

Estimating Quarterly Poverty Rates Using Labor Force Surveys: A Primer

*Mohamed Doudich, Abdeljaouad Ezrari, Roy Van der Weide,
and Paolo Verme*

This paper builds on the existing cross-survey imputation literature to provide up-to-date estimates of poverty when official estimates are deemed outdated. This is achieved by imputing household expenditure data into Labor Force Surveys (LFSs) with models that have been estimated using Household Expenditure Surveys (HESs). In an application to Morocco, where the latest official poverty rate is for 2007, estimates of poverty are obtained for all years (and quarters) between 2001 and 2009. It is found that the approach accurately reproduces the official poverty statistics for the two years these surveys are available. The imputation-based estimates furthermore reveal that poverty has consistently declined over the entire 2001–2009 period. This would suggest that poverty reduction in Morocco was not halted by the global financial crisis. While our focus is on head-count poverty, the method can be applied to any welfare indicator that is a function of household income or expenditure, such as the poverty gap or the Gini index of inequality. JEL codes: D6, H53, I3, R13

The estimation of poverty in any given country relies on household surveys that contain data on income or expenditure (Household Expenditure Surveys, or HESs for short). This data is hard to collect and requires elaborate and time-consuming questionnaires that result in costly surveys. For this reason, statistical agencies worldwide have taken to the practice of administering relatively small surveys (usually in between 5,000 and 10,000 households) administered at intervals of several years (usually every 4–5 years).

Paolo Verme (corresponding author) is a senior economist at the World Bank; his email address is pverme@worldbank.org. Mohamed Doudich is engineer general, High Commission for the Plan, Morocco; his email address is doudich@yahoo.fr. Abdeljaouad Ezrari is chief of department, High Commission for the Plan, Morocco; his email address is ezrari@yahoo.fr. Roy Van der Weide is an economist at the World Bank; his email address is rvanderweide@worldbank.org. The work was undertaken in the framework of the cooperation agreement between the World Bank and the High Commission for the Plan of Morocco and under the World Bank program “Growth, Employment and Poverty” (EW-P127927-ESW-BB). It also received financial support from the World Bank’s Knowledge for Change Program, for which the authors are grateful. The authors are also grateful to three anonymous referees, Peter Lanjouw, Tomoki Fujii, Nobuo Yoshida, Roy Katayama, Gladys Lopez, and participants in several workshops in Washington and Morocco where the paper was presented for useful comments. A supplemental appendix to this article is available at <http://wber.oxfordjournals.org/>.

THE WORLD BANK ECONOMIC REVIEW, VOL. 30, NO. 3, pp. 475–500

doi:10.1093/wber/lhv062

Advance Access Publication December 12, 2015

© The Author 2015. Published by Oxford University Press on behalf of the International Bank for Reconstruction and Development / THE WORLD BANK. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

This practice is sensible from a logistics and cost perspective but has two main drawbacks for the measurement of poverty. The first is that small surveys can provide reliable statistics only for highly aggregated areas such as rural and urban areas or large subnational regions. And the second is that poverty statistics can only be produced in conjunction with the HESs surveys every several years, leaving researchers with no information on poverty for the periods in between any two surveys or beyond the most recent survey.

To address these two shortcomings, we advocate the use of imputation methods. Imputation methods have a long history in statistics and economics and have been used to address a variety of missing data problems, see for example, [Rubin \(1978, 1987\)](#). While originally conceived to fill data gaps within surveys, these methods have also been extended to cross-survey imputation, where one survey is used to fill data gaps of another survey belonging to the same population. A recent review of these methodologies by [Ridder and Moffitt \(2007\)](#) shows how widespread these methodologies have become, and how they can be adapted to respond to different types of missing data problems. See also [Fujii and van der Weide \(2013, 2014\)](#) and the references therein.

In the context of poverty analyses, imputation methods have found numerous applications to address data gaps and statistical inference problems across space and time. For example, [Elbers et al. \(2002, 2003, 2005\)](#) combine census and survey data to estimate poverty and inequality for areas considered too small for statistical inference with survey data alone, an exercise known as “poverty mapping.” For an overview of the statistics literature on small area estimation, see [Rao \(2003\)](#) and [Elbers and van der Weide \(2014\)](#). One could also work with large surveys instead of censuses such as Demographic and Health Surveys (DHSs) to obtain imputation-based poverty estimates. Examples of the latter are [Stifel and Christiaensen \(2007\)](#) and [Grosse et al. \(2009\)](#), who used DHSs to estimate poverty in Kenya and Bolivia, respectively. More recently, [Christiaensen et al. \(2012\)](#) used data from Vietnam and China to create a synthetic panel out of repeated cross-section HESs using imputation methods (see also the recent study by [Dang et al. 2014](#)).

Cross-survey imputation techniques have also been used to resolve problems of comparability between surveys of the same type over time, for example due to changes in the questionnaires. [Kijima and Lanjouw \(2003\)](#) and [Tarozzi \(2007\)](#) have re-estimated poverty rates in India using imputed data in an effort to validate the official figures on poverty during what became known as the “great Indian poverty debate” (see e.g., [Deaton and Kozel, 2005](#)). Note that all of these approaches, the small area estimation approach put forward by [Elbers et al. \(2003\)](#) included, are nested as special cases in the multiple imputations framework (see [Rubin, 1987](#)).

These examples show that cross-survey imputation methods can be used effectively to improve both the levels of *disaggregation* and the *frequency* at which statistics of interest can be obtained. This is of great relevance to both research and policy making. Imagine a country that may have been affected by the 2008 global financial crisis and the 2001 crisis before that. Suppose however that the only

official poverty estimates that are available over this extended period of time are for the years 2001 and 2007, while the next official poverty estimates will not become available until 2014. This means that until 2014 policy makers will have no data on how poverty has evolved post-2007. By 2014, when the new official poverty estimates are published, the effects of the 2008 crisis may not be visible any longer, and the window of opportunity for intervention may have passed. The same argument can be applied to macroeconomic shocks that occurred prior to 2007, which include the 2001 financial crisis. Relying solely on official poverty estimates derived from these household surveys does not allow us to identify changes in poverty in between these survey years, which can be far apart.

In this paper, we build on the existing cross-survey imputation literature to improve the *frequency* of poverty estimates, that is, to provide up-to-date estimates of poverty when official estimates are deemed outdated. The approach takes advantage of existing household surveys that are available at a higher frequency than the HES. It is required that these alternative surveys have covariates in common with the HES that are sufficiently correlated with household expenditure (think of variables on demographics, education, and employment). In many countries, Labor Force Surveys (LFS) meet these criteria. We will demonstrate the approach by means of an application to Morocco, a country that represents an ideal test case. Morocco implemented consumption surveys in 2001 and 2007 and is expecting to complete the next survey in 2014. The official estimates derived from these two surveys suggest that poverty has been on the decline in Morocco. However, much has happened since 2007 as well as between 2001 and 2007, which includes two global financial crises (2001 and 2008).¹ Unfortunately, policy makers in Morocco are presently unable to verify whether poverty has continued its decline, has stagnated, or has possibly even reversed its trend in response to these events.

The empirical question we wish to address is whether we can use Labor Force Surveys (LFSs) to fill these gaps and, by doing so, connect the dots of poverty estimates in Morocco for the period between the last two expenditure surveys (2001–2007) and beyond. To the best of our knowledge, this is the first comprehensive exercise of cross-survey imputation that uses LFS data to construct a time-series of poverty estimates spanning a decade. Note that, if the proposed methodology proves successful, it can be applied to countries worldwide, wherever there is a need for it. In the event that LFS data are not available for a country or time-period of interest, one could consider exploring alternative household surveys, such as the Demographic and Health Survey (DHS).

Another potential advantage of LFSs is that they often provide nationally representative samples at a quarterly frequency. Since HESs are generally not

1. Note that poverty may also exhibit important seasonal fluctuations that the annual poverty estimates cannot reveal. A few countries have attempted to address this problem by producing statistics on poverty at an intra-annual level. For example, Peru experimented with the administration of quarterly consumption surveys, while Mexico produces a proxy of income poverty every month using Labor Force Surveys. However, collecting survey data quarterly is evidently very expensive, while many countries do not collect income data together with labor data, making these efforts difficult to replicate elsewhere.

representative at this frequency, relying on cross-survey imputation might provide a means of detecting intra-year fluctuations in poverty. Estimating quarterly poverty rates may seem redundant but there are several reasons why statistical agencies, researchers, and policy makers might be interested in increasing the frequency of poverty estimates. First, economic shocks often unfold over short periods of time. Having an early assessment of the impact of these transient shocks on poverty may be valuable to policy makers who have to decide on the need for compensatory measures. Second, it is conceivable that annual estimates of poverty hide seasonal fluctuations. Third, macroeconomic variables are generally available at a higher frequency than micro-economic welfare variables are. Increasing the frequency of poverty and inequality estimates, thereby synchronizing them with macro data, might expand the potential for longitudinal micro-macro research. Naturally, underlying such high-frequency imputation-based estimates of poverty will be a set of assumptions. For example, one will have to assume that intra-year changes in poverty can be traced back to changes in the poverty predictors used (such as changes in employment variables). Furthermore, the magnitude of intra-year fluctuations may be dampened by the fact that households will try to smooth their consumption over the course of a year.

