TEXTO PARA DISCUSSÃO

No. 576

Estimation of Poverty Transition Matrices with Noisy Data

Nayoung Lee
Geert Ridder
John Strauss

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

Estimation of Poverty Transition Matrices with Noisy Data*

Nayoung Lee[†], Geert Ridder, and John Strauss[‡]

August 16, 2010

Abstract

This paper investigates potential measurement error biases in estimated poverty transition matrices. We compare transition matrices based on survey expenditure data to transition matrices based on measurement-error-free simulated expenditure. The simulation model uses estimates that correct for measurement error in expenditure. This dynamic model needs error-free initial conditions that can not be derived from these estimates. We provide bounds on the initial-conditions parameters, when these initial conditions are obtained by projection, and we also obtain initial conditions on the assumption that there is no time-constant measurement error. We find that for both estimates of the initial conditions measurement error in expenditure data magnifies economic mobility in and out of poverty. Roughly 44% of households initially in poverty at time $t-1$ are found to be out of poverty at time t using expenditure data from the Korean Labor and Income Panel Study (KLIPS). However, when we remove measurement error through a model-based simulation, only between 32 and 40% of households initially in poverty are found to be out of poverty.

JEL Classifications: C81, I32, O15
Keywords: Measurement error, Economic mobility, Transition matrix

*We would like to thank Loren Brandt, Angus Deaton, Hyungsik Roger Moon, participants of the Econometric Society winter meetings, January 2010, the Northeast Universities Development Consortium (NEUDC) meetings, October 2009, CUHK seminar, October 2009, the Pacific Development (PACDEV) meetings, March 2010, the Society of Labour Economists (SOLE) World Congress, June 2010 and the Econometric Society World Congress, August 2010, for constructive comments. Geert Ridder acknowledges support from the NSF under grants SES 0819612 and 0819638 and hospitality from the Department of Economics, PUC, Rio de Janeiro.

†Lee: Department of Economics, Chinese University of Hong Kong; Shatin, New Territories, Hong Kong; Email: nayoung.lee@cuhk.edu.hk. ‡Ridder, and Strauss: Department of Economics, University of Southern California; KAP 300, University Park Campus, Los Angeles, CA 90089-0253; Email: ridder@usc.edu, and jstrauss@usc.edu.

1. Introduction

Income and consumption mobility as measures of economic mobility have been receiving substantially more attention in the social sciences, with the increasing availability of panel data. Economic mobility is both related to changes in economic welfare for individuals and to changes in inequality for a society. One important method used to measure economic mobility for the poor is to study poverty dynamics using expenditure data and poverty transition matrices. In addition, studies of consumption dynamics using transition matrices that cross different quintiles of the expenditure distribution is a related method that is useful to assess economic mobility, not just of the poor.[1]

Most studies of income and poverty dynamics, however, have ignored potential measurement error biases in the transition matrices, although the presence of measurement error in both income and expenditure survey data has been widely acknowledged (eg. Deaton, 1997, Bound et al., 2001). This paper quantifies the direction and magnitude of the bias that measurement error in surveyed expenditure creates in poverty transition matrices and in more general expenditure transition matrices that measure economic mobility, using longitudinal data from Korea.

In this study, we use an economic model of consumption dynamics developed by Lee (2009) to construct measures of simulated expenditure that, under the assumptions of our model, do not contain measurement error.[2] From these simulated data we construct poverty and expenditure quintile transition matrices, which are then compared to matrices that use the measured expenditure data, which are measured with error. We allow for fairly general types of measurement error, in particular we allow both for time-varying measurement

[1] In principle, there is a distinction between consumption and expenditures. Consumption, which is the correct economic concept, is the flow value of services from the good over the period in question. This is quite hard to measure for semi-durables and durables, for which data generally contain expenditures, not flows of service. However, in practice, expenditures also partly include consumption, one example being for farmers, for food consumed out of own production, which is generally included in expenditures, but is also consumption. So expenditures is really a mixture of concepts. In this paper we will use the two terms interchangeably.

[2] A number of assumptions underlying this model we are able to test, see Lee (2009).

error, of the type that is most often thought as being problematic, and for time-invariant measurement error, which can be non-classical.

Since we estimate in first differences our model for error-free expenditures, which is autoregressive of order 1 in levels, we need two initial error-free values for expenditure to perform the simulation. The methodological contribution of this paper is that we develop two methods, based on non-nested assumptions, to resolve this problem. The first approach estimates the initial values by projection. As we show, with the available data we cannot point identify the variance of the error-free projection in the year before the simulation period. However, we develop sharp and informative bounds on this variance and simulate error-free expenditures at the lower and upper bound. The second approach makes the rather strong assumption that there is no time-constant measurement error in consumption and also assumes that the expenditure process is in stationary equilibrium. Under these assumptions we can point identify the distribution of the initial values. Fortunately, the results of both approaches not so different, so that the results are reasonably robust to deviations from either set of assumptions.

This study uses data from the Korean Labor and Income and Panel Study (KLIPS) from 1998 to 2006. We find that the transition matrices based on survey data are biased when expenditure data are reported with error. In particular, in these data measurement error magnifies economic mobility into and out of poverty. Roughly 44% of households initially in poverty at time $t-1$ are found to be out of poverty at time t using the KLIPS expenditure data. However, when measurement error is removed through our model-based simulation, 32 to 40% of households initially in poverty are found to be out of poverty. As another way to look at the data, over the four year period, $2002-2005$, the measured expenditure data show that 36% of the households are poor in at least one year, but only 6% are poor in all years, and 12% poor in three or four years. Hence two-thirds of the poverty is transitory using these estimates. On the other hand, using our simulated, measurement error-free data, some 34% of households are estimated to be poor at least one of these four years, but nearly half of those, 15%, are poor at least in three of the four years, while only 6.5% are poor in only one year and 12% poor in one or two years. Hence the poverty that exists in Korea seems more permanent when measurement error is accounted for in the poverty dynamics.

We want to be careful in our conclusions though. In other settings, notably rural settings in lower income countries, we expect a good deal more economic uncertainty than in Korea, which is largely an urban population, due to the high variance in rainfall and other factors that are critical in determining rural incomes and consumption.[3] As a result we might expect a greater degree of poverty mobility in such areas. Yet, we guess that in those cases too measurement error may bias mobility estimates based on reported data in the direction that we found. Given the modest data requirements of our approach, it would be advisable to use it to correct for measurement error bias, if panel data are available.

The remainder of the paper is organized as follows: Section 2 briefly reviews the literature on poverty dynamics; Section 3 develops the empirical methodology; Section 4 describes the data; Section 5 shows our findings; and Section 6 concludes.

2. Studies of Poverty Dynamics

Recent research on poverty in developing countries has focused on its variability in addition to static measures. There exist very different models that researchers have used to estimate consumption and/or poverty mobility.[4] An important method that is used to estimate consumption or income mobility assumes an AR(1) model for consumption or income

[3] See Rosenzweig (1994) for evidence comparing the mean coefficient of variation (CV) of profit incomes of individual farmers in the ICRISAT Village Level Studies to the mean CV of labor earnings of young white males in the US from the National Longitudinal Survey of Youth (NLSY). The CV for incomes in rural India is three times higher than in the US. The ICRISAT data are very well known for being high quality, it is most unlikely that differential measurement error can explain this result.

[4] Consumption, or expenditures, are generally preferred to income to study poverty, both because income is thought to generally have much larger measurement errors than expenditures, and because households tend to smooth their consumption relative to their income, so that consumption is a better measure of long-run resources. Indeed, Lee (2009) estimated both income and expenditure dynamics models on the same Korean data that we use in this paper and was able to quantify that the variance for time-varying measurement error for income was approximately five times larger than for expenditure.

with exogenous covariates.[5] The estimate of the autoregressive coefficient can be taken as an estimate of income or consumption mobility (see for instance, Gottschalk and Spolaore, 2002; McCullough and Baulch, 2000; Luttmer, 2002; Antman and McKenzie, 2007a, 2007b; Lee, 2009; Glewwe, forthcoming), with a higher value pointing at persistence and a lower (positive) value to more mobility.[6] These dynamic models are sometimes estimated in levels and sometimes in first differences (the latter being preferred in order to take out unobserved household fixed effects and time-invariant measurement error) and require instrumental variables (IV) to account for the endogeneity of lagged expenditure. The plausibility of these instruments, as is often the case, can be debated (see Antman and McKenzie, 2007a, for a discussion).

Another measure of poverty dynamics are poverty transition matrices between two years and/or counts of the number of years out of the total for which an individual household is in poverty. Many studies have used these two measures to investigate the degree to which poverty is persistent, and in these studies strong movements in and out of poverty have been one of regularities (Gaiha and Deolalikar, 1993; Jalan and Ravallion, 1998; Dercon, 1998; Baulch and Hoddinott, 2000; Baulch and McCulloch, 2000; Dercon and Krishnan, 2000). Some studies have tried to distinguish between chronic and transitory poverty because different types of poverty can have different determinants and have very different policy implications (Jalan and Ravallion, 2000; Duclos et al., 2010; McCulloch and Baulch, 2000). As a consequence, economists have agreed that distinct anti-poverty policies for each type of poverty can be more efficient to alleviate targeted poverty than a single set of policies that do not distinguish types of poverty. For example, long-term investments for the poor like education are likely to be effective in reducing chronic poverty, while enhancing households' ability to smooth consumption by providing a social safety net is likely to be more important to reduce transient poverty.

[5] See Fields (2006) for a useful survey of income mobility and how it can be measured.

[6] In addition, one can interpret the autoregressive coefficient as signifying conditional (on the X's) stationarity in income or consumption provided the coefficient is less than 1 and trending if the coefficient is greater than one, as in growth models.

Poverty dynamics is also related to the poverty trap literature which, if true, implies permanent rather than transitory poverty. Permanently low incomes will lead to less asset accumulation, which may lead to a poverty trap. Households with a low endowment of assets are unable to translate these into higher incomes because they pursue low-risk and low-return activities (Dercon, 1998; Zimmerman and Carter, 2003; Barrett and McPeak, 2004).[7] Accordingly, some studies recently have emphasized a nonlinear relationship between current and lagged income to identify potential poverty traps, but studies generally do not find evidence that supports this (Lokshin and Ravallion, 2004; Antman and McKenzie, 2007b).

Surprisingly few studies, however, have investigated the effect of measurement error on poverty rates and transition probabilities. The gold standard of studies that have considered measurement error are for income and earnings, not expenditure, and use administrative data (considered measured without error) that are matched to survey data at the household or individual level (see Lee, 2009, for a more detailed discussion). These studies are mainly based on US panel data (see Bound et al., 2001, for an older survey). Studies such as Bound and Krueger (1991) and Pischke (1995), for instance, have found that measurement error in labor market earnings in the US is positively autocorrelated and negatively correlated with "true" earnings. Even though these findings are for earnings, not expenditure, to the extent that they may be relevant for this study, we allow for measurement error that is correlated with true expenditure. Tests in Lee (2009) fail to reject the hypothesis that measurement error for both income and expenditure is uncorrelated over time in our data after removing household fixed effects and time-constant measurement error.

As discussed in Lee (2009), there are a small number of studies that compare measured recalled expenditure in a survey to expenditure as measured by daily diaries (eg. Ahmed et al., 2006). These studies find discrepancies between these measures. While in these studies, diaries are assumed to measure true expenditure, there is evidence that even diaries contain

[7] However, many other studies show that a household's business can start at very low asset levels and grow (McKenzie and Woodruff, 2006, for instance).

measurement errors (Browning, Crossley and Weber, 2003).

Studies that have examined income or consumption dynamics while accounting for measurement error without the benefit of administrative records, include McGarry (1995), Fields et al. (2003), Antman and McKenzie (2007a, 2007b), Lee (2009), Gibson and Glewwe (2005) and Glewwe (forthcoming). Of these, only McGarry and Gibson and Glewwe estimate poverty transition matrices as we do here. Fields et al. estimate a growth model of change in log income on lagged log income and other covariates. Although the dependent variable is a first-difference, only lagged log income is included. Relative to our model log income lagged twice is omitted and this variable is correlated with time-invariant measurement error and household-level time-invariant unobservables. Fields et al. instrument lagged income with household asset variables, household location variables and characteristics of the household head such as age, education and employment status in the initial period. These variables are surely correlated with time-invariant omitted variables such as individual ability.[8]

Glewwe (forthcoming) and Gibson and Glewwe (2005), have only two years of data available to estimate consumption dynamics for Vietnam. Not surprisingly, these studies have to trade off much stronger assumptions against the shorter panel. In particular, the dynamic model has to be estimated in levels instead of first differences and Glewwe (forthcoming) uses body mass index as his IV for lagged expenditure. These studies thus cannot address time-invariant measurement error, unlike this paper; other strong assumptions are made as well.[9]

Antman and McKenzie (2007a, 2007b) take a different approach to the estimation of the dynamic model for income. They use a synthetic cohort approach, using quarterly, urban Mexican employment data. They construct cohorts based on birth year and level of

[8] Assets may also be correlated with time-invariant measurement error of income, if for example, income is consistently underestimated in order to keep it secret from the government and the higher the income and assets, the more the incentive to underreport.

