Estimation of Poverty Transition Matrices with Noisy Data

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

October 31, 2009

Abstract

This paper investigates potential measurement error biases in poverty transition matrices. We compare transition matrices based on survey expenditure data to measurement-error-free simulated expenditure, which is built on initial conditions and parameters estimated from a basic consumption dynamics model allowing for measurement error. We find that 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 16 and 18% of households initially in poverty are found to be out of poverty.

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

---

[*] 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 used to investigate poverty dynamics, 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 plays on poverty transition matrices and on more general expenditure transition matrices that measure economic mobility, using longitudinal data from Korea.

In this study, we extend 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 do contain measurement error. We allow for fairly general types of measurement error, in particular we allow both for random, time-varying error, of the type most often thought of as being a problem, and for time-invariant measurement error, which can be non-classical.

---

[1] In principle, there is a distinction between consumption and expenditures in that expenditures include purchases, which may be more or less than consumption. However, in practice, expenditures also partly include consumption, for food consumed out of own production in particular, so is really a mixture. 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).

This study uses data from the Korean Labor and Income and Panel Study (KLIPS) from 1998 to 2006. Results suggest that the transition matrices based on survey data are biased when expenditure data are reported with errors. In particular, in these data measurement error greatly 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, only 16 to 18% 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 24% are poor in only one or two years. Hence most of the poverty is transitory using these estimates. On the other hand, using our simulated, measurement error-free data, some 38% of households are estimated to be poor at least one of these four years, but nearly half of those, 18.5%, are poor each year and another 8% 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 much more permanent when measurement error is accounted for in the poverty dynamics. We explore the robustness of our results to certain modeling assumptions and find them to be robust.

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 would guess in those cases, too, that measurement error may have important biases on mobility estimates that ignore measurement error, very likely similar in direction to those we have found.

---

[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.

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 its static status. There exist very different models that researchers have used to estimate consumption and/or poverty mobility.[4] One important method in the literature used to estimate consumption or income mobility are dynamic models using an AR(1) model with exogenous covariates.[5] The estimate of the autoregressive coefficient from such a dynamic model can be taken as one 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 value of one corresponding to immobility and a lower positive number 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 some form of 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 good discussion).

Another type of research methodology often used to estimate poverty dynamics is to estimate poverty transition matrices between two years and/or to count the number of

---

[4] Consumption, or expenditures, are generally preferred to income to look at poverty, both because income is thought to generally have much higher measurement error 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 higher than for expenditure.

[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) convergence in income or consumption provided the coefficient is a positive fraction, divergence if the coefficient is greater than one, as in growth models.

years out of the total for which an individual household was in poverty. Many studies have used these two methods 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). Accordingly, 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; McCulloch and Baulch, 2000).[7] 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 for consumption smoothing by providing social safety nets is likely to be more important to reduce transient poverty.

Poverty dynamics is also related to the poverty trap literature which, if true, implies permanent rather than transitory poverty. Permanently low incomes will also lead to less asset accumulation, which may lead to poverty traps. Households with a low endowment of assets may have an inability to translate these into higher incomes because they may pursue low-risk and low-return activities (Dercon, 1998; Zimmerman and Carter, 2003; Barrett and McPeak, 2004).[8] 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 such evidence (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

---

[7] Jalan and Ravallion (1998) define 'transient poverty' as the poverty that can be attributed to inter-temporal variability in consumption and 'chronic poverty' as the poverty that persists in mean consumption over time. They construct an inter-temporal poverty measure using a household i's consumption stream, and measure chronic poverty in the same way but using mean consumption over time. Then, transient poverty is the remainder obtained by subtracting chronic poverty from poverty.

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

measurement error are for income and earnings, not expenditure, and use administrative data (considered truth) 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 they may be relevant for this study, our study allows for measurement error to be correlated with true expenditure, and tests in Lee (2009) fail to reject that measurement error for both income and expenditure is not serially correlated in the data we use (after removing household fixed effects).

As discussed in Lee (2009), there exist a small number of consumption studies that compare measured recalled expenditure asked by enumerators with expenditure as measured by daily diaries (eg. Ahmed et al., 2006), finding measurement error in expenditure. While in these studies, diaries are assumed to represent truth, there does exist evidence that even diaries contain measurement errors (Browning, Crossley and Weber, 2003).

Studies that have examined income or consumption dynamics while accounting for measurement error without administrative records to compare to, 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. This strategy makes the strong assumptions that there do not exist any time-invariant measurement error and also that no time-invariant omitted variables exist that are correlated with lagged income. 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.[9]

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 in this paper; other strong assumptions are made as well.[10]

Antman and McKenzie (2007a, 2007b) take a different approach to 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 education of the head of the household.[11] 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 when they include cohort fixed effects, they should be controlling for that source of 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, not too far above the estimates in Lee (2009). 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.[12] There they are testing for poverty traps, which they do not find evidence

---

[9] 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.