The application of this approach to Morocco shows encouraging results. We estimated both annual and quarterly poverty rates with LFSs for the period 2001–2009 using two models estimated using 2001 and 2007 expenditure data, respectively. Despite the fact that the two models are six years apart, we found that the models produced nearly identical poverty trends over the period under consideration. Note that the years 2001 and 2007 can be used to further validate the approach. For each of these years, the imputation-based poverty estimates (using any of the two prediction models) can be compared to the official survey direct estimates that are based on the HES alone. For both years, and regardless of the choice of prediction model used, the imputation-based estimates were able to reproduce the official poverty rates with remarkable accuracy. The *forward* approach (where 2007 poverty is estimated by imputing into the 2007 LFS using a model estimated with 2001 data) and the *backward* approach (where 2001 poverty is estimated using a model estimated with 2007 data) effectively led to the same poverty estimates. In a recent application to Uganda, Mathiassen (2013) found imputation-based estimates of poverty to be equally robust to the survey year chosen for estimation of the prediction model, although in that application poverty was predicted back into the HES, not into an alternative survey such as the LFS.²

We obtain a number of new insights for Morocco. The imputation-based estimates show that poverty consistently declined between 2001 and 2007 and that this decline continued beyond 2007 up to 2009. This confirms that the progress being made in Morocco with reducing poverty was not halted by the global financial crisis, possibly helped by increases in agricultural production during

2. For a full account of macroeconomic changes, economic reforms, and labor market changes that took place in Morocco between 2000 and 2009, see Verme et al. (forthcoming).

that same time period. The estimates also show an urban-rural convergence in poverty, with rural poverty falling faster than urban poverty, thereby reducing the urban-rural gap. Interestingly, the rates of poverty-reduction exhibit a fair degree of heterogeneity across different subregions and subgroups of the population. This heterogeneity found in the imputation-based poverty estimates is consistent with the survey direct estimates obtained for the years 2001 and 2007. Unfortunately, we were not able to identify meaningful changes in poverty within the years. Whether this is because the quarterly fluctuations were indeed of a limited magnitude or whether our approach falls short of capturing the fluctuations on that time-scale is a question we leave for future research.

The paper is organized as follows. Section I presents the cross-survey imputation methodology adopted. Section II introduces the HESs and LFSs data used. The empirical model, the validation tests, and the main results, namely our estimates of the poverty trends for the entire 2001–2009 period, are presented in sections III, IV and V, respectively. Section VI concludes.

I. METHODOLOGY

We adopt a standard imputation approach that is commonly used in the case of missing data. When a variable of interest is missing altogether in a given data set, one can still proceed with imputing this variable provided that a second data set representative of the same population is available that does contain the variable of interest. This second data set is needed to identify a prediction model that can be used to generate the imputed values in the primary data set. A prerequisite is that the two data sets share a set of covariates that are sufficiently correlated with the missing variable.

Consider the following standard linear regression model for log of household expenditure per capita:

$$\ln(y_{it}) = x_{it}^T \beta_t + u_{it}, \quad (1)$$

where x denotes a vector of independent variables (e.g., variables on demographics, education, employment, housing conditions, asset ownership) including the constant, u denotes a vector of independent errors with zero expectation, and the subscripts i and t indicate household i and time t . The superscript ‘T’ indicates matrix transpose.

We have two types of data sets: Household Expenditure Survey (HES) data and Labor Force Survey (LFS) data. Both types of surveys contain the regressors x , but only the HES contains the household expenditure y and only for selected years. In our case, we will consider the period 2001–2009, for which we have two years of the HES (2001 and 2007) and the full period of the LFS. The objective is to use model (1), estimated using the 2001 and 2007 HESs, to impute household expenditure into the LFSs for all available years (2001 to 2009) and

then use the imputed expenditure data to estimate poverty for the entire 2001–2009 period.

We will be relying on a number of assumptions:

Assumption 1 The model is time-invariant, that is, $\beta_t = \beta$, when all variables that measure monetary value (i.e., expenditures, incomes, and asset values) are expressed in constant prices.

Note that the model is estimated using data from one time period and then adopted for imputation in another time period. Under assumption 1, this disconnect will have no bearing on the results, since the model underlying the data used for estimation and the model underlying the data used for imputation are one and the same. If the model is however subject to some variation over time, that is, if the assumption does not hold true, then ignoring this variation will introduce a degree of model error. Note that this assumption can be tested both directly and indirectly, provided that one has access to more than one HES. For a direct test, one could use any test statistic that evaluates the difference between the model-coefficients estimated to the different HESs. An indirect test can be obtained by estimating the model using one of the two HES years and then applying this model to obtain an imputed poverty rate for the other year. Since both years also allow us to compute the actual poverty rates based on observed data, we are able to verify how well the observed data compares to the imputed data. The results of this test for our data can be found in section 5. If the test would indicate that the imputation-based poverty estimates are subject to a bias, then one has the option to relax assumption 1; one could explicitly incorporate a time-trend in the model coefficients. Estimation of the unknown model parameters in this case of course requires that one has at least two HESs. In the event that the modeler has access to only one round of the HES, then testing the validity of the assumption is no longer an option.

Remark 1 It would arguably be most practical if all variables that measure monetary values (such as expenditures and value of assets) were expressed in constant prices. If, for whatever reason, one prefers to measure values in time t prices, then it is recommended that one does not mix value with nonvalue independent variables (i.e., mixing value of assets with count of assets). To see why this matters, let us consider a stylized example where we refer to the model parameters as betas. Suppose that one variable measures the value of owned bicycles, while another variable simply checks whether the household owns a car in the form of a dummy variable. The beta associated with car ownership will then measure the value added by the car to the total household expenditure, while the beta associated with the bicycles will be unit-less as it simply passes on (by some factor) the value of the bicycles to the value of total household expenditure. Note that the beta attached to the car ownership will be expressed in time t prices. If we now apply this model to predict household expenditure at time $t + 1$, then the beta for the car times the car dummy measures a contribution in time t prices

while the beta for the bicycles times the value of the bicycles measures a contribution in time $t + 1$ prices (since the beta is unit-less and the value of the bicycles in this case is expressed in time $t + 1$ prices).

Assumption 2 The error term u is homoscedastic and normally distributed.

This assumption too can be relaxed. Nonnormality could be accommodated in a variety of ways; one could draw the errors from the empirical distribution of residuals (see e.g., [Filmer and Pritchett 2001](#); and [Elbers et al. 2003](#)), or one could fit a mixture distribution to the errors (see e.g., [Elbers and van der Weide 2014](#)). Heteroskedasticity could also be accommodated in a number of ways; one could work with a random coefficient model (see e.g., [Hsiao 1975](#); [Breusch and Pagan 1979](#)), or one could model the error variances more directly (see e.g., [Elbers et al. 2003](#)). Our recommendation is to start off simple with a model that is based on assumptions 1 and 2, but then consider relaxing these assumptions in case the data calls for it. When the imputation-based estimates are consistent with the survey direct estimates for the available benchmark/HES years, and this consistency is maintained when one further disaggregates the poverty estimates into different subgroups, then one could make a case for not adding more flexibility to the model. It is when the imputation-based estimates are off, that one is compelled to revisit the assumptions one by one in an effort to identify the source of the discrepancy. Incorrectly assuming normal errors, for example, certainly has the potential to introduce a meaningful bias when estimating poverty or inequality (see [Elbers and van der Weide 2014](#)). Arguably, the same could be said for ignoring heteroskedasticity. In other words, if the model used for imputation is subject to any serious misspecification, then one should expect to see a bias in the imputation-based poverty estimates. In our case, a model based on assumptions 1 and 2 fits the data remarkably well as our empirical results for Morocco will show.