[9] Glewwe discusses the issues raised by his use of BMI as an IV. He has to assume no reverse causality from BMI to future consumption through health effects on income. Furthermore BMI is a stock measure, so that current values of BMI may be caused by past values of consumption, another possibility that Glewwe has to rule out. Either possibility, if true, would invalidate BMI as a valid instrument.

education of the head of the household.[10] By constructing cohort cells with at least 100 observations in each, they argue that time-varying, random measurement error is averaged out, leaving time-invariant error. On the other hand, they are also removing true random shocks and thus possible understating true mobility. If the time-invariant error is fixed within a birth year/education cohort, then the inclusion of cohort fixed effects will control for the time-constant measurement error. They estimate autoregressive coefficients for quarterly data which are around 0.8 when measurement error is fully accounted for (Antman and McKenzie, 2007a). When they use annual data, as we do, their estimate drops to 0.54. In Antman and McKenzie (2007b), they estimate nonlinear dynamic models of income and consumption using the same data, using a cubic term in lagged income or expenditure.[11] There goal is to test for poverty traps and they do not find much evidence for the existence of such traps.

McGarry (1995) in a study very close in spirit to this paper, simulates income and uses these simulations study the effect of measurement error on the change in poverty status and poverty transitions for widows in the US. The model that she uses in her her simulations is based on the autocorrelated individual component model (or variance component model) advocated by Lillard and Willis (1978). This is a model without an AR(1) term in income, but with an individual random effect, and a second random effect that is interacted with a time trend. In addition, Lillard and Willis (and McGarry) introduce a random, individual AR(1) disturbance, plus an idiosyncratic error. The idiosyncratic error is the sum of random shocks to income and time-varying measurement error that, as in this paper, is assumed

[10] They use only households in their first year of the survey in order to reduce potential problems due to the make-up of cohorts changing over time because of migration.

[11] Measurement error makes these nonlinear studies much more difficult because the measurement error term will have interactions with the lagged income or consumption when the quadratic and cubic terms are expanded. When first differences or fixed effects are taken, these interactive terms will not be removed, even if the measurement error is time-invariant. This makes estimates such as Lokshin and Ravallion (2004) inconsistent. Pseudo-panels as in Antman and McKenzie (2007b) may help, as they argue it does, for random, time-varying error because of the averaging of such error over cohort members, but for non-classical errors or errors that are correlated with covariates related to the cohort definition, like age and education, it will not help. Cohort fixed effects will not help either because they will not remove interactions between time-invariant cohort-specific measurement error and time-varying lagged cohort consumption.

to be uncorrelated over time. McGarry cannot distinguish between the random shocks to true income and time-varying measurement error, as we do in this paper. She also does not allow for time-invariant measurement error. McGarry compares simulated incomes with and without the variance component that includes measurement error, and she concludes that the amount of permanent poverty of widows is underestimated if there is measurement error. In this paper, we simulate expenditure based on our model without measurement error, which is possible because can separately identify the true shocks and measurement error variances. We compare the error-free expenditure to reported expenditure data, and not a second simulation. The measured expenditure data include all types of measurement error.[12]

In sum, there are a relatively few studies that are serious in trying to correct for measurement error in estimating income or consumption dynamics. Apart from studies that use administrative data that can be matched to survey data, studies only correct for time-varying measurment error, not time-invariant error, and this under strong assumptions. Of these studies, only McGarry (1995) and Gibson and Glewwe (2005) attempt to construct poverty transition matrices based on error-free simulations.

3. Empirical Methodology

3.1. Transition matrix

Let C_{it}^{*} be the true *per capita* consumption (or *per capita* expenditure, pce) of household i in period t. We discretize consumption, so that the household has consumption level j in period t if $b_{j-1} \leq C_{it}^{*} < b_j$ with $b_0 = 0 < b_1 < \cdots < b_{m-1} < b_m = \infty$. The probability that

[12] Luttmer (2002) and Villanger (2003), in unpublished papers, investigate the effect of measurement error on poverty transition probabilities. Luttmer's paper is similar to McGarry's in spirit. Both construct their measurement error-free welfare measure (either income or expenditure) by subtracting simulated measurement error itself from survey data. However, in their studies, simulated measurement error is assumed to be random, classical error, thus independent of the error-free income or expenditure. Thus the difference between surveyed income or consumption and simulated random error includes non-classical error in addition to true consumption.

household i makes a transition from consumption level j in period $t-1$ to level k in period t is

$$p^*_{jkt} = \frac{\Pr(b_{j-1} \leq C^*_{i,t-1} < b_j, b_{k-1} \leq C^*_{it} < b_k)}{\Pr(b_{j-1} \leq C^*_{i,t-1} < b_j)} \tag{1}$$

The $m \times m$ matrix of transition probabilities is denoted by P^*_t and this matrix is the parameter of interest.

However, we do not observe the true consumption C^*_{it} but rather the mismeasured C_{it}. The relative measurement error is η_{it} so that

$$\ln C_{it} = \ln C^*_{it} + \eta_{it}. \tag{2}$$

We assume that the measurement error η_{it} can be decomposed into a time-invariant and a time-varying component[13]

$$\eta_{it} = e_i + v_{it}. \tag{3}$$

Measurement error in per capita consumption implies that in general the observed consumption transition matrix P_t differs from the true transition matrix P^*_t. The objective of this study is to estimate the true transition matrix from data on mismeasured per capita consumption.

We estimate the true transition matrix by simulation. We take the model and estimates from Lee (2009) and add an assumption on the distribution of the random shocks in that model. However, even with this additional assumption we cannot simulate the true transition matrix. The problem is that the model is an autoregressive model that does not specify initial conditions for true consumption. We follow two approaches to deal with this problem. In the first approach we project true initial consumption on covariates and use observed consumption to estimate the coefficients of the projection. For the simulation we also need the projection variance, i.e. the variance of the projection error for the true initial observations. This variance is not point identified, but we obtain bounds on the variance

[13] One of scenarios considered is that the time-constant measurement error is 0.

of this projection error and these bounds turn out to be informative. An advantage of this approach is that only weak assumptions on the measurement error processes and no assumptions on the stationarity of true consumption are needed. In the second approach we assume that there is no time-invariant measurement error and we also assume that the process for true consumption is in stationary equilibrium. Under these assumptions we can initialize the process without using projections to estimate the distribution of the initial observations. The two approaches are valid under non-nested assumptions and therefore we report the results of both to investigate the sensitivity of our estimates to our assumptions.

3.2. True Consumption

Lee (2009) specifies the following autoregressive model of consumption dynamics[14]

$$\ln C_{it}^* = \gamma \ln C_{i,t-1}^* + \beta' X_{it} + D_t + \alpha_i + \varepsilon_{it} \ , \qquad t \geq 1 \tag{4}$$

where X_{it} is a vector household demographic variables, D_t captures time-specific effects, i.e. a full set of year dummies, and α_i is a time-invariant unobserved household specific intercept. The corresponding model in observed household per capita consumption is after substituting equations (2) and (3) into equation (4),

$$\ln C_{it} = \gamma \ln C_{i,t-1} + \beta' X_{it} + D_t + \alpha_i + (1-\gamma) e_i + v_{it} - \gamma v_{i,t-1} + \varepsilon_{it}, \qquad t \geq 1. \tag{5}$$

The total composite error of this model is

$$\tau_{it} \equiv \alpha_i + (1-\gamma) e_i + v_{it} - \gamma v_{i,t-1} + \varepsilon_{it}. \tag{6}$$

The random shock in the true consumption equation ε_{it} is in the sequel referred to as the equation error. Both the equation error and the time-varying measurement error are assumed

[14] Her paper also discusses the procedure that was used to select this model.

to be serially uncorrelated.[15]

The estimation of the model (5) is complicated by the presence of unobserved household effects, including the time-invariant component of the measurement error, and the time-varying measurement error. Lee (2009) follows Arellano and Bond (1991) by first-differencing the model and using lagged consumption as instruments.

$$\Delta \ln C_t = \gamma \Delta \ln C_{t-1} + \beta' \Delta X_t + \Delta D_t + \Delta \tau_t, \qquad t \geq 2 \tag{7}$$

$$\Delta \tau_t \equiv \Delta v_t - \gamma \Delta v_{t-1} + \Delta \varepsilon_t, \qquad t \geq 2 \tag{8}$$

First differencing allows for arbitrary correlation between α_i and e_i and the independent variables. Because of the time-varying measurement error log consumption lagged two periods is not a valid instrument, but log consumption lagged three periods is valid. In addition she uses a lagged measure of income satisfaction as an external instrument. The assumptions made by Lee (2009) are

$$E[\varepsilon_{it} | \ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i, Z_{i0}, \ldots, Z_{i,t-1}, \alpha_i, e_i] = 0. \tag{9}$$

and

$$E[v_{it} | \ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i, Z_{i0}, \ldots, Z_{i,t-1}, \alpha_i, e_i] = 0, \tag{10}$$

with X_i, Z_i the vectors of observations on the time-varying independent variables and the external instruments respectively. This is a sequential exogeneity assumption on the lagged dependent variables and the external instruments and an assumption of strict exogeneity on the other explanatory variables, X, conditional on α_i and e_i. No assumptions on the conditional variances are needed for the consistent estimation of the regression parameters,

[15] Lee (2009) also considers an MA(1) specification for ε_{it}, but fails to reject the hypothesis of serial uncorrelatedness. Using only the external instrument, she also tests for serial correlation in the time-varying measurement error and fails to reject the hypothesis of no serial correlation.

i.e. under the assumptions made Lee estimates the parameters of the true consumption process (4) consistently.

For the simulation of the transition probability matrix we also need the variance of the equation error ε_{it}. Following Lee (2009) we assume that both the equation error and the time-varying measurement error are homoskedastic, i.e.

$$\mathrm{Var}(\varepsilon_{it}|\ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i) = \sigma_{\varepsilon}^2 \tag{11}$$

and

$$\mathrm{Var}(v_{it}|\ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i) = \sigma_{v}^2. \tag{12}$$

Moreover we assume that the errors are conditionally uncorrelated

$$E(\varepsilon_{it} v_{it}|\ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i) = 0. \tag{13}$$

Under these assumptions

$$E\left[(\Delta\tau_{it})^2\right] = 2\sigma_{\epsilon}^2 + 2(\gamma^2 + \gamma + 1)\sigma_v^2 \tag{14}$$

$$E[\Delta\tau_{it}\Delta\tau_{i,t-1}] = -\sigma_{\epsilon}^2 - (\gamma^2 + 2\gamma + 1)\sigma_v^2 \tag{15}$$

$$E[\Delta\tau_{it}\Delta\tau_{i,t-2}] = \gamma\sigma_v^2. \tag{16}$$

We can estimate the variance and covariances on the left hand side using the residuals of the estimated consumption equation, so that we can use these three moment conditions to estimate σ_{ϵ}^2 and σ_v^2 by minimum distance methods (see Lee (2009) for more details). [16]

With the estimated regression parameters and the equation error variance we can simulate the first difference equation (7) if we add the assumption that the equation error

[16] We could relax the homoskedasticity assumption by replacing estimates of unconditional by conditional (co)variances on the left hand side. This is not considered in this paper.

ε_{it} has a normal distribution, i.e. we simulate

$$\Delta \ln C_{it}^* = \hat{\gamma} \Delta \ln C_{i,t-1}^* + \hat{\beta}' \Delta X_{it} + \Delta D_t + \Delta \varepsilon_{it} \qquad t \geq 2 \tag{17}$$

with

$$\varepsilon_{it} \sim N(0, \hat{\sigma}_{\varepsilon}^2). \tag{18}$$

In the simulation we ignore the sampling variation in the parameter estimates.[17]

Because the transition probabilities are for the level of log consumption and not their changes we need to find appropriate initial values. We simulate the levels by

$$\ln C_{it}^* = \ln C_{i,t-1}^* + \Delta \ln C_{it}^*, \qquad t \geq 2. \tag{19}$$

Therefore the (joint) distribution of two initial observations must be known, e.g that of $\Delta \ln C_{i1}^*$ and $\ln C_{i1}^*$. To obtain these distributions of the initial values we consider two approaches: (i) projection, and (ii) no time-invariant measurement error plus stationarity.

