0.001
Atkinson measures			
A(0.5)	0.547	0.568	0.001
A(1)	0.565	0.580	0.001
A(2)	0.571	0.561	0.001
Variance of log NDC	0.570	0.600	0.000
Coefficient of variation	0.477	0.488	0.005
90–10 ratio	0.500	0.612	0.000
Generalized entropy			
GE(−1)	0.549	0.566	0.001
GE(0)	0.559	0.580	0.001
GE(1) (Theil's first measure)	0.524	0.541	0.001
GE(2)	0.474	0.488	0.005

Notes: see Fields (2001, p. 30) for formulae.
Calculations are after trimming top and bottom 1% of non-durable consumption.

among the poor, such as the Atkinson A(2) measure and the variance of logarithms. This is to be expected if the asset indicators do a better job of distinguishing among the poor than among the rich. There appears to be a robust relationship between I_c and non-durable consumption inequality regardless of the inequality measure used.

As an additional robustness check, we investigated whether the standard deviations of higher-order principal components had any relation to NDC inequality. Using all 30 indicators, the second principal component is defined as the linear combination of the variables that accounts for a maximal proportion of the remaining variance subject to being orthogonal to the first principal component. Higher order components are defined similarly. While, by construction, the index formed from the first component is orthogonal to an index of the second component, it may still be the case that inequality measures formed from different components could be correlated. However, a Spearman test could not reject independence of the state-level standard deviations of the first principal component with any of the state-level standard deviations of principal components two through five. Moreover, regressing the gini of NDC on the standard deviations of each of the first five principal components, one finds a positive and significant coefficient for the standard deviation of the first principal component, and can not reject that the coefficients on the standard deviations of the next four principal components are jointly zero. Therefore it appears that only the first principal component is necessary for measuring inequality.

6.1. Inclusion of other major assets

The housing quality, household infrastructure, and durables asset indicators considered here are the assets most commonly used in the literature when constructing proxies for wealth levels (see for example the survey in Montgomery et al. 2000). The main reason that the literature has tended to focus on these measures are that they are easy to collect and are the common measures available in many demographic surveys, as discussed in the introduction. Nevertheless, it is clear that there are several important classes of assets which are not considered. The first is financial assets, both in the form of cash, and also in the form of pension entitlements among upper-income households. Such information is rarely collected in surveys in developing countries, except through special surveys which then do not contain information on the many other variables of interest. The ENIGH survey from Mexico used in this study is no exception, and financial asset information is not available. Nevertheless, the presumption is that ownership of these assets will be strongly correlated with the other asset indicators, and in particular, that any substitution between financial assets and other forms of assets does not vary systematically across communities. If this is true, then the relative inequality measure I_c will still rank inequality across communities correctly. Alternatively, if auxiliary methods are available, the bootstrap method will still provide accurate measures of inequality in consumption, since the bootstrapped residuals will pick up left over variation due to financial assets.

Researchers may also want to consider the inclusion of land and human capital assets, depending on the type of inequality they wish to proxy for. In other cases (e.g., Birdsall and Londoño 1997), the interest may be on the separate effects of land inequality, human capital inequality, and income or asset inequality, and so one would not wish to include these other assets. The ENIGH survey used here does not collect information on land ownership, whereas most of the Demographic and Health surveys do. The ENIGH data is used here because it contains both asset indicator and consumption information, allowing for comparisons between measures. Nevertheless, to examine how robust our results are likely to be to the inclusion of land and human capital indicators, we employ data from the Mexican Migration Project (MMP), a collaborative research project based at the University of

Pennsylvania and the University of Guadalajara[22], which contains asset indicators, indicators of land ownership, and information on education of the household head, but does not contain consumption data. We use the MMP71 database, which contains data on approximately 200 households in each of 71 communities.[23]

Column 1 of Table 6 provides the scoring factors for the first principal component based on 17 asset indicators in the MMP which correspond closely to those used above.[24] Column 2 then adds three indicators of land ownership, which are dummy variables for ownership of private city land, private rural land, and ejido or community land. The first two land types have

Table 6. Robustness to inclusion of land ownership. Results from the Mexican migration project survey