Let $W(y, m; z)$ denote a welfare indicator that can be expressed as a function of all household expenditures y , household size m , and some poverty line z (note that not all welfare indicators will need all of these inputs; average income and many standard measures of inequality for example do not require a poverty line). We are interested in estimating the expected value of W , that is, $E[W]$, given a sample of households. In case we observe the expenditures for the households in the sample, then the standard estimator for $E[W]$ would be the sample direct estimator. For headcount poverty, which represents a popular example of the welfare indicator, the sample direct estimator would take the form:

$$\hat{H}_t = \sum_i w_{it} m_{it} 1(y_{it} < z_t) / \sum_i w_{it} m_{it},$$

where $1(\cdot)$ denotes the standard indicator function that equals 1 if the argument is true and 0 otherwise and where w denote the household sampling weights. Note that there may be other sampling design parameters that will feature in the

estimation of statistical precision. The only source of statistical error in this case is sampling error. Let us denote the estimate of the sampling variance by $U_n^{(0)}$, which declines with the sample size n .

Consider now the case where we do not observe expenditure, where we will be working with imputed expenditure data instead. Accordingly, sampling error will no longer be the sole source of error in this case. The imputation-based estimator will also be subject to model error. Imputing the expenditure data multiple times for the sample of households denotes a practical way of accounting for this added source of error. In each imputation round, we draw a new set of model parameters and household errors from their estimated distributions and use these to impute expenditure.³ If we repeat this R times, we obtain R simulated data sets, and consequently R estimates of the headcount poverty rate. The imputation-based estimator takes on the following form:

$$\tilde{H}_t^{(r)} = \sum_i w_{it} m_{it} 1(\tilde{y}_{it}^{(r)} < z_t) / \sum_i w_{it} m_{it},$$

where:

$$\tilde{H}_t^{(r)} = \sum_i w_{it} m_{it} 1(\tilde{y}_{it}^{(r)} < z_t) / \sum_i w_{it} m_{it},$$

where $\tilde{y}^{(r)}$ denote the simulated expenditures from imputation round r . Note that \tilde{H}_t is an estimate for $E[\tilde{H}_t^{(r)}]$. Let $U_n^{(r)}$ denote the estimated sampling variance associated with $\tilde{H}_t^{(r)}$. An estimate of the total variance (or standard error), which would account for both sampling error and model error, can be obtained by appealing to the law for total variance:

$$\begin{aligned} \text{var}[\tilde{H}_t^{(r)}] &= E[\text{var}[\tilde{H}_t^{(r)} | \tilde{y}^{(r)}]] + \text{var}[E[\tilde{H}_t^{(r)} | \tilde{y}^{(r)}]] \\ &\cong \frac{1}{R} \sum_r U_n^{(r)} + \frac{1}{R} \sum_r (\tilde{H}_t^{(r)} - \tilde{H}_t)^2. \end{aligned}$$

The first component in this variance decomposition captures the sampling variance while the second component captures the contribution to the variance that is due to model error (or imputation error).

3. Note also that Mathiassen (2013) opted to evaluate statistical precision analytically, accounting for both model error and sampling error, instead of appealing to multiple imputations. The potential advantage of this approach is that it is computationally fast, which would be most noticeable when dealing with large datasets. For the majority of household surveys, and given today's computational power, this advantage will be mostly lost. The downside is that it is rather inflexible; each new welfare measure would call for a new analytical derivation in order to obtain the appropriate standard errors, which is a nonnegligible cost.

Many of the popular statistical software packages, including Stata, provide multiple imputation routines, which can compute point estimates and corresponding standard errors. We conducted our empirical application using Stata's MI package. For an elaborate treatment of the multiple imputation (MI) approach we refer the reader to [Rubin \(1987\)](#). Note that the poverty mapping approaches such as the approach put forward by [Elbers et al. \(2003\)](#) are nested as a special case within the MI framework.

Remark 2 In theory it is possible that the standard errors associated with the imputation-based estimator are smaller than the standard errors associated with the sample direct estimator. It is important to realize however that any such improvement in statistical precision is obtained under the assumption that the model used for imputation accurately reflects the model underlying the real data. The imputation-based estimator has the potential to gain statistical precision over the direct estimator from the fact that it utilizes more information; it utilizes knowledge of the data-generating process as well as data on the covariates used for imputation (both of which are ignored by the sample direct estimator). The general rule in statistics is that more information will result in more precise estimates, provided that this information is correct and efficiently utilized. For mathematical proofs that imputation-based estimators can indeed be more precise than sample direct estimators, see for example, [Matloff \(1981\)](#) and [Fujii and van der Weide \(2013, 2014\)](#). It is of course possible (or even plausible) that the model used for imputation is subject to some degree of misspecification, in which case some of the information on which the imputation estimator is based is not entirely correct. Unfortunately, there is no obvious way of adjusting standard errors for potential model misspecification error. Standard errors are always obtained under the assumption that the model on which they are based is correctly specified. It is the modeler's responsibility to identify the appropriate model. Note that another circumstance under which the imputation-based estimator may top the sample direct estimator in terms of precision is when the household expenditures are imputed into a survey that is considerably larger than the HES (as this will reduce the sampling variance component), which generally applies in the case of LFSs.

Remark 3 The poverty line must be measured in the same time t prices as the expenditure (or income) data that was used as the dependent variable in the regression model. In other words, if a model is estimated with 2001 household expenditure data (measured in 2001 prices) and subsequently applied to a 2007 LFS, then the imputed data will represent 2007 expenditures measured in 2001 prices. As a result, the imputed expenditure data will have to be compared to the 2001 poverty line in order to obtain an estimate of 2007 poverty. Expressing all household expenditures in constant prices will simplify matters, as one will have only one poverty line to work with. Note that relative poverty lines can also be accommodated. For example, the poverty line may be defined as a percentage of

median expenditure. The median expenditure level in this case is obtained as the median of the imputed expenditure data.

Remark 4 The approach has the best chance of identifying changes in poverty over time if these changes can be traced back to changes in the observed independent variables (such as changes in education levels, employment characteristics, and housing conditions) as opposed to changes driven by exogenous shocks that are not well captured by the observed data. Naturally, the ability to capture short-term fluctuations in household welfare depends crucially on the availability of independent variables that exhibit fluctuations at that time-scale. Ideally these are variables that are either responsible for or have responded to the changes in household welfare. Among the variables typically available in LFSs (or similar household surveys), the following variables are arguably best suited for capturing medium- to short-term fluctuations: labor force status, type of employment, and selected asset variables. Examples of variables that are arguably less ideal for capturing short-term fluctuations include demographics, education, and dwelling characteristics. Not because of a lack of correlation with household expenditures (they are often found to be highly correlated) but because they tend to be more stable over time. Independent variables that are candidates for tracking short-term fluctuations are arguably smaller in number (compared to variables that are well-suited for tracking longer-term welfare). Whether the available data is sufficient for monitoring poverty at an annual and quarterly frequency is an empirical question.

Remark 5 Finally, while endogeneity (reverse causality) will generally bias estimates of the model parameters, it does not necessarily bias the imputed values. In fact, endogeneity may benefit the statistical precision of the imputed data since a nonzero correlation between the independent variables and the error term implies that the error term is now not entirely unpredictable.

II. DATA

We use two sets of surveys, the Household Expenditure Surveys (HESs) and the Labor Force Surveys (LFFs). The HESs are the surveys that contain our variable of interest (household expenditure) and they are used to construct and estimate the model that we rely on for imputation. The LFFs denote the surveys that are used to estimate poverty based on imputed data for time periods that are not covered by the HESs. Strictly speaking, the HESs in Morocco include two different surveys; the 2000–2001 National Survey on Consumption and Expenditure (NSCE) and the 2006–2007 National Living Standards Survey (NLSS). Both surveys measure household expenditure and are nationally and regionally representative as well as representative of urban and rural areas.

The 2000–2001 NSCE covers 15,000 households and was administered between November 2000 and October 2001 with multiple objectives in mind. It

was designed to measure household expenditure and to provide the necessary information to weigh the living-standard index constructed for Morocco and other national accounts aggregates. It was also designed to measure household consumption, nutrition, poverty, and inequality. The questionnaires include sections on socio-economic characteristics, habitat, energy, economic activities, education, health, transfers, subjective indicators of wellbeing, expenditure, durable goods, anthropometrics, nutrition, and also a module administered to the community to measure access to services.

The 2006–2007 NLSS covered 7,200 households and was administered between December 2006 and November 2007. The survey focused on household expenditure and revenues and was principally administered to measure poverty, inequality, and other dimensions of living standards. The questionnaire includes modules on socio-demographic characteristics, social mobility, habitat, expenditures, revenues, credits, transfers, education, health, employment, durable goods, and poverty perceptions.