Initial values by projection

We specify a linear relation between $\Delta \ln C_{i1}^*$ and $\ln C_{i1}^*$ and X_{i0}, X_{i1}. If we first-difference equation (4) and recursively substitute for $\Delta \ln C_{it-1}^*$we find that $\Delta \ln C_{i1}^*$ is a (linear) function of $X_{i1}, X_{i0}, X_{i,-1}, \ldots$. The same equation implies that $\ln C_{i1}^*$ also is a (linear) function of $X_{i1}, X_{i0}, X_{i,-1}, \ldots$ and, in addition, of α_i, the household effect that can be correlated with all X_{it}.[18] Therefore the linear relations

$$\Delta \ln C_{i1}^* = \delta_0 + \beta_0 X_{i0} + \beta_1 X_{i1} + \zeta_{i0} \tag{20}$$

[17] We normalize D_1 to be 0 and use the estimated value of ΔD_2 to derive D_2.

[18] It is also a function, of course, of lagged values of ϵ.

and

$$\ln C_{i1}^{*} = \delta_1 + \beta_2 X_{i0} + \beta_3 X_{i1} + \zeta_{i1}. \tag{21}$$

are linear projections of these relations on X_{i0}, X_{i1}. Here, ζ_{i0} and ζ_{i1} are the projection errors, with variances σ_0^2 and σ_1^2.[19][20] These coefficients of the projections and the variance of the projection error need to be estimated to simulate $\ln C_{i1}^{*}$ and $\Delta \ln C_{i1}^{*}$.

Substituting observed consumption $\ln C_{i1}$ and $\Delta \ln C_{i1}$, we have

$$\Delta \ln C_{i1} = \delta_0 + \beta_0 X_{i0} + \beta_1 X_{i1} + \Delta v_{i1} + \zeta_{i0} \tag{22}$$

and

$$\ln C_{i1} = \delta_1 + \beta_2 X_{i0} + \beta_3 X_{i1} + e_i + v_{i1} + \zeta_{i1}. \tag{23}$$

We obtain consistent estimates of $\delta_0, \delta_1, \beta_0, \beta_1, \beta_2, \beta_3$ if we assume $E[e_i | X_{i0}, X_{i1}] = 0$, i.e. the time-constant measurement error is mean independent of X_{i0}, X_{i1}. This is a strong assumption and if it fails the coefficients in (23) are biased but not those in (22).

We make two further assumptions on the projection errors that can be relaxed. We assume that the errors are homoskedastic and that they are normally distributed. In addition, because the initial observations are for period 1 we should allow for correlation between $\Delta\varepsilon_{i2}$ and ζ_{i0}, ζ_{i1} and these covariances must be estimated as well.

Define the errors in the estimation equations:

$$\begin{aligned} \tau_t &= \alpha + (1-\gamma)e + v_t - \gamma v_{t-1} + \varepsilon_t, \qquad t \geq 1 \\ \psi_0 &= \Delta v_1 + \zeta_0 \\ \psi_1 &= e + v_1 + \zeta_1 \end{aligned} \tag{24}$$

[19] Note that ζ_{i1} is a function of any part of the household fixed effect, α, that is uncorrelated with X_{i1} and X_{i0} as well as being a function of further lags in X and in ε.

[20] Using a projection estimator for initial conditions of dynamic panel data regression models was used by Bond and Windmeijer (2002), but see also Hsiao (1986).

The errors in the simulated equations are: $\Delta\varepsilon_t, t \geq 2, \zeta_0, \zeta_1$.

It is reasonable to assume

$$\mathbb{E}[\zeta_0 \Delta\varepsilon_t] = \mathbb{E}[\zeta_1 \Delta\varepsilon_t] = 0 \quad , t \geq 3$$

and that the projection errors, ζ_0, ζ_1, are uncorrelated with the time-varying measurement error, v. The following variances and covariances are needed for the simulation:

$$\sigma_\varepsilon^2,\ \sigma_0^2,\ \sigma_1^2,\ \sigma_{01},\ \mathbb{E}[\Delta\varepsilon_2\zeta_0],\ \mathbb{E}[\Delta\varepsilon_2\zeta_1]$$

The variance matrix of the random errors in the simulation is

$$\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{01} & \mathbb{E}[\zeta_0\Delta\varepsilon_2] & 0 & 0 & 0 \\ & \sigma_1^2 & \mathbb{E}[\zeta_1\Delta\varepsilon_2] & 0 & 0 & 0 \\ & & 2\sigma_\varepsilon^2 & -\sigma_\varepsilon^2 & 0 & 0 \\ & & & 2\sigma_\varepsilon^2 & -\sigma_\varepsilon^2 & 0 \\ & & & & 2\sigma_\varepsilon^2 & -\sigma_\varepsilon^2 \\ & & & & & 2\sigma_\varepsilon^2 \\ & & & & & \end{pmatrix} \tag{25}$$

Moment conditions and identification The paper has three moment conditions (14) – (16) that overidentify σ_v^2 and σ_ε^2. Moreover there are two moment conditions that exactly identify σ_0^2 and σ_{01}

$$\mathrm{Var}(\psi_0) = 2\sigma_v^2 + \sigma_0^2 \tag{26}$$

$$\mathrm{Cov}(\psi_0, \psi_1) = \sigma_v^2 + \sigma_{01} \tag{27}$$

There are also two moment conditions that exactly identify $\mathbb{E}[\zeta_0\Delta\varepsilon_2], \mathbb{E}[\zeta_1\Delta\varepsilon_2]$

$$\text{Cov}(\psi_0, \Delta\tau_2) = \mathbb{E}[\zeta_0\Delta\varepsilon_2] - (1+2\gamma)\sigma_v^2 \tag{28}$$

$$\text{Cov}(\psi_1, \Delta\tau_2) = \mathbb{E}[\zeta_1\Delta\varepsilon_2] - (1+\gamma)\sigma_v^2 \tag{29}$$

The variance of the projection error for the level equation ζ_1, however, is not point identified. To see this note that if we denote the variance of the observed projection error for the first period log consumption by

$$\omega = \text{Var}(\psi_1) = \text{Var}(e) + \sigma_v^2 + \text{Var}(\zeta_1) \tag{30}$$

then this can be solved for $\text{Var}(e) + \text{Var}(\zeta_1)$. The problem is that we cannot identify the variance of the time-constant measurement error. However we have information on this variance from the time-average of the errors of the log consumption equation. From this we can derive bounds on the variance σ_1^2. These are derived in three steps. First define

$$\kappa = \mathbb{E}\left[(\alpha + (1-\gamma)e)^2\right] = \sigma_\alpha^2 + (1-\gamma)^2\sigma_e^2 \tag{31}$$

From this definition

$$0 \leq \sigma_e^2 \leq \frac{\kappa}{(1-\gamma)^2} \tag{32}$$

Note that

$$\kappa = \mathbb{E}\left[\frac{1}{T-2}\sum_{t=3}^{T}(\tau_t - \mathbb{E}(\tau))(\tau_{t-2} - \mathbb{E}(\tau))\right] \tag{33}$$

This suggests the estimator

$$\hat{\kappa} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T-2}\sum_{t=3}^{T}(\hat{\tau}_{it} - \overline{\hat{\tau}}_i)(\tau_{t-2} - \overline{\hat{\tau}}_i)\right)$$

with $\overline{\hat{\tau}}_i$ the time average of the residuals $\hat{\tau}_{it}$.[21][22]

Finally, from the moment condition

$$\mathrm{Var}(\psi_1) = \sigma_e^2 + \sigma_v^2 + \sigma_1^2 \tag{34}$$

we obtain the bounds

$$\max\left\{\mathrm{Var}(\psi_1) - \sigma_v^2 - \frac{\kappa}{(1-\gamma)^2}, 0\right\} \leq \sigma_1^2 \leq \mathrm{Var}(\psi_1) - \sigma_v^2 \tag{35}$$

In addition to satisfying the bounds (35),the variance σ_1^2 must satisfy the additional restriction that the variance matrix Σ must be positive semi-definite. If at the lower bound that matrix is not positive semi-definite, i.e. if the smallest eigenvalue of that matrix is negative then the lower bound of σ_1^2 is that value that makes the smallest eigenvalue equal to 0. For that value the joint normal distribution of the errors in simulation is singular. The exact relation between these errors is given by the eigenvector corresponding to the 0 eigenvalue. If the eigenvector is c, then we have

$$c_1\zeta_0 + c_2\zeta_1 + c_3\Delta\varepsilon_2 + c_4\Delta\varepsilon_3 + c_5\Delta\varepsilon_4 + c_6\Delta\varepsilon_5 \equiv 0 \tag{36}$$

This allows us (if $c_2 \neq 0$) to express ζ_1 as a function of $\zeta_0, \Delta\varepsilon_2, \Delta\varepsilon_3, \Delta\varepsilon_4, \Delta\varepsilon_5$ so that only the latter variables must be simulated.

In the simulation we draw from the joint normal distribution of $\zeta_0, \zeta_1, \Delta\varepsilon_2, \ldots, \Delta\varepsilon_5$ given by (25). For the point identified parameters we substitute the point estimates and for the interval identified σ_1^2 we take the estimates of the upper and lower bound.

[21] Other moment conditions could be used in addition, but we do not do so here.

[22] A minor complication is that the data are not a balanced panel. The equation (33) can be easily adapted to the case that we have s observations for each household. The estimates of κ for the households that appear s times in the panel are averaged using the fraction in the sample as weights to obtain an overall estimate of κ.

Initial value by stationarity

We can avoid the use of projection for the initial observations if we assume that the true consumption process is in stationary equilibrium. Under that assumption the period 1 log consumption is equal to

$$\ln C_{i1}^{*} = \frac{\beta' X_{i1}}{1-\gamma L} + \frac{D_1}{1-\gamma L} + \frac{\alpha_i}{1-\gamma} + \phi_{i1} \tag{37}$$

with

$$\phi_{i1} = \frac{\varepsilon_{i1}}{1-\gamma L} \tag{38}$$

To use this equation in simulation we assume that

$$X_{it} = X_{i0} \ , \qquad t \leq 0 \tag{39}$$

$$D_t = 0 \ , \qquad t \leq 0 \tag{40}$$

These assumptions are in line with the assumption of stationarity of log consumption.

We also need an estimate of the household effect α_i. Because the household effect in the observed log consumption equation is $(1-\gamma)e_i + \alpha_i$ we could go two ways. We could assume that α_i is uncorrelated with the independent variables, i.e. we have a random effects model, and use the bounds on the variance of α_i implicitly derived in the previous section to do simulations for the two extreme cases. If we are reluctant to make the random effects assumption, then we can assume that the time-invariant measurement error is 0. Note that this corresponds to the case of maximum uncertainty in the distribution of the initial observation in (21), so that the variability of the true consumption process is largest minimizing the role of time-varying measurement error as an explanation of observed transitions. Under the assumption of no time-constant measurement error we have that α_i is the long run time average of the τ_{it} the error of the log consumption equation. Therefore an obvious estimator is

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T} \hat{\tau}_{it} \tag{41}$$

The equation error ϕ_{i1} in the initial condition is drawn from the normal distribution with the variance

$$Var(\phi_{i1}) = Var\left(\frac{\varepsilon_{i1}}{1-\gamma L}\right) = \frac{\sigma_\varepsilon^2}{1-\gamma^2}. \tag{42}$$

Because we have an estimate of the household effect α_i we can simulate the true log consumption in levels using equation (4), so that we only need the one initial condition.

3.3. Transition Matrices

Our goal is to estimate which fraction of the transitions observed in our sample is spurious, i.e. due to measurement error. If our model of the true error free consumption process is correct, then we could estimate the population true transition probabilities p_{jk}^* by simulating a large number of error free consumption paths for each household. We then could use these to estimate the true transition probabilities for each household and averaging of these household transition probabilities would give us the population p_{jk}^*. Note that strictly speaking this still would give us an estimate, since we would be averaging over a random sample and not over all members of the population. This is however not the relevant comparison. If we would observe consumption without error, then we would have a sample of the same size as our current sample and consisting of the same households, but with true instead of mismeasured consumption. The simulation of true consumption using our model will produce such a sample. We now compare the transition probabilities based on the sample with the mismeasured consumption $\hat{p}_{jk}$ to the transition probabilities estimated from the simulated true consumption paths $\hat{p}_{jk}^*$. The resultant estimated transition probabilities $\hat{p}_{jk}^*$ differ from those estimated using mismeasured consumption for two reasons. First, their average over (many) simulations, p_{jk}^* , differs from p_{jk} , which is the population transition probability for the mismeasured consumption paths. Second, $\hat{p}_{jk}^*$ and $\hat{p}_{jk}$ differ because of sampling variation due to the fact that for each household we observe only one mismeasured consumption path and one error free consumption path. Using our model it is easy simulate more than one sample of true consumption paths. Comparing the distribution of these $\hat{p}^*_{jk}$ to $\hat{p}_{jk}$ allows us assess the sampling variation (associated with sampling variation of the true consumption paths) of the fraction of observed transitions that is genuine, i.e. not due to measurement error. The average of this sampling distribution is equal to $p^*_{jk}/\hat{p}_{jk}$ and this ratio will have a sampling distribution because of sample variability in the mismeasured consumption paths.