Variables	Scoring factor for first Principal component (/Std. Dev)			Summary Statistics	
	Asset index (1)	Adding land (2)	Adding education (3)	Mean of variable	Std. dev of variable
Number of rooms/member	0.160	0.160	0.157	1.082	0.962
Brick and cement or tile roof	0.390	0.393	0.387	0.758	0.428
Dirt floor	−0.889	−0.889	−0.875	0.113	0.317
Wood or tile floor	0.537	0.538	0.536	0.434	0.496
Running water	0.840	0.834	0.815	0.941	0.236
Sewerage	0.656	0.657	0.649	0.768	0.422
Electricity	1.188	1.181	1.154	0.971	0.168
Telephone	0.595	0.595	0.597	0.251	0.434
Car	0.494	0.496	0.508	0.187	0.390
Van	0.381	0.373	0.365	0.183	0.387
Radio	0.554	0.549	0.541	0.906	0.292
Television	0.855	0.851	0.836	0.893	0.309
Sewing machine	0.450	0.446	0.434	0.475	0.499
Stove	1.027	1.025	1.004	0.922	0.269
Fridge	0.730	0.726	0.719	0.670	0.470
Washing machine	0.620	0.616	0.610	0.506	0.500
Stereo	0.571	0.568	0.566	0.454	0.498
Private city land		0.343	0.338	0.014	0.119
Private rural land		0.022	0.010	0.059	0.236
Ejido or community land		−0.203	−0.209	0.122	0.328
Head has completed primary education			0.076	0.208	0.406
Head has 7-9 years schooling			0.150	0.117	0.321
Head has 10-12 years schooling			0.320	0.068	0.252
Head has more than 12 years schooling			0.412	0.076	0.264
Correlation in mean with asset index mean across communities		0.9996	0.9990		
Correlation in Ic with asset index Ic across communities		0.9995	0.9981		

Source: Authors calculations from MMP71

Omitted roof category is adobe and wood, omitted floor category is cement, omitted education category is no education or incomplete primary education.

Land is classified as city or rural land based on the population of the municipality in 1990.

positive coefficients, showing they are associated with other forms of asset ownership, whereas households which have ejido or community land have lower asset index scores. Column 3 then adds indicators for the education of the household head, with more education associated with a higher index score. These expanded asset indices were then used to calculate relative inequality I_c across communities. The foot of Table 6 gives the correlation in I_c across communities between each of the expanded indices and the asset indicator index in Column 1. As can be seen, these correlations are extremely high, showing that the addition of land and human capital does not change very much the relative inequality rankings.[25] We therefore view this evidence as suggesting that the index used in this paper is providing a good proxy for inequality, and is likely to be robust to the inclusion of land.[26]

7. Application: School attendance and inequality

To illustrate that asset indicator-based measures of inequality can give similar results to inequality measured with non-durable consumption, we use the 1998 ENIGH survey to examine the effect of state-level inequality on school attendance of 14–18 year olds in Mexico. Elementary education is compulsory in Mexico, and is normally provided to children between six and fourteen years (SEP, 1999). Lower secondary education became compulsory in 1993, and is given over three years to the population generally aged 12 to 16 years who have completed elementary schooling. In the data we find 74% of 14 year olds report attending school, while only 29% of 18 year olds do. We thus concentrate on the age range 14–18 as the range over which most drop-out occurs.

The question of interest is whether, conditional on household income, household demographics, and state mean levels of income, children in states with more inequality are less likely to be attending school. Theoretically, there are several reasons why this might be the case. From a political economy viewpoint, more equal societies may provide more public goods, such as schooling, and distribute schooling resources more equally, increasing access to schooling. Relative income may determine housing and rental prices, and therefore affect the amount parents can pay for education (see Deaton 2003, for discussion of this mechanism for the health and inequality relationship). Inequality may also represent social divisions, which prevent individuals from lower social classes from obtaining good jobs even if they have an education, lowering the incentives to obtain education. In a dynamic context, Galor and Zeira (1993) show that the distribution of wealth can determine who can invest in education in the presence of capital market imperfections and indivisibilities in human capital investment, with more equal societies enabling more individuals to invest in education. However, Garcia-Peñalosa (1995) suggests that in poor countries, the effect of inequality can be to raise the returns from schooling since skilled labour is scarce, and therefore inequality can increase the proportion of individuals receiving schooling.[27]

Table 7 presents marginal effects from a probit of the probability of an individual aged 14–18 still attending school, separately for males and females, for each of three measures of inequality across states: state-level NDC ginis, state-level predicted NDC ginis from the bootstrap prediction method on all 30 indicators with 25 groups, and state relative inequality I_c measures using