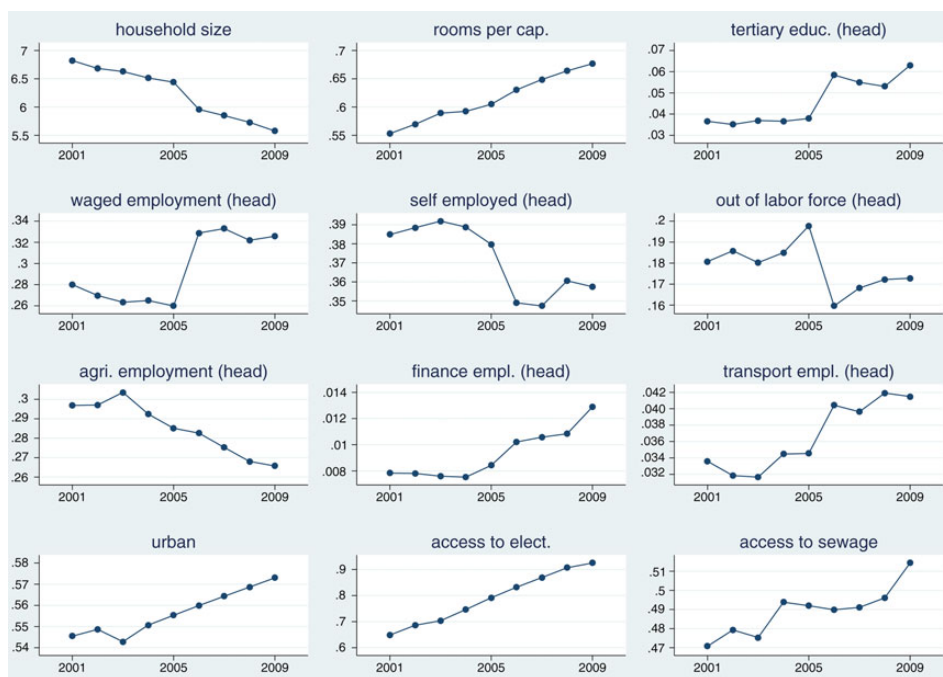
The Labor Force Survey (LFS) of Morocco is a household survey originally launched in 1976. The survey has been developed in four successive phases (1976–1982, 1984–1993, 1995–2005, and 2006–) with progressive improvements in the sampling frame, sample size, questionnaire design, and administration. In 2007, it introduced Computer Assisted Personal Interviewing (CAPI) devices, which allow for data verification and error corrections in real time. Each quarterly sample of the LFS is representative at the national and regional level and, within regions, at the urban and rural level.

All HESs and LFSs in Morocco are based on the master sample frame that is derived from the latest population census; the 1994 census for surveys conducted up to 2005, and the 2004 census for surveys conducted after 2005. The surveys also share the same stratification structure. For urban areas, strata include the region, province, city size (large, medium, and small) and type of housing (“lux”, “modern”, “old medina”, “new medina,” and “clandestine”). For rural areas, the strata are regions and provinces.

The 2001 NCSE follows a two-stage sampling process. In the first stage, 1,250 Primary Sampling Units (PSUs) of approximately 300 households each were extracted from the 1994 population census. In the second stage, a dozen households per PSU were extracted randomly to constitute the final sample. The 2007 NLSS follows a three-stage sampling process. In the first stage, 1,848 PSUs of approximately 600 households each were selected from the 2004 population census. The second stage subdivided each PSU into twelve Secondary Sample Units (SSUs) representing about fifty households each, and then randomly selected six of the twelve SSUs from each PSU. In the third stage, a constant number of households were selected randomly from each SSU. The LFS follows a sampling process that is similar to that of the 2007 NLSS.

Let us have a first look at the LFS data. To have any chance of capturing changes in poverty over time, one needs a set of poverty predictors that can match these changes. Figure 1 plots the time-trends for a selection of twelve

FIGURE 1. LFS Variables Over the Period 2000–2009



Source: LFSs (2001–2009).

variables available in the LFS (all of which are expected to be highly correlated with household expenditures).⁴ In the case of Morocco, these variables are found to exhibit a remarkable variation over time.

The direction of the time-trends is consistent with a process of modernization of the economy. What clearly stands out is: (1) a large reduction in household size; (2) a fair degree of urbanization, which is accompanied by an increase in access to electricity and sewage; (3) a steady reduction in employment in agriculture combined with a steady increase in employment in sectors such as finance and transport; and (4) an increase in higher education attainment.⁵ On the basis of this, one would expect to see a continued decline in poverty, although it is hard to say without having access to the predicted poverty data, whether the rate of poverty reduction has increased, decreased, or has largely remained unchanged.

4. Note that the model parameters can either be drawn from the estimated asymptotic distribution, or the parameters can be re-estimated after bootstrapping the HES data. The latter option is computationally more intensive but might provide more accurate estimates in case asymptotic results do not hold, i.e., if the sample size of the HES is not sufficiently large.

5. The LFS and HES also collect data on earnings, which would arguably be a strong correlate of household expenditures and incomes. While we have the earnings data for the HES, the High Commission for the Plan of Morocco has not released the LFS data on earnings on the grounds that it is of insufficient quality.

III. EMPIRICAL MODEL

We split the data into urban and rural and fit a model to each separately. The labor markets, sector decomposition, returns to education, living conditions, the availability and use of infrastructure, and the price of transport tend to be different between urban and rural areas, which may be expected to lead to different models. In what follows, it can be seen for example that employment in agriculture matters in rural (both statistically and economically highly significant) but not in urban Morocco. The reverse holds true for employment in the financial sector. Further geographic heterogeneity is captured by interacting selected independent variables with macro-region dummy variables, in addition to including macro-region fixed effects.

While we avoid fully automatic model-selection procedures, we do adopt a modeling strategy. First, we group the independent variables into subgroups: demographics (household size, head age, and marital status), education (of head as well as other household members), employment (labor force status and sector of employment for the head as well as other members), and dwelling unit characteristics. Each of these groups of independent variables is regressed on (log) per capita household expenditure (with macro-region fixed effects but without region interactions), which gives us a first idea of how the different types of variables rank as predictors of household expenditure. At this stage, the differences between urban and rural can already be seen. In addition to differences in regression coefficients, education and employment variables are found to be better predictors of expenditures for urban than for rural households. In both urban and rural areas, the dwelling unit variables rank as the strongest predictors. In urban areas this is followed by education, in rural areas the demographic variables rank as the second strongest predictor. In fact, in the rural regressions the education variables rank as the weakest predictor of household expenditures. Next, we start combining the groups of independent variables in order of their predictive strength, where after each iteration we drop variables that cease to be significant. Once variables from all groups are represented, we explore whether any improvements in the goodness-of-fit can be obtained by including interactions with macro-region dummies to allow for geographic heterogeneity. Throughout the procedure, we are also alert on potential multi-collinearities and counterintuitive regression coefficients.

We will work with both the 2001 and the 2007 HESs data. While we allow the models to be different, by building models that best fit the data for any given year, the models we identified are found to be closely related (in the sense that they include nearly the same set of explanatory variables).

As an experiment, a second set of models is obtained by adding a handful of variables on durable asset ownership and housing conditions that are available in the HES but not in the LFS. Because these are not LFS variables, we cannot consider these models for imputation into the LFS. We can, however, use them to impute consumption poverty into the HES and, subsequently, assess how

TABLE 1. Summary Statistics, Urban and Rural Models 2001 and 2007

	2001		2007	
Statistic	Urban	Rural	Urban	Rural
R2	0.59	0.43	0.58	0.42
R2 (assets)	0.64	0.48	0.63	0.48
# vars	52 (57)	45 (50)	58 (63)	51 (56)
Obs	7888	6355	4266	2796

Source: HESs (2001, 2007); the numbers in between brackets denote the number of independent variables in the models where additional “asset” variables have been included.

imputed poverty data compares to observed poverty data. The purpose of this exercise is to verify whether adding these variables to the model significantly improves the statistical precision of the imputed data. Other studies have found that durable asset ownership and housing conditions are particularly powerful predictors of poverty (see e.g., [Christiaensen et al. 2012](#)). If the same holds true for Morocco, then an argument can be made for adding these variables to future rounds of the LFS. Note that considering a variety of different models, with and without asset variables and each estimated to 2001 as well as 2007 data, also provides us with an assessment of model sensitivity.