In this paper we do not try to assess the sampling variation in $\hat{p}^*_{jk}/\hat{p}_{jk}$. Instead we draw a single sample (of size about 4000) from the distribution of true consumption paths (one for each household) and use $\hat{p}^*_{jk}/\hat{p}_{jk}$ as an estimate of the fraction observed transitions between consumption intervals based on mismeasured consumption that is genuine, i.e. would occur if we would have accurate measures of consumption. As explained above, drawing several samples from the distribution of true consumption would give an impression of the sampling variability of this fraction due to sampling variation in true consumption, but not due to sampling variability in mismeasured consumption. For now we ignore this issue and report the fraction genuine transitions for a single sample of mismeasured consumption paths and a single sample of true, i.e. without measurement error, consumption paths. Only the full sampling distribution of $\hat{p}^*_{jk}/\hat{p}_{jk}$ over samples of mismeasured and samples of true consumption gives an impression of the likely variation of the fraction of spurious transitions due to sampling. Even then we ignore sampling variability in the parameter estimates which is expected to have a small contribution to the variability of the estimate of the fraction genuine transitions.

4. Data

4.1. Variables and Sample Size

The data used for this study come from the Korean Labor and Income Panel Study (KLIPS), from 1998 to 2006. This study uses household expenditure for investigating poverty dynamics or economic mobility. As discussed earlier, researchers have agreed that expenditure (or consumption) is a better basis for measuring economic welfare and poverty in particular, and this extends to studies of mobility.

Household Expenditure Variables

KLIPS reports household expenditure in two ways: through an aggregate reporting of average monthly household expenditures over the past year, using a single question that covers all expenditure items (including autoconsumption of foods) and through the more common disaggregated method, which is based on details of household expenditure. However, even for the latter, KLIPS suffers from a lack of disaggregation of expenditure categories. Other panel surveys usually have more categories for expenditure data; some like the Living Standards Measurement Surveys (LSMS) may have up to one hundred categories, with much detail for foods, but KLIPS only has 11 (for the second wave) to 20 (for the ninth wave) categories. Household expenditures are measured in the survey by both methods only in the second, fourth and following waves. The survey asks for total household expenditures in the first and third waves, but excludes the disaggregated details. The average monthly household expenditures based on aggregate reporting is thus chosen for our main analysis so that we may have more years of data. All expenditures are converted into annual measures in year 2000, won. In KLIPS, there is little difference between the aggregate and disaggregate levels of expenditure.[23] Though there are fewer incentives to under-report survey consumption compared to income, substantial recall errors are assumed because of the lack of documented

[23] See Appendix Table 2. Only two households report zero consumption. Log of household per capita expenditure (pce) is taken and the two households who report zero consumption are excluded.

records for expenditures by households and because expenditure is asked about aggregated groups, which we suspect will lead to measurement error.

Other Variables

A set of household characteristics is controlled for the basic estimations and expenditure simulations. These include household size, the fraction of elderly people, educational level of head of household, sex of head of household, age of the head of household and its square, a locality indicator to show whether the respondent resides in Seoul, and a non-spouse indicator to show whether the household head has a spouse living in the household. The main summary statistics are reported in Table 1.

As explained in Section 3, The two-step GMM estimation of equation (7) uses in addition to three period and past lags of the dependent variable, lags of the household head's measured satisfaction regarding their household income as instruments. The income satisfaction variable comes from the response of each household head to the question 'how much are you satisfied with your household net income', and each individual responds according to degree of satisfaction on a 1 to 5 scale, with "1" being very satisfied and "5" being very dissatisfied. Lower scores, therefore, measure higher satisfaction. This question is asked at the individual level for each year except for the first wave.

Two period lagged income satisfaction is used as an external instrument in the GMM estimation. This is useful because when we use period three and further back lags of the dependent variable as our internal IVs for the time difference in log pce between periods $t-1$ and $t-2$ (as we must when we allow for time-varying measurement error) we sometimes encounter a weak instruments problem. External instruments help to avoid this in our case (see Lee, 2009). Equations (9) and (10) in Section 3 show the assumptions we must maintain in order to consistently use income satisfaction as an instrument. In the case of income satisfaction, conditional on the fixed effect and time-invariant measurement error, we must assume that past values of income satisfaction are uncorrelated with the current equation error (or expenditure shock), but that future values of income satisfaction may be correlated with current expenditure shocks.

Sample Size

The estimation of equation (7) requires at least four years' data because of potential time-varying measurement error. The availability of instruments used in this study is reported in Appendix Table A2. Total annual household expenditure, asked directly, is available from 1997 to 2005.[24] On the other hand, income satisfaction data are available only from 1999 to 2006. The overlapping periods for this analysis are only from 1999 to 2005. Because the Arellano and Bond method requires up to lagged period t-3 for instruments under the assumption of time-varying measurement error, the expenditure equation can only cover the years $2002, 2003, 2004$ and 2005. Consequently, 2000 and 2001 data are used to estimate initial conditions, and 2002 - 2005 data are used to construct measurement error-free expenditure for the later years, as explained in Section 3.

In this study, the simulation of expenditure without measurement error is carried out for the same households that are used for the estimation of the basic standard consumption dynamics model, equation (5).[25]

4.2. Consumption Classes

The goal in our study is to compare movements into and out of poverty, or more generally movements across quintiles of the distribution of real per capita expenditure, comparing surveyed and simulated expenditure data. Accordingly, this study starts constructing by 2×2 and 5×5 transition matrices with two and five consumption classes respectively. The former provides absolute poverty transition probabilities, while the latter allows us to look

[24]It is not available for 2006 because the question asks monthly average during the prior year.

[25]See Table 2 and Appendix Table 1. Using our simulation method, it is actually possible to simulate expenditures even if the household did not report expenditures for the current year (but did so for past years). Since we compare surveyed expenditures with our simulations, we drop the few household/years for which this is true.

at the expenditure transition probabilities across quintiles.

Studies for other countries generally use a poverty threshold, in particular an official poverty line, as a boundary of two consumption classes for 2×2 transition matrices. However, there is no official poverty line in Korea. Most researchers who study poverty in Korea use the Minimum Cost of Living (MCL) announced annually by the government as a poverty threshold (Park, 2001).[26] Like other researchers, this study uses the MCL as a boundary of consumption classes for 2×2 transition matrices. However, the MCL differs by household size. In this study, the MCL is calculated in proportion to the average household size in 2002, which is 3.43 for samples used in this study. The MCLs in 2002 are 73.7 and 92.8 for three and four person households (household, not *percapita*) respectively before taking logarithms, and so 87.0 for a 3.43 person household.[27] In 2000 won the 2002 MCL is 23.9 per capita. On the other hand, for 5×5 matrices, the boundaries are based on the quintiles of *percapita* expenditure (pce) in KLIPS, 2002. Either the poverty threshold or quintiles are fixed in real currency over time from 2002 to 2005, as are the *percapita* expenditures.

Any poverty threshold could have an important role for both static and dynamic analysis of poverty because where the threshold is set could affect results of both the poverty rate and transition (see Davidson and Duclos, 2000, for example, for a method to avoid this dependence for static poverty analyses). One advantage of examining transitions between quintiles, as well as between in and out of poverty, is that the thresholds differ and so we can examine to some degree the extent to which this matters.

5. Results

5.1. Simulated Consumption

This study requires estimates of several regressions, and these estimates can be classified

[26] The Ministry of Health and Welfare in Korea releases the MCL each year.
[27] This is about $8 per day per capita in 2002.

by their roles. An estimate of the autoregressive coefficient from the dynamic consumption model of equation (5) while informative, is not sufficient to construct the transition matrices. Because the model is a first or second order stochastic difference equation, we need initial conditions. These initial conditions are distributions of initial observations and these distributions can be estimated, or we can assume stationarity with additional assumptions, particularly that no time-invariant measurement error exists.

Table 3 shows the results of our base estimations using the initial conditions projection methodology. The GMM estimate of γ in equation (7), the autoregressive coefficient on log pce, is .375.[28] The variance of the time-varying measurement error (.032) is 60 percent as large as the variance of the equation error (.054), which suggests that time-varying measurement error exists in the reported KLIPS consumption data and has a substantial magnitude. However, the table shows a weak correlation between the projection error, ζ, in either the differenced or initial level equation (equations (22) and (23)), and the differenced equation error, $\Delta\varepsilon$, in equation (7) (see equations (28) and (29)). The projection errors, ζ_{i0} and ζ_{i1}, are also weakly correlated, with an estimated covariance of .027.

As described in Section 3, we obtain initial observations as draws from linear projections with projection errors, ζ_0 and ζ_1, that have a joint normal distribution with mean 0 and a variance-covariance matrix that has to be estimated. First, the variance of the projection error for the differenced initial condition, σ_0^2, is estimated as .116. The variance of the projection error for the level initial condition, σ_1^2, is not point identified. The bounds that we derive can be estimated, and estimates of the upper and lower bound of the variance together with the implied estimates of the unconditional variance of the time-constant measurement error and of the household effect, α, are shown in Table 3. The variance of the level projection error, σ_1^2, lies between .081 and .197. The lower bound is a constrained estimate that makes the 6x6 covariance matrix just positive semi-definite, as discussed in Section 3.[29] Our estimate of the eigenvector corresponding to the constrained lower bound is also reported.

[28] See Appendix Tables 3 and 4 for the other estimates of equations (7), (20) and (21)

[29] That is the lowest eigenvalue is 0; numerically it is $5.15E-12$.

Note that when σ_1^2 is largest, the time-invariant measurement error variance is 0, and σ_1^2 is smallest if the time-constant measurement error is as large as possible given the inequality constraint in equation (35). We can see that the estimated variance of the time-invariant measurement error in the lower bound case is over 3.5 times the variance of the time-varying measurement error.[30] This seems high for expenditures, but there has never been a quaniti-fication of this comparison, so we don't really know. The time-invariant measurement error may include a part that is non-classical, related to true expenditures, such as mean reversion, but a part, too, that may be related to time-invariant characteristics of the respondents, such as education and age (which obviously varies over time, but over a very short time period such as our four years, may be considered as time-invariant for practical purposes). The variance of the household fixed effect, α, is also large, especially compared to the variance of ϵ, the true equation error in equation (4), and to the variance of ζ_1. It is also larger for the upper bound case compared to the lower bound estimate. However, also notice that for the lower bound case the variance of α is 50 percent higher compared to the variance of the time-invariant measurement error. Given the many omitted variables in the consumption dynamics equation, one would expect that the variance of the fixed effect should be larger than of the time-invariant measurement error. Whether this order of magnitude makes sense the reader will have to decide. One advantage of both the upper bound projection and the stationary distribution estimates is that the fixed effect dominates the time-invariant measurement error, though obviously that is because the latter is assumed to be zero in both cases, which is also an assumption that many might not believe.

To impose stationarity, we need to estimate the the implied stationary equation error variance, that is equal to .063 (Table 3). We also need estimates of the household fixed effect, α, the two-step GMM estimates of equation (7) and the estimate of σ_ϵ^2 from the minimum distance estimation. The α_i are estimated from the residuals of the estimation of equation

[30] Remember that in the estimation of the dynamic consumption model, all time-invariant factors, including measurement error, are removed; the time-invariant measurement error variance is only used in the projection of intitial conditions for the 2001 level.

(5) (see equation (41)) for each household. The mean of α_i is 3.9 and its variance .264 (Table 3). The variance is somewhat higher than the variance of α_i for the upper bound projection error variance case-henceforth the upper bound projection case (Table 3). Recall that under our stationarity assumptions, we also have to assume no time-invariant measurement error.

Figure 1 plots the densities for our simulated initial expenditure and for the actual data as a difference between 2000 and 2001 and a level for $t = 2001$. Summary statistics are presented in Appendix Table A.5. Based on the upper bound projection case, the two distributions in Figure 1 seem close, and this implies that there is more spread in the surveyed consumption for 2001 when the lower bound on the projection error variance is used (the lower bound projection case). In this case our estimate of the time-invariant measurement error variance (which of course does not contribute to the simulated pce) is highest. The density for the stationarity simulation shows a slightly higher variance for 2001 than the surveyed data.