Table 7. Inequality and education in Mexico

	Individual probit model probability that a 14-18 year old attends school						State OLS regression state expenditure per capita on education		
	Males			Females					
Ln household income	−0.736	−0.734	−0.665	−0.382	−0.382	−0.341			
	(4.19)**	(4.15)**	(3.82)**	(2.96)**	(2.95)**	(2.65)**			
Ln hh income squared	0.046	0.046	0.043	0.029	0.029	0.027			
	(4.70)**	(4.68)**	(4.35)**	(3.93)**	(3.94)**	(3.65)**			
Ln hh size	−0.109	−0.105	−0.109	−0.145	−0.142	−0.140			
	(3.08)**	(2.96)**	(3.07)**	(4.23)**	(4.17)**	(4.09)**			
(# boys 0–4)/hh size	−0.649	−0.619	−0.633	−1.048	−1.062	−1.094			
	(3.46)**	(3.30)**	(3.37)**	(6.17)**	(6.25)**	(6.44)**			
(# boys 5–9)/hh size	−0.108	−0.098	−0.099	−0.006	0.002	−0.009			
	(0.74)	(0.67)	(0.68)	(0.04)	(0.01)	(0.06)			
(# girls 0–4)/hh size	−0.460	−0.465	−0.482	−0.538	−0.547	−0.546			
	(2.44)*	(2.46)*	(2.55)*	(3.39)**	(3.44)**	(3.43)**			
(# girls 5–9)/hh size	−0.469	−0.496	−0.484	−0.138	−0.133	−0.130			
	(3.24)**	(3.43)**	(3.34)**	(0.91)	(0.88)	(0.87)			
Dummy for age 14	0.463	0.467	0.467	0.453	0.455	0.459			
	(13.15)**	(13.28)**	(13.29)**	(11.95)**	(11.99)**	(12.12)**			
Dummy for age 15	0.349	0.353	0.352	0.341	0.340	0.343			
	(11.51)**	(11.66)**	(11.64)**	(10.66)**	(10.59)**	(10.70)**			
Dummy for age 16	0.214	0.216	0.218	0.194	0.194	0.195			
	(6.60)**	(6.67)**	(6.73)**	(5.87)**	(5.87)**	(5.89)**			
Dummy for age 17	0.089	0.089	0.089	0.110	0.108	0.110			
	(2.68)**	(2.67)**	(2.67)**	(3.25)**	(3.17)**	(3.23)**			
Dummy for federal district	0.239	0.199	0.187	0.244	0.222	0.259	−245.694	−225.494	−235.386
	(4.51)**	(3.83)**	(3.34)**	(4.87)**	(4.63)**	(5.07)**	(1.67)	(1.53)	(1.48)
State mean NDC	−0.035			−0.029			16.189		
(in 1000s)	(4.17)**			(3.56)**			(1.08)		

State NDC gini	−0.997			−0.624			−934.153		
	(2.17)*			(1.39)			(1.02)		
Bootstrap method state mean NDC (in 1000s)		−0.052			−0.044			38.585	
		(5.54)**			(4.71)**			(2.25)*	
Bootstrap method state gini		−1.776			−0.976			459.536	
		(2.83)**			(1.59)			(0.36)	
State mean asset index			−0.063			−0.047			34.850
			(5.60)**			(4.21)**			(1.50)
State relative inequality (I_c)			−0.370			0.155			−184.946
			(2.46)*			(1.04)			(0.57)
# Observations	2617	2617	2617	2756	2756	2756	32	32	32
Mean of dependent variable	0.533	0.533	0.533	0.486	0.486	0.486	217	217	217
Pseudo-R2	0.1711	0.1747	0.1749	0.1900	0.1927	0.1922			
R2							0.207	0.196	0.124
Effect at the mean of an increase in inequality equivalent to moving from the federal district to Chiapas									
NDC	−0.140*			−0.087			−131		
Bootstrap NDC		−0.194**			−0.106			50	
Asset indicator IC			−0.115*			0.048			−58

Notes: Absolute value of t-statistic in parentheses, * Significant at 5%; ** Significant at 1%
Probit coefficients are the marginal effect at the mean for continuous variables, and the discrete change from a move from 0 to 1 at the mean for dummy variables.
Probits also include dummy variables for cities (100,000 + population), and rural areas (2,500 or less population), and the proportions of males and females aged 10-14, and 15-19 relative to household size.
Source: Own calculations from ENIGH 1998; SEP (1999) for education spending by state.