Table 1 shows selected descriptive statistics for the four regression models (urban versus rural plus 2001 versus 2007). A number of characteristics are apparent: (a) the models provide reasonably good in-sample fits of the data judging by the adjusted R-squared, (b) the urban models fit the data better than the rural models, which is typical for these type of regression models, and (c) as expected, adding the five durable assets and housing variables significantly improves the in-sample fit. Whether this also translates into better out-of-sample fits is examined in the next section.

Urban Model

Table 2 presents the urban models for Morocco (for both years; with and without additional variables). The “U* x *” variables denote interactions between selected independent variables and the macro-region dummy variables. The subregions that are obtained when the macro-regions are disaggregated into urban and rural parts will be referred to as “domains” (see appendix 2 in the supplemental appendix, available at <http://wber.oxfordjournals.org/>).

The estimates of the model coefficients are found to be largely coherent: (a) per capita expenditure decreases with household size with a declining marginal effect; (b) the “returns to education” are all positive with higher returns for higher education levels (i.e., tertiary > secondary > primary education coefficient); (c) unemployment enters negatively, while waged employment, self-employment, and being an employer all enter the regression positively; (d) public sector employment too enters positively, while employment in the BTP sector (construction) is associated with lower standard of living, which is consistent

TABLE 2. Urban Model

Variable	Without additional assets		With additional assets	
	2001	2007	2001	2007
Domain U2	-0.233***	-0.132	-0.178***	-0.122
Domain U3	-0.041	-0.160**	-0.031	-0.070
Domain U4	-0.103**	-0.191**	-0.043	-0.145*
Domain U5	0.000	-0.108	0.127***	-0.026
Household size	-0.090***	-0.113***	-0.140***	-0.169***
Household size ²	0.002***	0.005***	0.005***	0.008***
Log age (head)	0.082***	0.108***	0.070***	0.082***
Married (head)	0.096***	0.140***	0.070***	0.114***
Primary (head)	0.101***	0.071***	0.069***	0.037*
Secondary (head)	0.231***	0.187***	0.149***	0.114***
Tertiary (head)	0.489***	0.439***	0.382***	0.352***
Employed (head)	-0.055	-0.191*	-0.043	-0.193**
Unemployed (head)	-0.196***	-0.308***	-0.144***	-0.300***
Self-employed (head)	0.098***	0.191***	0.159***	0.265***
Employer (head)	0.296***	0.403***	0.226***	0.353***
Employer (count)	0.385**	0.749***	0.394***	0.721***
Public (count)	0.315***	0.300***	0.249***	0.240***
BTP (head)	-0.097***	-0.127***	-0.077***	-0.091***
Finance (head)			0.135*	0.160**
Finance (count)			0.580***	0.447**
Waged (count)	0.222***	0.253***	0.264***	0.339***
Primary 1 (count)	0.122***	0.145***	0.039	0.106***
Primary 2 (count)	0.420***	0.369***	0.266***	0.234***
Secondary (count)	0.639***	0.485***	0.412***	0.355***
Tertiary (count)	0.684***	0.795***	0.470***	0.636***
Rooms per cap	0.602***	0.679***	0.410***	0.488***
Rooms per cap ²	-0.059***	-0.071***	-0.034***	-0.047***
Electricity	0.183***	0.154***	0.084***	0.052
Sewage	0.065***	0.131***	0.057***	0.079*
Drinking water	0.145***	0.067	0.087***	0.031
Flush toilet			0.058*	0.074
Kitchen			0.075***	0.027
Douche			0.228***	0.225***
Tv			0.145***	0.112***
Parabole			0.227***	0.209***
U2 × unemp (hd)	-0.104	0.141	-0.052	0.182
U3 × unemp (hd)	-0.157	0.159	-0.090	0.144
U4 × unemp (hd)	0.086	0.298**	0.076	0.270*
U5 × unemp (hd)	-0.137	0.085	-0.084	0.111
U2 × waged (count)	-0.144*	-0.258**	-0.118	-0.223**
U3 × waged (count)	-0.057	-0.040	-0.087	-0.097
U4 × waged (count)	-0.068	-0.315***	-0.060	-0.273***
U5 × waged (count)	0.116	0.033	0.106	-0.008
U2 × public (hd)	-0.037	0.043	-0.066	-0.009
U3 × public (hd)	0.009	-0.015	0.031	-0.036
U4 × public (hd)	-0.135**	-0.054	-0.108**	-0.054
U5 × public (hd)	-0.149**	-0.161***	-0.094	-0.140**

(Continued)

TABLE 2. Continued

Variable	Without additional assets		With additional assets	
	2001	2007	2001	2007
U2 × drinkwater	0.035	−0.038	0.028	−0.066
U3 × drinkwater	0.047	0.242***	0.041	0.221***
U4 × drinkwater	−0.047	0.173**	−0.079**	0.111
U5 × drinkwater	0.142***	0.248***	0.087**	0.174**
U2 × sewage			0.035	0.082
U3 × sewage			−0.246***	−0.245***
U4 × sewage			0.014	0.022
U5 × sewage			−0.186**	−0.160*
U2 × roompc	0.241***	−0.027	0.146**	−0.063
U3 × roompc	0.077	0.136	0.086*	0.048
U4 × roompc	0.441***	0.159	0.278***	0.085
U5 × roompc	0.331***	0.080	0.136*	0.018
U2 × roompc ²	−0.057***	0.021	−0.021	0.032
U3 × roompc ²	−0.001	−0.041*	−0.009	−0.012
U4 × roompc ²	−0.143***	−0.065**	−0.087***	−0.036
U5 × roompc ²	−0.071**	0.004	0.010	0.013
Constant	8.095***	8.187***	8.261***	8.427***
adj-R ²	0.591	0.579	0.619	0.619
Obs	7888	4266	7888	4266

Source: HESs (2001, 2007).

with the BTP being a low-wage sector; (e) employment in the financial sector is clearly beneficial but only in 2007, it was not yet significant in 2001; (f) size of the house as measured by the number of rooms per capita is strongly positively associated with total household expenditure, although the marginal effect declines for larger houses as expected; (g) similarly, households equipped with electricity, sewage, in-house clean drinking water (as well as the added durable assets and housing variables) are found to have higher total expenditure on average; (h) the significance of the interactions with the region dummy variables shows that the significance of the above mentioned variables is stronger for some areas than for others.

Rural Model

Table 3 presents the rural models for Morocco (for both years; with and without additional variables). Similarly to the urban model, the “R* x*” variables denote interactions between selected independent variables and the macro-region dummy variables. Note that the domains R1 to R4 refer to the rural parts of macro-regions 1 to 4. (A table with the definitions of the macro-regions can be found in appendix 2 of the supplemental appendix available at <http://wber.oxfordjournals.org/>).

Also here we find that the estimated model coefficients are largely coherent. For the variables that are shared by the rural and urban models, the signs of the

TABLE 3. Rural Model

	Without additional assets		With additional assets	
	2001	2007	2001	2007
Domain R2	0.269***	0.010	0.319***	0.105
Domain R3	0.060	-0.143	0.127*	-0.073
Domain R4	0.224***	-0.086	0.239***	0.008
Household size	-0.092***	-0.162***	-0.115***	-0.197***
Household size ²	0.003***	0.007***	0.004***	0.008***
Married (head)	0.109***	0.181***	0.088***	0.147***
Primary (head)	0.066***	0.055**	0.054***	0.029
Secondary (head)	0.147**	0.236***	0.079	0.219***
Tertiary (head)	0.449***	0.271**	0.405***	0.133
Unemployed (count)	-0.420***	0.422*	-0.389**	0.468**
Self-employed (count)	0.122***	-0.137*	0.156***	-0.112
Employer (count)	1.841***	1.519***	1.614***	1.329***
Agriculture (count)	0.170***	0.212***	0.182***	0.253***
Transport (count)	0.704***	0.829***	0.550***	0.587***
Commerce (count)	0.604***	0.556***	0.449***	0.483***
Public (head)	0.354***	0.283***	0.276***	0.184**
Waged (head)	-0.115***	-0.166***	-0.104***	-0.146***
Waged (count)	0.365***	0.393***	0.301***	0.418***
Primary 1 (count)	0.170***	0.167***	0.092**	0.094**
Primary 2 (count)	0.542***	0.480***	0.371***	0.322***
Secondary (count)	0.849***	0.599***	0.695***	0.348***
Tertiary (count)	1.145***	0.723***	0.950***	0.631***
Rooms per cap	0.533***	0.286***	0.418***	0.162**
Rooms per cap ²	-0.070***	-0.016	-0.051***	-0.002
Electricity	0.206***	0.236***	0.087***	0.054**
Sewage	0.041	0.493	-0.006	0.418
Drinking water	0.063***	0.079	0.046**	0.063
Flush toilet			0.129***	0.096***
Kitchen			0.029**	0.047*
Douche			0.140***	0.230***
Tv			0.163***	0.153***
Parabole			0.156***	0.227***
R2 × hhld size	-0.024***	-0.006	-0.024***	-0.004
R3 × hhld size	0.004	0.019*	0.002	0.026**
R4 × hhld size	-0.006	0.012	-0.001	0.013
R2 × unemp (count)	-0.055	-0.490	-0.122	-0.547*
R3 × unemp (count)	0.487**	-0.848***	0.493**	-0.946***
R4 × unemp (count)	0.075	-1.146***	0.026	-1.066***
R2 × waged (count)	-0.279***	-0.173	-0.245**	-0.199
R3 × waged (count)	-0.018	-0.369***	0.055	-0.411***
R4 × waged (count)	-0.501***	-0.392***	-0.435***	-0.451***
R2 × public (hd)	-0.181	-0.020	-0.123	-0.021
R3 × public (hd)	-0.159	-0.156	-0.123	-0.045
R4 × public (hd)	-0.098	-0.384***	-0.070	-0.283**
R2 × drinkwater			-0.044	-0.070
R3 × drinkwater			0.275***	0.233**
R4 × drinkwater			0.091	0.041