Simulated expenditures from 2002 to 2005 are sequentially constructed starting with simulated initial conditions and estimates of equation (7) and then using equations (17) and (19). Figure 2 plots the results and Appendix Table A.5 shows the summary statistics. Simulated expenditures using projection have virtually the same means as surveyed expenditure data, but unlike the expenditure data, the variances increase a small amount over time starting with 2002. One possible explanation may be that inequality is increasing, but is hidden by measurement error. It is also possible, on the other hand, that the simulated variance is approaching the stationary one.[31] Imposing stationarity, we can avoid having variances rise over time, and indeed we see that the densities largely overlap with the density for measured pce, but have slightly higher variances.

Random measurement error is generally believed to inflate the variances of consumption data. According to the literature, households are supposed to appear to have higher mobility with this type of measurement error. However, this study indicates that surveyed consump-

[31] It can be shown that the projection error variance cannot be equal to the stationary variance, so that the simulation will involve a transition to stationarity with changing variances.

tion has roughly equal variances compared to simulated consumption under the upper bound projection and stationarity assumptions, and a lower variance compared to the lower bound projection case. One possible explanation is that surveyed consumption in reality consists not only of time-varying measurement error but also of time-invariant measurement error. Bound and Krueger (1991) argue that households in the US at the top of the income distribution underreport their true income while households at the bottom over-report, and thus the distribution of surveyed income is compressed. The same logic may apply to these expenditure data, although there does not exist direct evidence for expenditure.

Appendix Table A.6 presents poverty headcount rates based on the minimum cost of living as a threshold, using expenditure data and simulated expenditure. For this study estimated headcount measures.are quite similar for the surveyed data and the simulations using the upper bound projection and stationary distribution estimates, while headcount rates using the lower bound projections are lower by approximately 5.3 percentage points.[32] Here, as Figure 2 shows, a comparison of the distribution of measured expenditure data with simulated data using projections of initial conditions for the lower bound case indicates a smaller variance of the simulated expenditure data. This results in a lower headcount estimate.

5.2. Mobility

The main focus of this study is on poverty dynamics, or the movement into and out of poverty. Table 4 shows the number and percentage of households experiencing poverty by years spent in poverty out of the four years possible from 2002 to 2005. Baulch and Hoddinott (2000) in their review note that the number of households characterized as 'sometimes poor' is larger than those that are 'always poor' in other studies, and the surveyed expenditure data in this study tells the same story. The simulated expenditure data show a somewhat different story.

[32] McGarry (1995) concludes that cross-sectional estimates of poverty rates of widows in the US are not biased by measurement error. As discussed in Section B, she only considers time-varying error.

Over the four year period, 2002-2005, the measured expenditure data show that 36% of the households are poor in at least one year, but only 6% are poor in all years, only 17 percent of the ever poor, while 24% (two-thirds of the ever poor) are poor in only one or two years. Hence most of the poverty is transitory using these estimates. Using our simulated, measurement error-free data based on the upper bound projection case, some 34% of households are estimated to be ever poor, but 25% of those, 9.2%, are poor each year and another 5.8% in three of the four years. Hence almost half the the ever poor (44%) are poor in all four years or in three of the four, while only 11.3% are poor in only one year and 19% poor in one or two years. This is a quite different balance, much more evenly split, than the surveyed expenditures show. Using the lower bound projection case, the results are different, a smaller fraction of the population is ever poor, only 26%, and of these 18% are poor in only one or two years (just over two-thirds). When we impose stationarity, a slightly higher fraction are ever in poverty, 35%, and of these 7.6% are in poverty all four years, while 6.1% are for three years. Thus, nearly 40% are in poverty 3 or 4 years, a little under the estimates from our upper bound projection case.

This study thus indicates that the fraction of 'always poor' households with these Korean data is downward biased by time-varying measurement error, whether stationarity is assumed or not for estimating initial conditions, the bias being greater when stationarity is not assumed. This result means that "chronic" poverty is understated when such measurement error is not corrected. However, allowing for both time-varying and time-invariant measurement error in projecting the initial conditions changes the picture. Apparently in our case, time-invariant measurement error offsets time-varying error.

Tables 5 and 6 present 2×2 and 5×5 transition matrices respectively, which are our main interest. The 2×2 poverty transition matrices show the persistence of poverty, by showing the probabilities of a household staying poor or moving from poor to non-poor. The 5×5 transition matrices consist of the probabilities of movements between quintile pce classes from t-1 to t.[33]

Most striking, for both the 2×2 and 5×5 matrices, the probabilities that households stay at the bottom are downward biased by measurement error; in the 2×2 case up to around 11.5 percentage points (56.5% versus 68%) for the simulation using the upper bound projection, a 20% difference.[34] The simulation using the lower bound projection case shows a higher mean movement out of poverty, 40%, but still marginally lower than estimates from the surveyed data. The estimate from the stationary distribution model is closer to the estimate from the upper bound projection case. So, again, it appears that allowing for time-invariant measurement error in our initial conditions projections offsets the impact of time-varying error. This makes sense in our context because as we saw from Figures 1 and 2, the tails of the level pce distribution are smallest for the lower bound projection model. This will result in those simulated observations that are below the poverty line, being closer to the poverty line compared to models that have fatter tails. This in turn will result in more transitions out of poverty under this model, which is what we see. For transiting into poverty, all the simulations show a very small fraction, between 5.8% and 6.8%. These estimates are lower than in the surveyed data, 7.8%, or between 13 and 25% lower.

In Table 6, when we examine the transition out of the bottom pce quintile the probabilities are quite close to those for transiting out of poverty, and the impact of measurement error is very similar as well, again with a downward bias for the probability of staying in the bottom quintile. Note that the probability of transiting into one of the top two quintiles from the bottom quintile is markedly lower for the simulations using the upper bound projection and stationary case estimates, although using the lower bound projection case the fractions are closer to the measured data.

[33] These probabilities are averaged across years. This study first calculates the cell size for each pair of adjacent years, then averages those across the years and finally calculates the row percentages. Year-by-year transition matrices are reported in Appendix Tables A7 and A8. Major differences across the pairs of years are not observed.

[34] McGarry (1995) also suggests a downward bias in the probabilities that households stay at the bottom, and the difference (74-63=11%) is quite similar. Again, McGarry compares simulated income with and without measurement error, not focusing on surveyed income. Though she also presents the probability for surveyed income, it is 74% which basically suggests no bias from measurement error at all.

The probability that households stay at the top of the distribution is somewhat biased downwards, by up to 7 percentage points, in the 5×5 matrix, when the upper bound projection estimate is used, but is roughly the same when we impose stationarity and is biased upwards when the lower bound projection estimate is used.. Being in the top quintile at time $t-1$, the odds of staying there are 61 - 76%, once measurement error is corrected using projections, and 71% when stationarity is imposed. The transition probabilities of the middle classes do not seem to be as much affected by measurement error. This is predicted by Baulch and Hoddinott (2000); measurement error biases in transition matrices are particularly problems for the poorest and richest categories, where negative and positive measurement errors cannot offset each other.

Finally, Table 7 summarizes the findings. In terms of mobility, especially for the 5×5 pce quintile transition matrix based on survey data, the probability of movements by two or more quintiles for households at the bottom is considerably larger, by at least 40 percent, using survey data than using simulated consumption with the upper bound projection or stationarity cases, although projection with the lower bound projection case provides a very different result. For all diagonals in the transition matrices the probability of moving is also overstated when measurement error is not corrected,[35] though the bias appears to be smaller than for those who start at the bottom.

6. Conclusion

We investigate whether transition matrices based on survey data are biased when expenditures are reported with errors. Measurement error-free expenditures are simulated based on parameters estimated from a basic model of consumption dynamics allowing for general types of measurement error. Initial conditions are estimated in two ways, by linear projection, and by imposing stationarity, and using the parameter estimates from the dynamic

[35] Again, except for the case of the loewr bound projection simulation and then only for the 5x5 transition matrix.

model of consumption. When we impose stationarity we need to assume that no time-invariant measurement error exists, because we are estimating the *percapita* expenditures in levels and so need an estimate of the household fixed effect. For this we can only identify the sum of the household fixed effect and the time-invariant measurement error, not the components separately. Using projections for the initial conditions, we can estimate pce in differences plus the level of initial conditions. The estimation of consumption dynamics in differences allows us to eliminate time-invariant measurement error and any household fixed effect in addition to dealing with time-varying measurement error. It also allows for a more general dynamic relationship between lagged and current consumption, not confining the source of the dynamics to be a serially correlated consumption shock. Consequently, our study presents different results from the literature.

This study has shown that with the KLIPS data from Korea, time-varying measurement error magnifies economic mobility into and out of poverty. The probability of remaining poor is downward biased by such measurement error. However, allowing for time-invariant measurement error in our projection of initial conditions of levels offsets the time-varying error in our case. The resulting differences from survey data in the probability of transiting out of poverty is at least 3.5 percentage points as a lower bound estimate, allowing for both types of measurement error, but may be as high as 11.5 percentage points, or 20 percent when we restrict measurement error in initial conditions to be only time-varying. The number of years spent in poverty is also downward biased by measurement error. Many studies up to date have purported to find substantial economic mobility, but this study suggests that the estimated high mobility in Korea is overstated due to measurement error.

Looking at the pce quintile (5×5) transition matrices has an advantage in exploring the potential difference of the impact of measurement error on transition probabilities for each class of expenditure. This study finds that the magnitudes of biases differ among classes. Among the probabilities of immobility for households which remain on the diagonals in a transition matrix, the probability of staying at the bottom is most affected by measurement error. For the 5×5 transition matrix, the second most affected probability is staying at the top. However, the diagonals of the matrices, which indicate economic immobility, seem to be more affected by measurement error when the number of classes increases.

In view of these results, for Korea at least, permanent poverty seems to be somewhat more important than people might have realized based on transition probabilities from uncorrected survey data. This suggests that policies aimed at lowering persistent poverty that focus on factors like education and health that can permanently raise people out of poverty, are still very important. As noted in the introduction, people residing in lower income countries with much larger rural populations are likely to experience higher income risk than the current, largely urban, population in Korea, and if they have problems smoothing consumption, then expenditures will be more variable relative to mean expenditure than we see in the KLIPS data. Still, measurement error in expenditures will be a problem for surveys in such settings just as it is in KLIPS, and likely will lead to an overstatement of consumption mobility, just as we have found.

References

[1] Antman, Francisca and David McKenzie. 2007a. “Earnings mobility and measurement error: A pseudo-panel approach”, Economic Development and Cultural Change, 56(1):125-162.

[2] Antman, Francisca and David McKenzie. 2007b. "Poverty Traps and Nonlinear Income Dynamics with Measurement Error and Individual Heterogeneity." Journal of Development Studies, 43(6): 1057-1083.

[3] Arellano, Manuel and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." The Review of Economic Studies, 58(2): 277-97.

[4] Barrett, Christopher and John McPeak, 2006. Poverty, Inequality and Development. Springer US.

[5] Baulch, Bob and John Hoddoinott. 2000. "Economic Mobility and Poverty Dynamics in Developing Countries." The Journal of Development Studies, 36(6): 1-24.

[6] Bond, Stephen and Frank Windmeijer. 2002. "Projection Estimators for Autoregressive Panel Data Models", The Econometrics Journal, 5(2): 457-79.

[7] Bound, John, Charlie Brown and Nancy Mathiowitz. 2001. “Measurement error in survey data”, in J.J. Heckman and E. Leamer (eds.), Handbook of Econometrics, Volume 5, Amsterdam: North Holland Press.

[8] Bound, John and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" Journal of Labor Economics, 12: 345-68.

[9] Browning, Martin, Thomas F. Crossley, and Guglielmo Weber. 2003. "Asking Consumption Questions in General Purpose Surveys." Economic Journal, 113: F540-F67.

[10] Datt, Garurav and Martin Ravallion. 1992. “Growth and redistribution components of changes in poverty measures: A decomposition with applications to Brazil and India in the 1980s”, Journal of Development Economics, 38:275-295.

[11] Davidson, Russell and Jean-Yves Duclos. 2000. “Statistical inference for stochastic dominance and for the measurement of poverty and inequality”, Econometrica, 68:1435-1464.

[12] Deaton, Angus S. 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy: Johns Hopkins University Press.

[13] Dercon, Stefan. 1998. "Wealth, Risk and Activity Choice: Cattle in Western Tanzania." Journal of Development Economics, 55: 1-42.

[14] Dercon, Stefan and Pramila Krishnan. 2000. "Vulnerability, Seasonality and Poverty in Ethiopia." Journal of Development Studies, 36(6): 25-53.

[15] Duclos, Jean-Yves, Abdelkrim Arrar and John Giles. 2010. "Chronic and Transient Poverty: Measurement and Estimation with Evidence from China." Journal of Development Economics, 91(2):266-277.

[16] Fields, Gary. 2006. "The Many Facets of Economic Mobility", in M.McGillivray (ed.), Inequality, Poverty and Well-Being. New York: Palgrave Macmillan.