the overall asset index. The coefficients on the other control variables, such as log household income, age dummies, and household demographics are very similar across inequality measures, suggesting that if our aim was merely to control for inequality while examining the influence of other variables, the methods presented here will give results similar to using the NDC gini. With all three inequality measures, we find that inequality is significantly negatively related to school attendance for boys, but not significantly related to school attendance for girls. The units of measurement for the gini and the I_c differ, so the bottom of Table 7 gives the predicted effect of moving from the level of inequality of the Federal District (one of the most equal states) to the inequality level of Chiapas (one of the most unequal states). The estimated effect for males is reasonably close for all three measures, ranging from a 0.12 drop in the probability of school attendance using the I_c measure to a 0.19 drop using the bootstrap prediction method, with the actual NDC gini predicting a 0.14 drop. The estimated effects for females are all insignificantly different from zero.[28] In this application then, using the asset indicator measures of inequality would lead one to very similar conclusions to using inequality in non-durable consumption.

The last three columns of Table 7 examine one of the possible mechanisms for this relationship, by examining whether more unequal states spend less per capita on education. In 1992 Mexico reintroduced federalism into its education system. On average, states in 1996 provided 19% of total government educational expenditure, with the Federal government providing the remainder. However, there was a lot of variation across states, ranging from Oaxaca, which contributed just 1.9% of the total education expenditure within its territory, to Baja California Norte, which contributed 39.5% (SEP, 1999). The Federal District enjoys a special status, so we have included a dummy variable for this area. We then regress state educational expenditure per capita[29] on state mean levels of wealth or NDC, and on the same measures of inequality as in the probit. All three measures give that inequality does not have a significant effect on state total expenditure on education. Coupled with our finding that inequality matters for school attendance of boys, but not for girls, this suggests that a simple public good provision story is not the mechanism under which inequality is affecting schooling. As labour force participation for males is greater, the effect of inequality on incentives to undertake schooling may potentially explain the difference. We leave examination of these alternative mechanisms to future work.

8. Conclusions

This paper has found that asset indicators can be used to provide reasonable measures of inequality when no data on income or consumption is available. Although the relationship between the asset index and non-durable consumption is stronger in levels than in inequality terms, the results for inequality are still strong enough to conclude that asset indicators provide a viable method of inequality measurement. Care needs to be taken in practice to ensure that sufficient indicators are used to prevent clumping of households at just a few levels of the asset index, and to avoid truncation and thereby allow inequality among the poor to be measured. If inequality in non-durable consumption is truly the object of interest, and an auxiliary survey is available, then the bootstrap prediction method provided here can be used to predict inequality in the main survey of interest. Using all indicators separately in the prediction regression, and groups in the bootstrapping stage, gave the best results.

When the inequality measures are applied to Mexican data to examine the relationship between school attendance and state-level inequality, we find a significant effect of inequality on the probability that a boy aged 14–18 attends school, with more inequality being associated with lower attendance. The effect of the difference in inequality between the Federal District and Chiapas is estimated to be a 0.12–0.14 difference in school attendance rates for boys. Inequality is not found to have a significant effect on the school attendance of girls, or on state spending per capita on education, suggesting that a simple political economy story of inequality affecting educational provision is not the cause of the relationship found. More research is needed to examine the influence of other mechanisms for this effect.

The ability to measure inequality from asset indicators suggests a number of other possible research applications. The Demographic and Health Surveys (DHS) could be used to examine the influence of inequality on a wide range of socioeconomic outcomes, such as health care provision, health outcomes, and fertility choice. The analysis of Filmer and Pritchett (1999) could be extended to examine the relationship between inequalities in education and inequalities in wealth. In McKenzie and Rapoport (2004), the method provided in this paper is used with the Mexican Migration Project data (which has detailed migration data and asset indicators, but no income or consumption data) to examine the relationship between inequality and migration in Mexico. Given renewed interest in the functional aspects of inequality in economic development, and the cost effectiveness of collecting asset indicator information, it is likely that the methods here will be of use in a number of other potential applications.