(Continued)

TABLE 3. Continued

	Without additional assets		With additional assets	
	2001	2007	2001	2007
R2 × sewage			−0.588	−0.515
R3 × sewage			−0.504	−0.507
R4 × sewage			−0.441	−0.431
R2 × roompc	−0.205**	0.048	−0.224***	0.010
R3 × roompc	0.224***	0.358***	0.124	0.280**
R4 × roompc	0.171**	0.370***	0.168**	0.306***
R2 × roompc ²	0.041*	−0.009	0.048**	−0.003
R3 × roompc ²	−0.031	−0.087**	−0.001	−0.064*
R4 × roompc ²	−0.028*	−0.065**	−0.028*	−0.051*
Constant	8.168***	8.737***	8.199***	8.786***
adj-R2	0.429	0.404	0.478	0.469
Obs	6355	2796	6355	2796

Source: HESs (2001, 2007).

coefficients generally match. Let us highlight some aspects that differentiate the rural model from the urban model: (a) employment in agriculture, transport, and commerce all enter the regression positively (sectors that are found to be less significant in urban Morocco); (b) returns to education are lower in rural compared to urban Morocco, as expected.

IV. VALIDATION TESTS

Before imputing expenditure poverty into all available years of the LFS, this section considers two different tests of the proposed methodology. Both tests use only 2001 and 2007 data, so that imputation-based estimates can be compared to official estimates based on observed data. The first test is conducted within HESs samples (so no LFS data is used here), while the second test considers cross-survey imputation using both the HESs and the LFSs. All imputation-based estimates of poverty, including the standard errors, are obtained using Stata's multiple imputation package (i.e., the “mi” package), where the number of imputations is set to 100.

In the first test, we conduct cross-imputations using only the HESs, by estimating the expenditure model in 2001 and then imputing expenditure poverty in 2007 and vice versa (table 4). This means that we do not have to worry about comparability between HESs and LFSs. It also allows us to test whether the additional durable assets and housing variables (that are not available in the LFS) yield better out-of-sample predictions.

The official poverty estimates are listed in the first column of table 4, which shows that poverty has almost halved over the six-year period, from 15.3 to 8.9 percent. Our imputation-based estimates of poverty are able to capture this trend remarkably well. What is equally encouraging is that the estimates

TABLE 4. Validation Tests Results, within HESs

Year	Official poverty estimates	No assets		With assets	
		2001 Model	2007 Model	2001 Model	2007 Model
2001	15.3 (0.54)	15.3 (0.56)	16.2 (0.72)	15.3 (0.56)	17.4 (0.77)
2007	8.9 (0.61)	9.6 (0.75)	8.9 (0.63)	8.8 (0.70)	8.9 (0.63)

Source: HESs (2001, 2007). Imputation-based estimates and standard errors obtained with Stata's MI.

TABLE 5. Validation Tests Results, Cross-Survey (HES-to-LFS)

Year	Official poverty estimates (%)	2001 Model	2007 Model
2001	15.3 (0.54)	15.3 (0.47)	15.3 (0.68)
2007	8.9 (0.61)	8.5 (0.30)	8.8 (0.37)

Source: HESs (2001, 2007). Imputation-based estimates and standard errors obtained with Stata's MI. Standard errors in parentheses. Note that the standard errors (SEs) associated with the imputation-based estimators are in some cases smaller than the SEs for the survey direct estimates. See remark 2 in section 2 for a discussion of (and intuition for) this observation.

obtained with the 2001 and 2007 models are very close, despite the six-years gap, which suggests that the assumption of a time-invariant model is not an unreasonable assumption in the case of Morocco. Finally, what this first test also shows is that the extended model (with assets) does not yield an obvious improvement in poverty estimates. Adding the asset variables improves model 2001's estimate of the 2007 poverty rate but not model 2007's estimate of the 2001 poverty rate, where we lose some precision. No changes are observed for the other estimates.

With the second test we impute into the LFS but still only for the HES years, so that also in this case we are able to compare our estimates to the official poverty estimates (table 5). Note that we are estimating annual poverty rates as this is the level at which the HES is representative, which means that we will be pooling the four quarters of the LFSs. Also, the imputation-based estimates match the "true" poverty rates remarkably well. It is particularly striking how accurately we are able to estimate poverty in 2007 based on a model estimated using 2001 expenditure data, and vice versa. Note also that the accuracy of the out-of-sample predictions gives us little reason to relax the model assumptions by adding more flexibility to the models.

Finally, the standard errors (SEs) associated with the imputation-based estimators are in some cases smaller than the SEs for the survey direct estimates. See remark 2 in section 2 for a discussion of (and intuition for) this observation.

V. POVERTY ESTIMATIONS 2001–2009

This section presents our main findings. We use both the 2001 and 2007 household expenditure models to estimate annual as well as quarterly poverty rates for the period 2001 to 2009 by imputing household expenditure in all rounds of the LFSs.

National and Urban-Rural Poverty Trends

Figure 2 shows the poverty trend at the national level. Note the consistency between the two different models. The two corresponding curves closely follow each other. There is also no discontinuity before and after 2007 when the new survey and computerized data collection systems were introduced. Interestingly, both imputation-based estimates find that poverty in Morocco has continued its decline beyond the 2007–2008 global financial crisis.

We also examined the intra-year fluctuations in the estimated poverty data. Specifically, we inspected the data for patterns that might be attributed to seasonality and made an attempt to link some of the larger quarterly fluctuations in poverty to similarly large fluctuations in independent variables. Results not reported here indicate that, while there appears to be some evidence of seasonality, the magnitude of these seasonal fluctuations is rather small. And while the larger quarterly fluctuations in poverty can indeed be traced back to fluctuations in selected independent variables, it is not clear whether these fluctuations capture genuine changes in underlying conditions or whether they should be attributed to sampling error. A more detailed examination might shed some more light on this question, which is beyond the scope of the present paper.⁶

Figure 3 disaggregates the poverty estimates into an urban and a rural trend. This reveals the divide between urban and rural standards of living, as expected, but also that this divide has been shrinking over time. Note also that the gap between the 2001 and 2007 curves has now widened somewhat due to the smaller samples but that the difference is still very small.

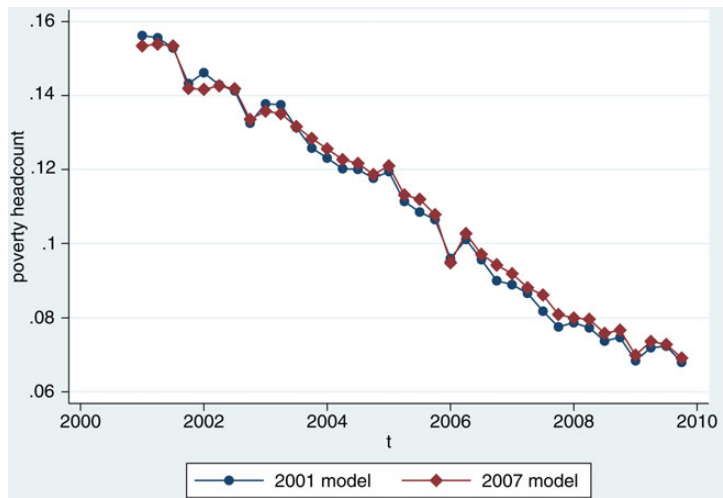
A Further Disaggregation of the Poverty Trends

Figure 4 further disaggregates the urban and rural poverty trends into macro-regions. These macro-regions consist of groupings of Morocco's original sixteen regions.⁷ For this level of disaggregation we are estimating poverty at an annual

6. A number of variables appear to exhibit a discontinuity between 2005 and 2006, most notably household size, higher education and selected employment variables. We conjecture that this may in part reflect the transition from the 1994 to the 2004 population census for the master sampling frame.