[17] Fields, Gary, Paul Cichello, Samuel Freije, Marta Menendez and David Newhouse. 2003. "Household income dynamics: A four country study", Journal of Development Studies, 40(2):30-54.

[18] Foster, James E. and Wolfson, Michael C. 1994. "Polarization and the Decline of the Middle Class: Canada and the US." Mimeo.

[19] Gaiha, Raghav and Anil B. Deolalikar. 1993. "Persistent, Expected and Innate Poverty: Estimates for Semi-Arid Rural South India, 1975–84." Cambridge Journal of Economics, 17: 409-421.

[20] Gibson, John and Paul Glewwe. 2005. "Analysis of Poverty Dynamics", in (Eds.), Handbook on Poverty: Concepts, Methods and Policy Use, United Nations.

[21] Glewwe, Paul. forthcoming. "Was Vietnam's economic growth in the 1990s pro-poor?, Economic Development and Cultural Change, forthcoming.

[22] Gottschalk, Peter, and Enrico Spolaore. 2002. "On the evaluation of economic mobility", Review of Economic Studies, 69:191-208.

[23] Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. "Estimating Vector Autoregressions with Panel Data." Econometrica, 56(6): 1371-95.

[24] Jalan, Jyotsna and Martin Ravallion, 1998. "Transient Poverty in Postreform Rural China." Journal of Comparative Economics, 26: 338-357.

[25] Jalan, Jyotsna and Martin Ravallion, 2000. "Is Transient Poverty Different? Evidence for Rural China." Journal of Development Studies, 36(6): 82-99.

[26] Lee, Nayoung. 2009. "Measurement Error and Its Impact on Estimates of Income and Consumption Dynamics", IEPR Working Paper, 08-11, University of Southern California.

[27] Lillard, Lee A. and Robert J. Willis. 1978. "Dynamic Aspects of Earning Mobility." Econometrica, 46(5): 985-1012.

[28] Lokshin, Micheal and Ravallion, Martin. 2004. "Household Income Dynamics in TwoTransition Economies." Studies in Nonlinear Dynamics and Econometrics, 8(3): Article 4.

[29] Luttmer, Erzo. 2002. "Measuring Economic Mobility and Inequality: Disentangling Real Events from Noisy Data." Mimeo.

[30] McGarry, Kathleen. 1995. "Measurement Error and Poverty Rates of Widows", The Journal of Human Resources, 30(1): 113-134.

[31] Park, Neung-Hoo. 2001. "Poverty Rate and Poverty Line in Korea", Mimeo. World Bank Institute Workshop on Strengthening Poverty Data Collection and Analysis.

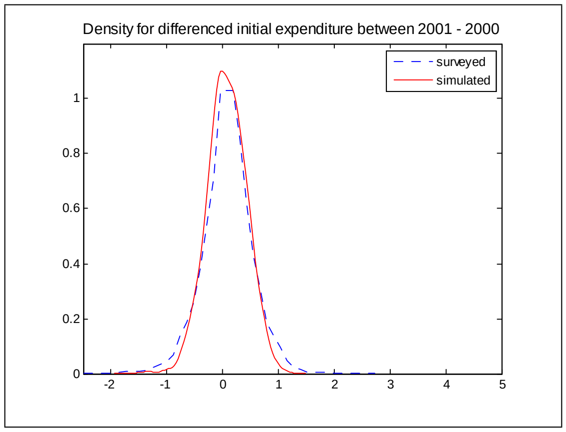
[32] Rosenzweig, Mark. 1994. "Human capital accumulation, the family and economic development", in S. Asefa and W.-C. Huang (eds.), Human Capital and Economic Development, Kalamazoo: W.E. Upjohn Institute for Employment Research.

[33] Villanger, Espen. 2003. "The Effects of Disasters on Income Mobility: Bootstrap Inference and Measurement Simulations." Mimeo.

[34] Zimmerman, Frederick and Michael Carter. 2003. "Asset Smoothing, Consumption Smoothing and the Reproduction of Inequality Under Risk and Subsistence Constraints", Journal of Development Economics, 71: 233-260.

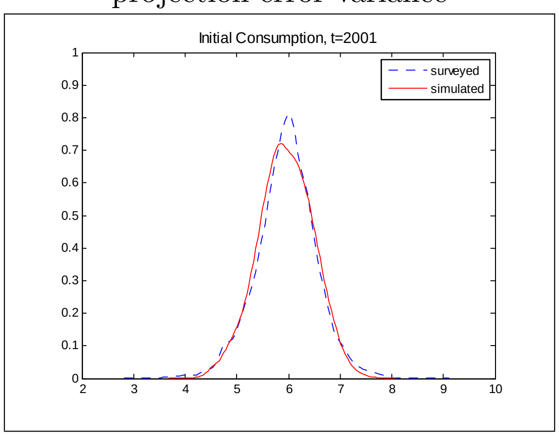
Figure 1: Simulated Initial Expenditures

Differenced initial expenditure between 2001- 2000

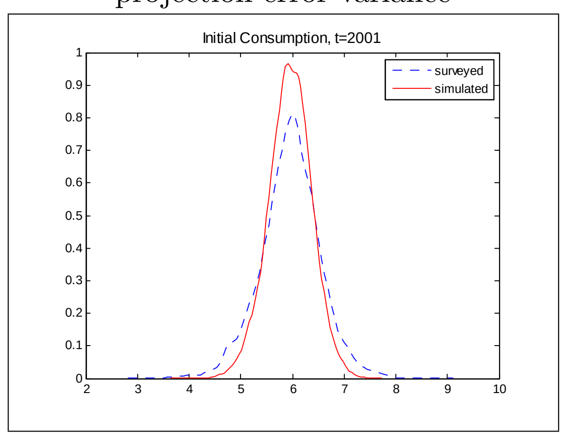


A. Level initial expenditures in 2001, from projections

A1. upper bound of
projection error variance



A2. lower bound of
projection error variance



B. Initial expenditure in 2001, from stationary distribution

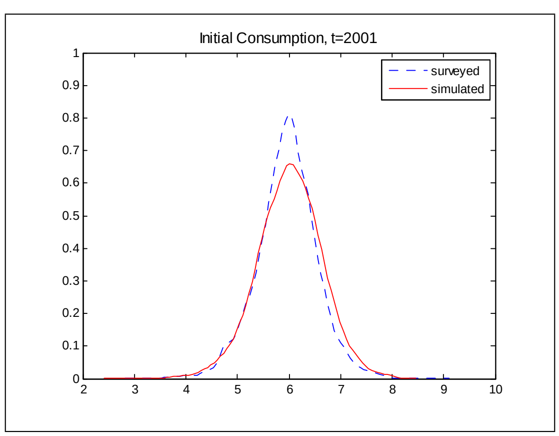
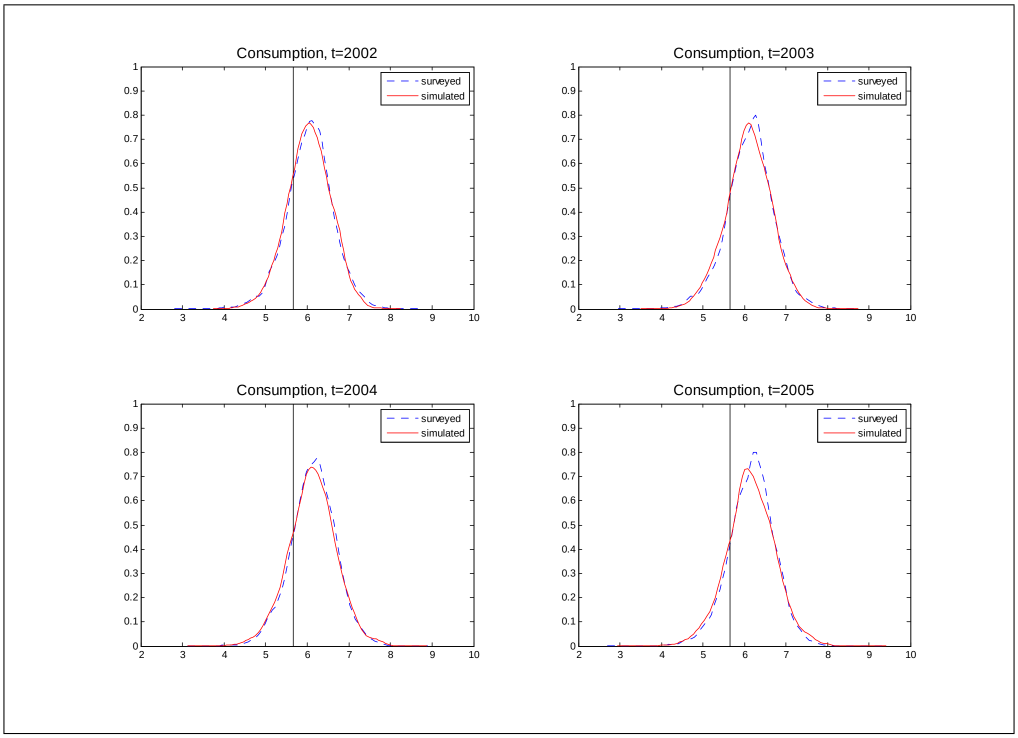


Figure 2: Density of Simulated Expenditure from 2002 to 2005

A1. Initial expenditures based on the upper bound of projection error variance



A2. Initial expenditures based on the lower bound of projection error variance

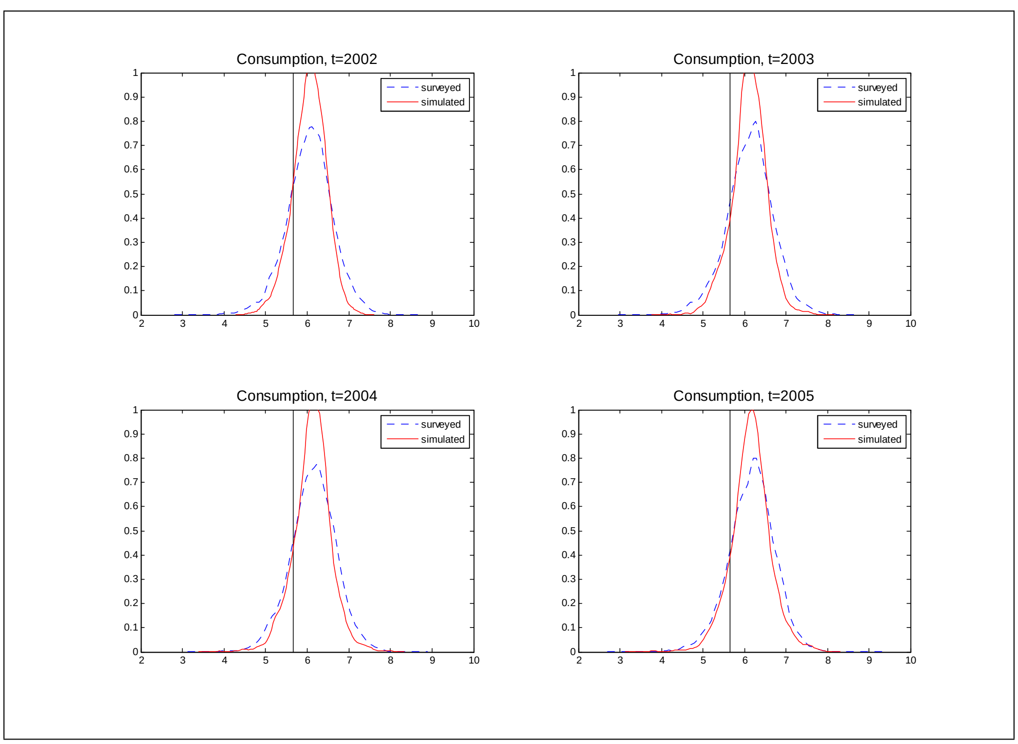
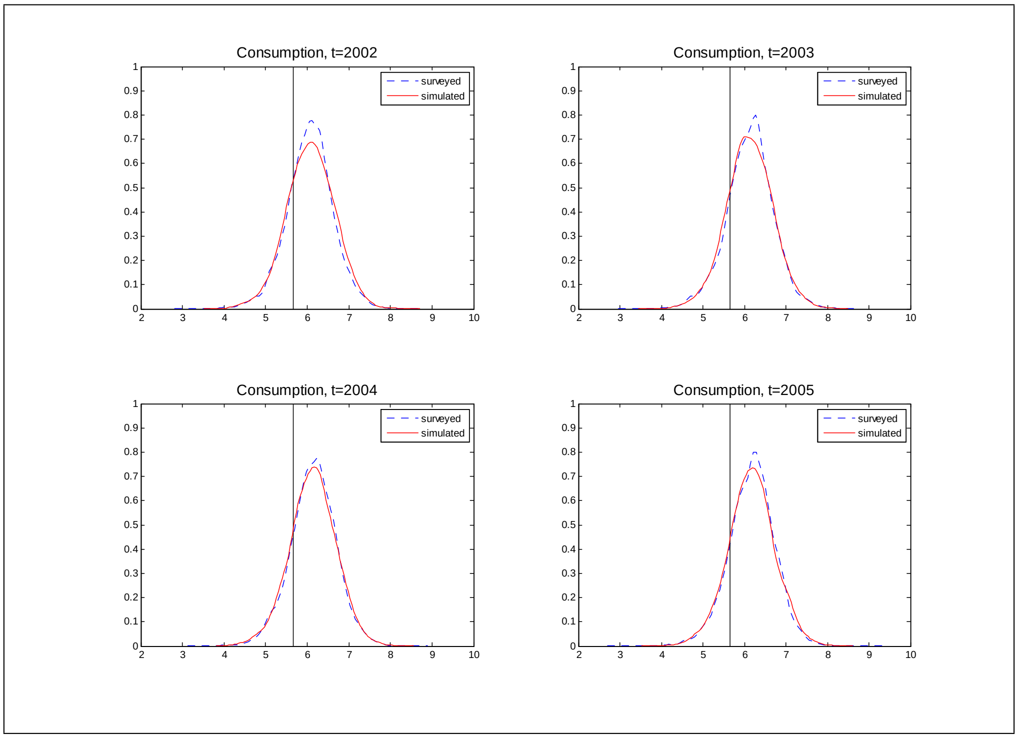


Figure 2: Continued

B. Initial expenditure from stationary distribution



Vertical lines indicate the minimum cost of living (MCL).