Appendix

Proof (Lemma 1). The first three conditions are easily seen. To show that the transfer property holds, let $\widetilde{y}_i = y_i - \delta$ and $\widetilde{y}_j = y_j + \delta$. Then

$$
\begin{aligned}
&(\widetilde{y}_i - \overline{y}_c)^2 + (\widetilde{y}_j - \overline{y}_c)^2 \\
&\quad = (y_i - \overline{y}_c)^2 + (y_j - \overline{y}_c)^2 - 2\delta^2\left[(y_i - \delta) - y_j\right] \\
&\quad < (y_i - \overline{y}_c)^2 + (y_j - \overline{y}_c)^2 \text{ since } y_i - \delta > y_j + \delta
\end{aligned}
$$

From this it follows that $\widetilde{\sigma}_c^2 < \sigma_c^2$, where $\widetilde{\sigma}_c^2$ is the sample variance in community c after the transfers. With some algebra one can then show that this implies that $\widetilde{I}_c < I_c$ unless there is only one community, in which case $\widetilde{I}_c = I_c = 1$.

Endnotes

1 Data are available at http://www.measuredhs.com/
2 See www3.who.int/whs/ for details.

[3] See http://www.ssc.upenn.edu/lamp/ for the LAMP data and http://www.pop.upenn.edu/mexmig for the MMP data.

[4] See www.idea.int/balkans.

[5] See Skoufias et al. (1999) for a thorough description of the selection process used and an evaluation of the targeting performance of the discriminant analysis carried out using the asset indicators.

[6] Thanks to a referee for making this point.

[7] If data on rental prices for households of different characteristics were available, standard hedonic regression methods could be used to estimate prices of housing infrastructure. However, rental values are not commonly found in the types of surveys used as motivation for our paper. Even if rental values were available, they are less likely to cover the poorest houses, making valuation of basic infrastructure difficult.

[8] See Chapt. 3 of Everitt and Dunn (2001) for a good introduction to the methodology used to calculate the vector a.

[9] This is our motivation for concentrating on principal components analysis. In practice we found very high correlations (around 0.99) between the score of the first principal component and the score of the first principal factor using principal factor analysis. Such a result is to be expected if the specific variances are small, e.g., if the variability in fridge ownership not shared with variability in ownership of other assets is small (see p. 287 of Everitt and Dunn 2001 for further discussion of this point).

[10] To see this note that the principal component score for household i is $y_i = a'\widetilde{\mathbf{x}}_i$, and hence the variance of y_i equals $a'Va$, where V is the correlation matrix of the $\mathbf{x}$. But then since λ is the eigenvalue of V corresponding to the eigenvector a, we have $Va = \lambda a$, and hence $a'Va = a'\lambda a = \lambda$ since by restriction $a'a = 1$.

[11] See, for example, Fields (2001), Chapt. 2 for a standard exposition of these properties.

[12] See Deaton (1997), Chapt. 1.2 for a good discussion of these issues.

[13] Since most of the variables in x are likely to be dummy variables, it is unlikely to be possible to include many non-linear terms in x to allow further for non-linearities in the mapping from wealth to non-durable consumption. Squared terms for variables such as the number of rooms in the house can be included in x here.

[14] If the residuals ε_i represent only transitory consumption fluctuations, then this measure may be thought of as measuring longer-run inequality.

[15] For example, with 17 binary asset indicator variables we would have $2^{17} = 131{,}072$ cells!

[16] We do not use some of the least common assets, such as ownership of boats, vehicles pulled by animals, video game machines and handmills, although our results are robust to the addition of these extra variables.

[17] Kernel densities were estimated in STATA using the Epanechnikov kernel with the default, data-determined bandwidth. The number of bins in the histograms are chosen as $\min\left(\sqrt{N}, \frac{10\ln(N)}{\ln(10)}\right)$, where N is the number of households in the sample.

[18] The height of the bars in the histogram is scaled so that the sum of their areas equals one, which is why particular bars in the utilities index histogram have height greater than one.

[19] The correlation with the actual NDC gini was 0.768 for the bootstrap on all 30 assets with $G = 2$ groups, with rank-order correlation of 0.803, which are about 0.05 lower than the correlations with 10–100 groups shown in Table 4.

[20] STATA was used to calculate these bootstrapped confidence intervals, with 100 replications used in the bootstrap.

[21] See, e.g., Fields (2001, p. 30) for the formulae for these measures.

[22] Full details of the methodology and the data can be accessed at http://www.pop.upenn.edu/mexmig.

[23] See McKenzie and Rapoport (2004) for more discussion of this data and the asset indicators in this data.

[24] The MMP asks about only a subset of the durables goods measured in the ENIGH, and the housing quality indicators combine roofing information together and omit characteristics of the walls.