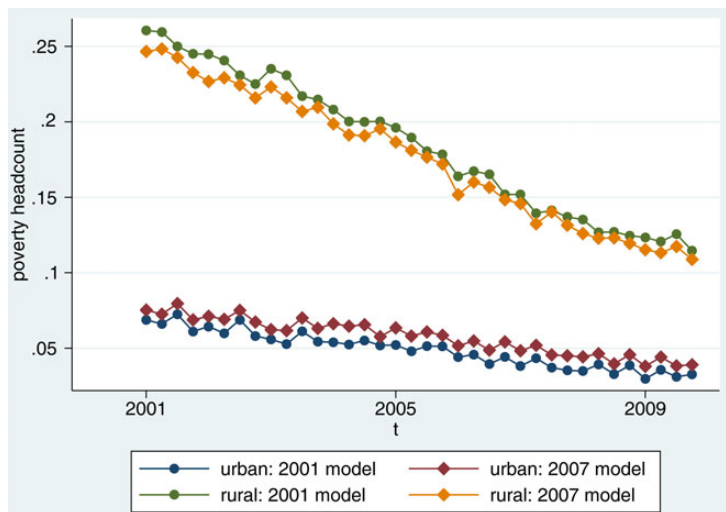
7. It may reasonably be expected that local food prices continue to display substantial seasonality (see e.g., Kaminski et al. 2014). Accounting for this variation in food prices could help explain some of the seasonal variations in household consumption. Unfortunately, sufficiently disaggregated data on food prices is rarely available (see e.g., Gibson et al. 2015). Also in our application to Morocco, food prices are excluded due to lack of data. It is conceivable that the imputation-based estimates would be more successful in identifying intra-year fluctuations in poverty if local food prices were included as part of the prediction model.

FIGURE 2. Quarterly Poverty Estimates 2001–2009



Source: LFSs (2001–2009) and HESs 2001 and 2007.

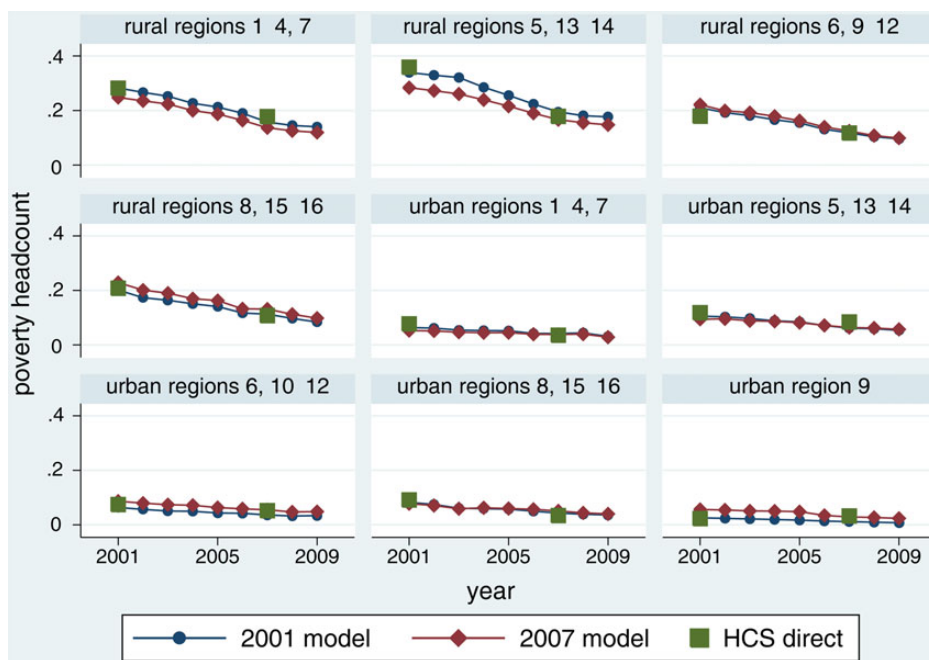
FIGURE 3. Quarterly Poverty Estimates 2001–2009, Urban and Rural Areas



Source: LFSs (2001–2009) and HESs 2001 and 2007.

frequency. Note that this also allows us to include sample direct estimates for the years 2001 and 2007 that will serve as a benchmark (this is not possible for the quarterly estimates since the HESs are not nationally representative at this frequency). What is apparent from these estimates is that practically all macro-regions of Morocco show a declining trend in poverty. Yet, these estimates also reveal a significant degree of heterogeneity in the rate of poverty reduction.

FIGURE 4. Quarterly Poverty Estimates 2001–2009 by Urban Regional Group



Source: LFSs (2001–2009) and HESs 2001 and 2007.

In addition, we disaggregated the poverty trends by selected household characteristics, namely by household size, age of the head of household, and sector of employment of the head. The corresponding figures are presented in the supplemental appendix (available at <http://wber.oxfordjournals.org/>). While these estimates also show a uniform decline in poverty, some subgroups can be seen to stand out with an above average reduction in poverty. These include the larger households, households with heads of average age or older, and those with heads employed in agriculture and BTP (construction). As is to be expected, these are also the subgroups who started off with a relatively high poverty rate in the base period. Importantly, all estimated trends in poverty reduction that are uncovered by the imputed data are found to be in agreement with the survey direct estimates for the years 2001 and 2007.⁸

Despite these encouraging results, some caution is warranted when considering the imputation-based estimates. While the impact of the recent global financial crisis may have been offset by positive domestic developments, it is conceivable that not all changes brought on by the crisis are well captured by the labor force surveys. For example, while households might have been able to hold

8. These groupings were determined by the High Commission for the Plan of Morocco based on population density; the names of the regions corresponding to the region codes ranging from 1 to 16 can be found in appendix 2.

on to employment, they might have re-adjusted consumption patterns by shifting to lower quality goods and/or reducing consumption altogether. Also, not all variables that may signal labor market distress, such as changes in contractual arrangements or changes in working time, are used in the consumption model so that these forms of labor market adjustments are not reflected in the imputation-based estimates. In sum, when estimating poverty rates beyond the last available consumption survey, it is important to be mindful of what is captured and what is not captured by the model and always validate estimates when a new consumption survey becomes available.

VI. CONCLUDING REMARKS

The paper has shown that Labor Force Surveys (LFSs) can be used effectively to estimate poverty when Household Expenditure Surveys (HESs) are not available. We have used cross-survey imputation methods to estimate poverty rates for Morocco at an annual and quarterly frequency for the period 2001–2009, when the most recent official estimate of poverty is for the year 2007. While our application has focused on the head-count poverty rate, the exact same approach can be applied to estimate other welfare measures including the poverty gap, poverty severity, and the Gini inequality index. All that is required is that the welfare measure can be expressed as a function of household expenditure (or income).

The empirical results are encouraging. We constructed first an expenditure model based on the 2001 HES data and a second model based on the 2007 HES data. We then used the 2001 model to estimate poverty with the 2001 and 2007 HES data and, vice versa, used the 2007 model to estimate poverty for both the 2001 and 2007 years. The imputation-based poverty estimates are found to be close to the official poverty estimates, regardless of which model is used. Next, we used the 2001 and 2007 consumption models to estimate poverty with the LFSs for the period 2001–2009; here we also obtained almost overlapping quarterly poverty trends, even when we disaggregated by urban and rural areas (and other subgroups).

This exercise provides several new insights for the study of poverty in Morocco. The imputation-based estimate of the poverty trend for this entire period shows that Morocco has been able to withstand global shocks (arguably helped by positive domestic changes), an observation that could not have been verified using the 2001 and 2007 HESs alone. Disaggregating the poverty trend into an urban and a rural series confirms a significant convergence in poverty for Morocco: the urban-rural gap declined from about 18 percentage points in 2001 to 7 percentage points by the end of 2009. When we disaggregate the poverty estimates further by sub-national regions (and by selected household characteristics) we find evidence of significant heterogeneity in the poverty reduction rates.