Table 1: Summary Statistics - Mean and Standard Deviation

Variable	Year 2000	2001	2002	2003	2004	2005
Log (per capita expenditure)	5.87	5.95	6.06	6.11	6.13	6.16
	(0.56)	(0.57)	(0.56)	(0.56)	(0.53)	(0.54)
Household size	3.53	3.58	3.43	3.39	3.35	3.28
	(1.34)	(1.35)	(1.31)	(1.32)	(1.32)	(1.32)
Male aged over 65	0.05	0.05	0.06	0.07	0.07	0.08
	(0.14)	(0.14)	(0.16)	(0.17)	(0.17)	(0.18)
Female aged over 55	0.15	0.15	0.17	0.18	0.19	0.2
	(0.26)	(0.26)	(0.28)	(0.28)	(0.29)	(0.29)
Sex of head	1.14	1.15	1.15	1.16	1.16	1.17
	(0.35)	(0.36)	(0.36)	(0.37)	(0.37)	(0.38)
Education of head	10.21	10.21	10.17	10.17	10.15	10.16
	(4.44)	(4.43)	(4.45)	(4.45)	(4.44)	(4.44)
Seoul dummy	0.24	0.23	0.23	0.22	0.21	0.21
	(0.42)	(0.42)	(0.42)	(0.41)	(0.41)	(0.41)
Nonspouse dummy	0.20	0.20	0.20	0.22	0.22	0.24
	(0.40)	(0.40)	(0.40)	(0.41)	(0.41)	(0.42)
Age of head	49.56	50.29	51.42	52.86	53.82	54.76
	(12.89)	(12.91)	(12.83)	(12.78)	(12.67)	(12.56)
Obs #	3,053	3,068	3,072	2,951	2,851	2,774

Observations for households analyzed in this study
Standard deviations in parentheses

Table 2: Sample Size

Year	Expenditure from survey	Requirements for simulation	Expenditure from Simulation
2001-2000	3,074	Expenditure from survey at 2000 and 2001	3,053
		Household characteristics, X at 2000 and 2001	
		A. Initial expenditure from projection	
2001	3,241	Expenditure from survey at 2001	3,068
		Household characteristics, X at 2000 and 2001	
		B. Initial expenditure from stationary distribution	
		Household characteristics, X at 2001	3,229
2002	3,359	Differenced initial condition, 2001 - 2000	3,072
		Level initial condition at 2001	
		Household characteristics, X at 2002 and 2001	
2003	3,241	Differenced simulated expenditure, 2002 - 2001	2,951
		Simulated expenditure at 2002	
		Household characteristics, X at 2003 and 2002	
2004	3,144	Differenced simulated expenditure, 2003 - 2002	2,851
		Simulated expenditure at 2003	
		Household characteristics, X at 2004 and 2003	
2005	3,104	Differenced simulated expenditure, 2004 - 2003	2,774
		Simulated expenditure at 2004	
		Household characteristics, X at 2005 and 2004	

The samples are restricted to the samples that are used to estimate the basic standard consumption dynamics model, equation (5).
For comparison between the cases of initial expenditures from projection and stationary distribution, the samples sizes are further restricted to the case of initial expenditures by projection

Table 3: Basic Results

Parameter	Estimate	
γ	0.375	
σ^2_ϵ	0.054	
σ^2_ν	0.032	
A. Initial observations by projections		
$\kappa = Var(\alpha_i{+}(1-\gamma)e_i)$	0.226	
$Var(\psi_0)$	0.180	
$Var(\psi_1)$	0.229	
$Cov(\psi_0, \psi_1)$	0.091	
$E[\zeta_0 \Delta\varepsilon_2]$	-0.077	
$E[\zeta_1 \Delta\varepsilon_2]$	-0.069	
	Projection errors	
	A1. upper bound	A2. lower bound
$Var(\zeta_{i0})$	0.116	0.116
$Var(\zeta_{i1})$	0.197	0.081 (constrained)
$Var(e_i)$	0.000	0.116
$Var(\alpha_i)$	0.226	0.180
c_1		0.244
c_2		0.374
c_3		0.654
c_4		0.490
c_5		0.327
c_6		0.163
B. Initial observation from stationary distribution		
$Var(\phi_{i1})$	0.063	
$\bar{\alpha}_i$	4.587	
$Var(\alpha_i)$	0.242	

Other estimates of equation (5) and their standard errors are in Appendix Table A.3.

Table 4: Number of years spent in poverty from 2002 to 2005

Number of years in which poor	Never	1	2	3	Always
(1)Observed expenditure					
Obs #	1,758	405	272	167	168
%	63.37	14.60	9.81	6.02	6.06
(2) Simulated true expenditure					
A1. Initial expenditures based on the upper bound of projection error variance					
Obs #	1,825	314	215	162	254
%	65.79	11.32	7.75	5.84	9.16
A2. Initial expenditures based on the lower bound of projection error variance					
Obs #	2,037	321	175	117	120
%	73.43	11.57	6.31	4.22	4.33
B. Initial observation from stationary distribution					
Obs #	1,811	353	228	168	210
%	65.29	12.73	8.22	6.06	7.57

The number of samples which exist in 2005 (and so all over the years) are 2,774.
% of HHs, for a certain number of years in which poor, is calculated based on these samples.

Table 5: 2x2 Poverty Transition Matrices for Two Consecutive Years

	Poverty status at t	
Poverty status at t-1	Poverty	Not in poverty
Poverty		
(1) Observed expenditure	56.45	43.55
(2) Simulated expenditure		
A1. upper bound of projection error variance	67.92	32.08
A2. lower bound of projection error variance	59.83	40.17
B. stationary distribution	64.78	35.22
Not in poverty		
(1) Observed expenditure	7.81	92.19
(2) Simulated expenditure		
A1. upper bound of projection error variance	6.78	93.22
A2. lower bound of projection error variance	5.81	94.19
B. stationary distribution	6.53	93.47

Row percentages are presented.
Averaged probability for all years (2002-2005)
Poverty status is based on the Minimum Cost of Living (MCL).
MCL = 5.7 , around \$8 per day in 2000.
Expenditure is converted by currency in 2000.

Table 6: 5x5 Expenditure Quintile Transition Matrices for Two Consecutive Years

		Class at t				
Class at t-1		1	2	3	4	5
1	(1) Observed expenditure	58.11	24.45	9.66	6.16	1.62
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	67.06	22.35	8.27	1.57	0.75
	A2. lower bound of projection error variance	61.77	20.72	10.43	5.25	1.84
	B. stationary distribution	62.15	27.70	7.86	2.16	0.13
2	(1) Observed expenditure	21.11	34.57	24.26	15.68	4.38
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	20.71	36.71	28.15	12.28	2.16
	A2. lower bound of projection error variance	19.09	27.47	25.26	19.16	9.03
	B. stationary distribution	18.92	40.28	29.51	9.66	1.62
3	(1) Observed expenditure	7.84	18.77	33.20	30.76	9.44
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	4.84	20.08	36.39	30.18	8.51
	A2. lower bound of projection error variance	7.63	20.60	26.35	28.05	17.37
	B. stationary distribution	3.77	23.99	36.15	30.10	6.00
4	(1) Observed expenditure	3.66	8.48	20.20	42.21	25.45
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	1.18	8.27	22.78	38.64	29.13
	A2. lower bound of projection error variance	2.49	12.76	22.49	30.83	31.44
	B. stationary distribution	0.65	6.80	23.60	45.19	23.76
5	(1) Observed expenditure	1.20	2.28	5.17	22.24	69.11
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	0.20	0.71	4.56	18.15	76.37
	A2. lower bound of projection error variance	0.75	4.32	11.90	21.99	61.04
	B. stationary distribution	0.35	0.98	5.10	22.71	70.86

Row percentages are presented
Averaged probability for all years (2002-2005)
Expenditure classes are based on the 2002 quintile (1: poorest).

Table 7: Summary of Transition Matrices

	Percent of households that:			Percent of households in bottom that:		
	Remain on diagonal	Move by one quintile	Move by two or more quintiles	Remain in bottom	Move by one quintile	Move by two or more quintiles
(1) Observed expenditure						
2×2	85.02	14.98		56.45	43.55	
5×5	47.63	37.52	14.85	47.63	37.52	14.85
(2) Simulated true expenditure						
A1. Initial expenditures based on the upper bound of projection error variance						
2×2	87.95	12.05		67.92	32.08	
5×5	51.83	37.77	10.41	51.83	37.77	10.41
A2. Initial expenditures based on the lower bound of projection error variance						
2×2	89.44	10.56		59.83	40.17	
5×5	41.86	37.83	20.31	41.86	37.83	20.31
B. Initial observation from stationary distribution						
2×2	87.62	12.38		64.78	35.22	
5×5	50.54	40.49	8.97	50.54	40.49	8.97

Appendix

Table A1: Sample Size Construction

for the estimation of equation (5)

	Year				
Variable	2002	2003	2004	2005	Total
Expenditure at t	3,516	3,638	3,637	3,639	14,430
- Expenditure at t-1	3,333	3,354	3,459	3,495	13,641
	(183)	(284)	(178)	(144)	(789)
- Expenditure at t-2	3,135	3,191	3,220	3,337	12,883
	(198)	(163)	(239)	(158)	(758)
- Expenditure at t-3	3,050	3,005	3,073	3,121	12,249
	(85)	(186)	(147)	(216)	(634)
- Other covariates at t	3,013	2,984	3,051	3,103	12,151
	(37)	(21)	(22)	(18)	(98)
- Other covariates at t-1	2,999	2,965	3,048	3,097	12,109
	(14)	(19)	(3)	(6)	(42)
- HH income satisfaction at t-1	2,974	2,942	3,023	3,035	11,974
	(25)	(23)	(25)	(62)	(135)
- HH income satisfaction at t-2	2,954	2,923	2,993	3,026	11,896
	(20)	(19)	(30)	(9)	(78)
- HH income satisfaction at t-3	2,931	2,903	2,975	2,991	11,800
	(23)	(20)	(18)	(35)	(96)
- Outliers	2,930	2,902	2,974	2,990	11,796
	(1)	(1)	(1)	(1)	(1)

Marginal loss of observations in parenthesis

Table A2: Data Availability

Year	Income	Expenditure (1)	Expenditure (2)	Income Satisfaction
1997		Yes		
1998	Yes	Yes	Yes	
1999	Yes	Yes		Yes
2000	Yes	Yes	Yes	Yes
2001	Yes	Yes	Yes	Yes
2002	Yes	Yes	Yes	Yes
2003	Yes	Yes	Yes	Yes
2004	Yes	Yes	Yes	Yes
2005	Yes	Yes	Yes	Yes
2006				Yes

Expenditure (1) refers to directly-asked expenditure and expenditure (2) refers to aggregated one from disaggregated questions

Table A3: Two-step GMM Estimates after First Differencing
Pooled over years, t=2002, 2003, 2004 and 2005