[25] There was also only small changes in the level of the I_c, with the addition of land and education slightly increasing I_c for a few communities and slightly reducing it for a few others.

[26] The ENIGH survey does contain data on the education of the household head. As with the MMP, adding this to the overall index hardly changed the levels or rankings of the I_c's across

states: the correlation in I_c across communities between the overall index and this index with head's education added was 0.998.

[27] Current school attendance will also affect *future* inequality since income will vary with education. As this application is mainly illustrative, we treat inequality as predetermined and do not consider reverse causation issues here.

[28] Overall levels of school attendance are lower for girls than for boys. See Parker and Pederzini (2000) for an examination of the causes of these gender differences in education in Mexico.

[29] Total State expenditure on education for 1996 is taken from SEP (1999), and divided by state populations in 1995, taken from INEGI (2001).

References

Banerjee AV, Newman AF (1993) Occupational Choice and the Process of Development. *Journal of Political Economy* 101(2):274–298

Bardhan P, Bowles S, Gintis H (1999) Wealth Inequality, Wealth Constraints and Economic Performance. In: Akkison AB, Bourguignon F (ed), *Handbook of Income Distribution*. Elsevier, Amsterdam

Birdsall N, Londoño JL (1997) Asset Inequality Matters: An Assessment of the World Bank's Approach to Poverty Reduction. *American Economic Review Papers and Proceedings* 82(2):32–37

Blundell R, Pistaferri L, Preston I (2003) *Partial Insurance, Information, and Consumption Dynamics*. Mimeographed, Stanford University

Deaton A (2003) Health, Inequality, and Economic Development. *Journal of Economic Literature*, 41(1):113–158

Deaton A (1997) *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. The World Bank, Washington, DC

Deaton A, Grosh M (2000) Consumption. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study, Vol. 1*. The World Bank, Washington, DC

Everitt BS, Dunn G (2001) *Applied Multivariate Data Analysis*, 2nd ed. Oxford University Press, New York

Fields GS (2001) *Distribution and Development: A New Look at the Developing World*. The MIT Press, Cambridge, MA

Filmer D, Pritchett LH (2001) Estimating Wealth Effects Without Expenditure Data-Or Tears: An Application to Educational Enrollments in States of India. *Demography* 38(1):115–132

Filmer D, Pritchett LH (1999) The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries. *Population and Development Review* 25(1):85–120

Galor O, Zeira J (1993) Income Distribution and Macroeconomics. *The Review of Economic Studies* 60:35–52

Garcia-Peñalosa, C (1995) The Paradox of Education or the Good Side of Inequality. *Oxford Economic Papers* 47:265–285

Instituto Nacional de Estadística, Geografía e Inform ática (INEGI) (2001) *Indicadores Sociodemográficos de México*. INEGI, Aguascalientes

McKenzie DJ, Rapoport H (2004) Network effects and the dynamics of migration and inequality: theory and evidence from Mexico. Stanford Center for International Development Working Paper 201

Minujin A, Bang JH (2002) Indicadores de Inequidad Social. Acerca del Uso del Indice de Bienes' para la Distribución de los Hogares. *Desarrollo Económico* 42(165):129–146

Montgomery MR, Gragnolati M, Burke KA, Paredes E (2000) Measuring Living Standards With Proxy Variables. *Demography* 37(2):155–174

Parker, SW, Pederzini C (2000) Gender Differences in Education in Mexico. World Bank Departmental Working Paper, Report No. 21023

Pollitt E, Gorman KS, Engle PL, Martorell R, Rivera J (1993) 'Early Supplementary Feeding and Cognition. *Monographs of the Society for Research in Child Development* 58(7): 235

Schroeder DG, Kaplowitz H, Martorell R (1992) Patterns and Predictors of Participation and Consumption of Supplement in an Intervention Study in Rural Guatemala. *Food and Nutrition Bulletin* 14(3):191–200

Secretaría de Educación Pública (Ministry of Public Education) (SEP) (1999) Profile of Education in Mexico. http://www.sep.gob.mx/work/appsite/Perfil_Ingles/index_english.htmw, accessed June 16, 2003

Skoufias E, Davis B, Behrman J (1999) An Evaluation of the Selection of Beneficiary Households in the Education, Health and Nutrition Program (PROGRESA) of Mexico, Final Report. mimeo. International Food Policy Research Institute, Washington, DC

Skinner J (1987) A Superior Measure of Consumption from the Panel Study of Income Dynamics. *Economics Letters* 23:213–216