Morocco is not the only country that can benefit from this approach. Some of the largest low- and middle-income countries in the world could make use of HES-LFS cross-survey imputation. This applies to all regions of the world. For

example, the Malaysia Department of Statistics implements LFSs annually but HESs only every five years; the Philippine Statistics Authority carries out LFSs biannually while HESs are administered every three years, and in China, according to the National Bureau of Statistics, LFSs are conducted annually while HESs are conducted every three years. Statistics South-Africa administers a quarterly Labor Force Survey while HESs are administered every five years and the Chilean National Institute of Statistics administers LFSs quarterly but HESs only every ten years.⁹

There are a number of outstanding empirical questions that could be explored with further work. For example, what time-gap between estimation and application of the model can be handled satisfactorily? As the LFS moves further away from the last administered HES, the assumption that the model has not changed over time becomes increasingly optimistic. Can other moments of the income distribution, such as income inequality, be estimated with comparable accuracy? Here, fitting expenditures for higher-income households becomes more important. Are upward and downward trends in welfare captured equally well? Improvements in standards of living are possibly more easily captured than deteriorations. One is more likely to observe upgrades in education, dwellings, and assets than the reverse. While it is certainly possible for these variables to show a decline on average, change of this type may come with a delay. Changes in labor market conditions, such as an increase in the unemployment rate, are arguably first to respond in bad times. Whether this introduces an asymmetry in our ability to capture upwards and downward movements in poverty is an empirical question.

Can we generally say something about the conditions under which the approach will do well and not so well? Success of the approach largely depends on the extent to which the forces responsible for changes in welfare are reflected in the variables that we observe in the LFS. If a decline in welfare is largely due to a reduction in aid flows and social assistance, and this data is not represented in the LFS, then chances are that we will be under-estimating the change in welfare when relying on imputed data. Similarly, success of the approach may be compromised if certain events have introduced a shift in the underlying model that cannot be identified with the data at hand (see e.g., [Christiaensen et al. 2012](#)). Another factor that cannot be discarded of course is data quality. It should be noted that Morocco is not a typical developing country; it has a solid statistical system in place, which produces high quality data. It is our hope that similarly encouraging results can be obtained in poorer countries with poorer data, where imputation-based methods are arguably needed the most.

Finally, we can think of a number of extensions and applications. One option worth considering is to expand the set of independent variables by appealing to

9. The University Alberto Hurtado in Chile implements every three years a socio-economic survey on behalf of the Ministry of Planning as part of the regional SEDLAC project. This survey is also used for the estimation of poverty.

global geo-referenced databases; think of variables that measure night-time-light, rainfall, vegetation, and land use. Depending on the country, these data may improve our ability to capture short-term fluctuations in welfare. Another possibility is to use the estimated poverty (and inequality) data as independent variables in regression analysis, which may help identify the drivers of changes in poverty. This application is also explored in [Elbers et al. \(2005\)](#). The key difference between their approach and ours is that in our case the welfare variables are imputed into a second survey (the LFS) instead of a population census, which introduces sampling error as an additional source of error (in addition to model error). Provided that this dual error structure is accounted for, the results from [Elbers et al. \(2005\)](#) should directly carry over. Another possible application is poverty forecasting. Using the most recent LFSs, Morocco today can count on twelve years of poverty (and inequality) estimates for a cross-section of subnational regions. It is worth exploring whether this panel can be exploited to produce poverty forecasts for the future by fitting dynamic panel data models to the newly obtained data.

REFERENCES

- Breusch, T., and A. Pagan. 1979. "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica* 47 (5): 1287–94.
- Christiaensen, L., P. Lanjouw, J. Luoto, and D. Stifel. 2012. "Small Area Estimation-Based Prediction Methods to Track Poverty: Validation and Applications," *Journal of Economic Inequality* 10 (2): 267–97.
- Dang, H. A., P. Lanjouw, J. Luoto, and D. McKenzie. 2014. "Using Repeated Cross-Sections to Explore Movements in and out of Poverty," *Journal of Development Economics* 107: 112–28.
- Deaton, A. 2008. "Price Trends in India and their Implications for Measuring Poverty," *Economic and Political Weekly (Mumbai)*, February 9: 43–49.
- Deaton, A., and V. Kozel. 2005. "Data and Dogma: The Great Indian Poverty Debate," *World Bank Research Observer* 20 (2): 177–99.
- Doudich, M., and A. Ezzrari. 2008. "Dynamique de la pauvreté au Maroc 1985–2007," *Les Cahiers du plan*, No. 26, Haut Commissariat au Plan, Morocco.
- Elbers, C., J. O. Lanjouw, and P. Lanjouw. 2002. "Micro-level Estimation of Welfare." World Bank Policy Research Working Paper, Washington, DC, No. 2911.
- . 2003. "Micro-level Estimation of Poverty and Inequality," *Econometrica* 71 (1): 355–64.
- . 2005. "Imputed Welfare Estimates in Regression Analysis," *Journal of Economic Geography* 5 (1): 101–18.
- Elbers, C., and R. van der Weide. 2014. "Estimation of Normal Mixtures in a Nested Error Model with an Application to Small Area Estimation of Poverty and Inequality," World Bank Policy Research Working Paper, No. 6962.
- Filmer, D., and L. Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data – or Tears: An Application to Educational Enrollments in States of India," *Demography* 38 (1): 115–32.
- Fujii, T., and R. van der Weide. 2013. "Cost-effective Estimation of the Population Mean Using Prediction Estimators," World Bank Policy Research Working Paper, No. 6509.
- . 2014. "Can Financial Costs be Reduced by Predicting Rather than Collecting (Part of the) Data?" mimeo, The World Bank, Washington, DC.
- Gibson, J., T. Le, and B. Kim. 2015. "Prices, Engel Curves and Time-Space Deflation: Impacts on Poverty and Inequality in Vietnam," mimeo.

- Grosse, M., S. Klasen, and J. Spatz. 2009. "Matching Household Surveys with DHS Data to Create Nationally Representative Time Series of Poverty: An Application to Bolivia," No 21, Courant Research Centre: Poverty, Equity and Growth—Discussion Papers, Courant Research Centre PEG.
- High Commission for the Plan. 2000–2009. National Accounts, Haut Commissariat au Plan, Maroc, http://www.hcp.ma/Comptes-Nationaux-Croissance-Base-1998_a166.html.
- Hsiao, C. 1975. "Some Estimation Methods for a Random Coefficient Model," *Econometrica* 43 (2): 305–25.
- Kaminski, J., L. Christiaensen, and C. L. Gilbert. 2014. "The End of Seasonality? New Insights from Sub-Saharan Africa," World Bank Policy Research Working Paper, No. 6907.
- Kijima, Y., and P. Lanjouw. 2003. "Poverty in India during the 1990s: A Regional Perspective," World Bank Policy Research Working Paper, No. 3141.
- Mathiassen, A. 2013. "Testing Prediction Performance of Poverty Models: Empirical Evidence from Uganda," *Review of Income and Wealth* 59 (1): 91–112.
- Matloff, N. S. 1981. "Use of Regression Functions for Improved Estimation of Means," *Biometrika*, 68: 685–89.
- Rao, J. N. K. 2003. "Small Area Estimation," Wiley-Interscience, Hoboken, New Jersey, USA.
- Ridder, G., and R. Moffitt. 2007. "The Econometrics of Data Combination." In James J. Heckman, and Edward E. Leamer, ed., *Handbook of Econometrics*, Volume 6, Chapter 75, Elsevier.
- Rubin, D. B. 1978. "Multiple Imputations in Sample Surveys - A Phenomenological Bayesian Approach to Nonresponse," The Proceedings of the Survey Research Methods Section of the American Statistical Association, pp. 20–34. With discussion and reply.
- . 1987. "Multiple Imputations for Nonresponse Surveys," Wiley, New York, USA.
- Stifel, D., and L. Christiaensen. 2007. "Tracking Poverty over Time in the Absence of Comparable Consumption Data," *World Bank Economic Review* 21 (2): 317–41.
- Tarozzi, A. 2007. "Calculating Comparable Statistics from Incomparable Surveys, with an Application to Poverty in India," *Journal of Business and Economic Statistics*, 25 (3): 314–36.
- Verme, P., A. Gadiry-Barry, M. Taamouti, and J. Guennouni. forthcoming. "Labor Mobility, Economic Shocks and Jobless Growth. Evidence from Panel Data in Morocco." *Middle East Development Journal*.