Dependent variable: $\triangle$Log expenditure t	Coefficient
$\triangle$Log expenditure $t-1$	0.376**
	(0.075)
$\triangle$Household size	-0.294**
	(0.015)
$\triangle$Male aged over 65	0.042
	(0.093)
$\triangle$Female aged over 55	0.140*
	(0.068)
$\triangle$Sex of head	0.005
	(0.078)
$\triangle$Education of head	0.011
	(0.013)
$\triangle$Seoul dummy	-0.031
	(0.055)
$\triangle$Nonspouse dummy	0.093
	(0.050)
$\triangle$Age of head	0.005
	(0.014)
$\triangle$Square age of head	0.000
	(0.000)
$\triangle$Year dummy (2002)	0.023
	(0.013)
$\triangle$Year dummy (2003)	0.032
	(0.022)
$\triangle$Year dummy (2004)	0.004
	(0.028)
$\triangle$Year dummy (2005)	0.016
	(0.032)
Hansen J statistics	9.815
(p value)	0.57
N	11,796

All covariates are first differenced (denoted by $\triangle$)
External IVs: HH income satisfaction of head at year t-2 and t-3
Internal IVs: log expenditure. at t-3 and earlier
** significant at 1%, * significant at 5%

Table A4: OLS estimates for Initial Conditions

Dependent variable:	Dependent Variable	
	ΔLog expenditure at 2001	Log expenditure at 2001
Household size in 2001	-0.317**	-0.191**
	(0.026)	(0.029)
Male aged over 65 in 2001	-0.065	-0.050
	(0.161)	(0.182)
Female aged over 55 in 2001	0.147	-0.136
	(0.139)	(0.157)
Sex of head in 2001	0.021	-0.207*
	(0.092)	(0.104)
Education of head in 2001	0.009	0.026*
	(0.010)	(0.011)
Seoul dummy in 2001	-0.030	-0.018
	(0.073)	(0.082)
Nonspouse dummy in 2001	-0.015	-0.067
	(0.075)	(0.085)
Age of head in 2001	-0.017	0.030
	(0.014)	(0.016)
Square age of head in2001	0.000	0.000
	(0.000)	(0.000)
Household size in 2000	0.306**	0.034
	(0.026)	(0.029)
Male aged over 65 in 2000	-0.142	-0.518**
	(0.162)	(0.182)
Female aged over 55 in 2000	-0.234	-0.111
	(0.137)	(0.154)
Sex of head in 2000	0.025	0.122
	(0.093)	(0.105)
Education of head in 2000	-0.009	0.024*
	(0.010)	(0.011)
Seoul dummy in 2000	0.009	0.122
	(0.073)	(0.082)
Nonspouse dummy in 2000	-0.040	0.039
	(0.076)	(0.086)
Age of head in 2000	0.007	0.004
	(0.014)	(0.016)
Square age of head in 2000	0.000	0.000
	(0.000)	(0.000)
Constant	0.413*	5.322**
	(0.135)	(0.152)
R-squared	0.073	0.288
N	3,053	3,068

** significant at 1%, * significant at 5%

Table A.5 Statistics for Surveyed and Simulated Expenditures

Year	Obs #	Mean	Std. Dev.	Min	Max
(1) Observed expenditure					
2001-2000	3,053	0.084	0.441	-2.605	2.486
2001	3,068	5.954	0.567	3.543	8.149
2002	3,072	6.057	0.558	3.335	8.458
2003	2,951	6.114	0.555	3.300	8.375
2004	2,851	6.125	0.534	3.446	7.869
2005	2,774	6.163	0.540	3.014	9.078
(2) Simulated true expenditure					
A1. Initial expenditures based on the upper bound of projection error variance					
2001-2000	3,053	0.081	0.357	-1.707	1.268
2001	3,068	5.946	0.536	4.022	7.734
2002	3,072	6.045	0.525	4.069	7.906
2003	2,951	6.103	0.550	3.851	8.401
2004	2,851	6.119	0.575	3.481	8.530
2005	2,774	6.155	0.595	3.292	9.066
A2. Initial expenditures based on the lower bound of projection error variance					
2001-2000	3,053	0.078	0.356	-1.650	1.365
2001	3,068	5.951	0.418	3.903	7.466
2002	3,072	6.053	0.398	4.522	7.350
2003	2,951	6.112	0.420	3.998	7.925
2004	2,851	6.121	0.451	3.632	8.096
2005	2,774	6.154	0.482	3.386	8.050
B. Initial observation from stationary distribution					
2001	3,068	6.007	0.632	2.786	8.085
2002	3,072	6.079	0.578	3.868	8.342
2003	2,951	6.118	0.561	3.788	8.154
2004	2,851	6.128	0.556	4.159	8.224
2005	2,774	6.166	0.560	3.857	8.274

Table A.6 Poverty Rates from 2002 to 2005

Poor (below MCL)		2002	2003	2004	2005	Avg.
(1) Observed expenditure						
	Obs #	719	551	510	446	557
	%	23.41	18.67	17.89	16.08	19.11
(2) Simulated true expenditure						
A1. Initial expenditures based on the upper bound of projection error variance						
	Obs #	689	582	570	523	591
	%	22.43	19.72	19.99	18.85	20.30
A2. Initial expenditures based on the lower bound of projection error variance						
	Obs #	463	388	388	363	401
	%	15.07	13.15	13.61	13.09	13.75
B. Initial observation from stationary distribution						
	Obs #	697	583	527	470	569
	%	22.69	19.76	18.49	16.94	19.84

Table A7: 2×2 Poverty Transition Matrices for Two Consecutive Years

Poverty status at t-1	Poverty status at t	
	Poverty	Not in poverty
Investigated Years: from 2002 to 2003		
Poverty		
(1) Observed expenditure	53.69	46.31
(2) Simulated expenditure		
A1. upper bound of projection error variance	64.47	35.53
A2. lower bound of projection error variance	56.11	43.89
B. stationary distribution	63.21	36.80
Non in poverty		
(1) Observed expenditure	7.96	92.04
(2) Simulated expenditure		
A1. upper bound of projection error variance	6.66	93.35
A2. lower bound of projection error variance	5.58	94.42
B. stationary distribution	6.90	93.11
Investigated Years: from 2003 to 2004		
Poverty		
(1) Observed expenditure	59.02	40.98
(2) Simulated expenditure		
A1. upper bound of projection error variance	71.28	28.72
A2. lower bound of projection error variance	62.33	37.67
B. stationary distribution	66.90	33.10
Non in poverty		
(1) Observed expenditure	8.46	91.54
(2) Simulated expenditure		
A1. upper bound of projection error variance	7.31	92.69
A2. lower bound of projection error variance	6.37	93.63
B. stationary distribution	6.44	93.56

Table A7: Continued

Poverty status at t-1	Poverty status at t	
	Poverty	Not in poverty
Investigated Years: from 2004 to 2005		
Poverty		
(1) Observed expenditure	57.55	42.46
(2) Simulated expenditure		
A1. upper bound of projection error variance	68.65	31.35
A2. lower bound of projection error variance	61.77	38.24
B. stationary distribution	64.50	35.50
Non in poverty		
(1) Observed expenditure	6.99	93.01
(2) Simulated expenditure		
A1. upper bound of projection error variance	6.36	93.64
A2. lower bound of projection error variance	5.47	94.54
B. stationary distribution	6.27	93.73

Row percentages are presented.
Poverty status is based on the Minimum Cost of Living (MCL).
MCL = 5.7 , around $8 per day in 2000.
Expenditure is converted by currency in 2000.

Table A8: 5×5 Expenditure Quintile Transition Matrices for Two Consecutive Years

		Class at 2003				
Class at 2002		1	2	3	4	5
1	(1) Observed expenditure	56.33	25.87	11.74	4.95	1.10
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	64.76	23.78	9.78	1.01	0.67
	A2. lower bound of projection error variance	56.07	23.25	12.31	5.64	2.74
	B. stationary distribution	60.10	28.96	8.75	2.02	0.17
2	(1) Observed expenditure	22.52	32.28	23.78	15.12	6.30
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	17.92	36.35	29.18	14.16	2.39
	A2. lower bound of projection error variance	15.75	29.15	28.14	19.60	7.37
	B. stationary distribution	20.47	37.77	29.62	11.31	0.83
3	(1) Observed expenditure	7.46	21.56	31.68	27.03	12.27
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	4.06	17.60	37.06	32.15	9.14
	A2. lower bound of projection error variance	5.89	21.21	25.42	28.62	18.86
	B. stationary distribution	3.74	25.68	34.18	30.78	5.61
4	(1) Observed expenditure	4.49	8.64	19.86	41.45	25.56
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	1.18	8.91	19.50	37.98	32.44
	A2. lower bound of projection error variance	2.71	11.86	20.85	30.51	34.07
	B. stationary distribution	0.68	6.28	23.77	43.46	25.81
5	(1) Observed expenditure	1.53	3.74	5.60	23.09	66.04
	(2) Simulated expenditure					
	A1. the upper bound of projection error variance	0.17	0.68	4.61	17.24	77.30
	A2. the lower bound of projection error variance	0.85	3.42	12.48	24.44	58.80
	B. stationary distribution	0.69	0.52	5.53	22.97	70.29

Table A8: Continued

Class at 2003		Class at 2004				
		1	2	3	4	5
1	(1) Observed expenditure	59.96	23.05	8.98	7.03	0.98
	(2) Simulated expenditure					
	A1. the upper bound of projection error variance	69.51	21.78	6.53	1.78	0.40
	A2. the lower bound of projection error variance	67.40	18.38	8.53	5.03	0.66
	B. stationary distribution	64.59	25.96	7.65	1.81	0.00
2	(1) Observed expenditure	22.08	34.72	25.47	13.77	3.96
	(2) Simulated expenditure					
	A1. the upper bound of projection error variance	24.09	38.06	26.52	9.92	1.42
	A2. the lower bound of projection error variance	23.00	24.17	24.56	18.52	9.75
	B. stationary distribution	18.60	43.33	27.54	8.42	2.11
3	(1) Observed expenditure	8.47	17.50	35.54	30.94	7.55
	(2) Simulated expenditure					
	A1. the upper bound of projection error variance	5.78	20.67	35.55	29.25	8.76
	A2. the lower bound of projection error variance	8.42	19.83	26.32	29.83	15.61
	B. stationary distribution	3.79	24.31	38.62	27.41	5.86
4	(1) Observed expenditure	3.12	9.50	22.12	40.97	24.30
	(2) Simulated expenditure					
	A1. the upper bound of projection error variance	1.02	8.33	23.81	40.65	26.19
	A2. the lower bound of projection error variance	3.25	13.96	24.19	29.06	29.55
	B. stationary distribution	0.79	6.80	23.58	45.25	23.58
5	(1) Observed expenditure	1.29	1.93	5.47	22.83	68.49
	(2) Simulated expenditure					
	A1. based on the highest projection error	0.43	0.72	5.35	21.27	72.21
	A2. based on the highest projection error	0.87	4.76	12.70	22.37	59.31
	B. from stationary distribution	0.18	1.40	5.44	23.33	69.65

Table A8: Continued

Class at 2004		Class at 2005 1	2	3	4	5
1	(1) Observed expenditure	58.14	24.33	8.04	6.60	2.89
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	67.34	21.24	8.22	2.00	1.20
	A2. lower bound of projection error variance	63.35	19.88	9.94	4.97	1.86
	B. stationary distribution	62.16	27.98	6.88	2.75	0.23
2	(1) Observed expenditure	18.02	37.58	23.52	18.68	2.20
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	20.62	35.70	28.60	12.42	2.66
	A2. lower bound of projection error variance	19.07	29.07	22.09	19.30	10.47
	B. stationary distribution	17.59	39.86	31.42	9.16	1.97
3	(1) Observed expenditure	7.62	16.91	32.53	34.76	8.18
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	4.71	22.22	36.54	29.00	7.53
	A2. lower bound of projection error variance	8.70	20.74	27.41	25.56	17.59
	B. stationary distribution	3.77	21.96	35.68	32.08	6.52
4	(1) Observed expenditure	3.46	7.37	18.65	44.06	26.47
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	1.34	7.56	25.04	37.31	28.74
	A2. lower bound of projection error variance	1.49	12.42	22.35	32.95	30.80
	B. stationary distribution	0.49	7.28	23.46	46.76	22.01
5	(1) Observed expenditure	0.80	1.27	4.46	20.86	72.61
	(2) Simulated expenditure					
	A1. upper bound of projection error variance	0.00	0.72	3.74	15.83	79.71
	A2. lower bound of projection error variance	0.56	4.62	10.64	19.61	64.57
	B. stationary distribution	0.17	1.04	4.33	21.84	72.62

Row percentages are presented
Expenditure classes are based on the 2002 quintile (1: poorest).

Departamento de Economia PUC-Rio
Pontifícia Universidade Católica do Rio de Janeiro
Rua Marques de Sâo Vicente 225 - Rio de Janeiro 22453-900, RJ
Tel.(21) 35271078 Fax (21) 35271084
www.econ.puc-rio.br
flavia@econ.puc-rio.br