[10] Glewwe discusses the issues raised by his use of BMI as an IV. He has to assume no reverse causality of BMI on 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.

[11] 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.

[12] Measurement error makes these nonlinear studies much more difficult because the measurement error

for.

McGarry (1995) in a study very close in spirit to this paper, simulates incomes and then poverty status and transitions, of widows in the US, including and excluding measurement error. The model that she uses to estimate 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 white noise. McGarry allows the white noise disturbance to contain random, time-varying measurement error, but assumes that the AR part of the error does not include any. There is only one white noise error allowed, which McGarry admits may also contain true white noise shocks. This complicates her analysis because then she is unable to identify each separately, unlike the model that we use. In her analysis, McGarry does not allow for time-invariant measurement error, for instance through the individual random effect, unlike what we do here. She examines poverty transitions based on the incomes simulated from the variance components of distribution of income that she estimates. By comparing simulated incomes with and without the variance component that includes measurement error, she concludes that the amount of permanent poverty of widows is underestimated by measurement error. In our case, we simulate expenditure, not income, based on our model without measurement error, and then compare that to measured expenditure data, not a second simulation. The measured expenditure data should include all types of measurement error.[13]

---

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 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.

[13] More recently, Luttmer (2002) and Villanger (2003) investigate the effect of measurement error on poverty transition probabilities. They construct their measurement error-free welfare measure (either income or expenditure) by simply subtracting simulated measurement error itself from survey data. However, in their studies, simulated measurement error is generated to be independent of the surveyed measure, and so

In sum, there are a relatively few studies that are serious in trying to correct for measurement error in estimating income or consumption dynamics. Outside of studies that have available administrative data to match to survey data, all of those using income, not expenditure data, most of these are only able to correct for random, time-varying error, not time-invariant error, and the former only under strong assumptions in many cases. Of these studies, only McGarry (1995) and Gibson and Glewwe (2005) attempt to construct poverty transition matrices based on model simulations with and without time-varying error.

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 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

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

Like most other studies, we assume that the measurement error η_{it} can be decomposed into

---

inevitably to be dependent on the true error-free expenditure or income. This is fundamentally wrong. Note that classical or random measurement error should be independent of true expenditure or income.

a time-invariant and a time-varying component[14]

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

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 the true consumption. We follow two approaches to deal with this problem. In the first approach we project the true initial observations 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 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 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.

---

[14] One scenario considered is that the time-constant measurement error is 0.

3.2. True Consumption

Lee (2009) specifies the following autoregressive model of consumption dynamics[15]

$$\ln C^*_{it} = \gamma \ln C^*_{i,t-1} + \beta' X_{it} + D_t + \alpha_i + \varepsilon_{it} \ , \qquad t \geq 2 \tag{4}$$

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),

$$\ln C_{it} = \gamma \ln C_{i,t-1} + \beta' X_{it} + D_t + \alpha_i + (1-\gamma)e_i + v_{it} - \gamma v_{i,t-1} + \varepsilon_{it}, \qquad t \geq 2. \tag{5}$$

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 to be serially uncorrelated[16].

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. This 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

---

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

[16] 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.

lagged three periods is valid. In addition she uses income satisfaction as an external instrument. The assumptions made by Lee (2009) are

$$E[\varepsilon_{it}|\ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i, Z_{i0}, \ldots, Z_{i,t-1}, \alpha_i, e_i] = 0. \tag{7}$$

and

$$E[v_{it}|\ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i, Z_{i0}, \ldots, Z_{i,t-1}, \alpha_i, e_i] = 0, \tag{8}$$

with X_i, Z_i the vectors of observations on the time-varying independent variables and the external instruments. 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 conditional on α_i and e_i. No assumptions on the conditional variances are needed for the consistent estimation of the regression parameters, 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.

$$\text{Var}(\varepsilon_{it}|\ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i) = \sigma_\varepsilon^2 \tag{9}$$

and

$$\text{Var}(v_{it}|\ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i) = \sigma_v^2. \tag{10}$$

Moreover we assume that the errors are conditionally uncorrelated

$$E(\varepsilon_{it} v_{it}|\ln C_{i0}, \ln C_{i1}, ..., \ln C_{i,t-1}, X_i) = 0. \tag{11}$$

Under these assumptions

$$E\left[(\Delta\tau_{it})^2\right] = 2\sigma_\epsilon^2 + 2(\gamma^2+\gamma+1)\sigma_v^2 \tag{12}$$

$$E[\Delta\tau_{it}\Delta\tau_{i,t-1}] = -\sigma_\epsilon^2 - (\gamma^2+2\gamma+1)\sigma_v^2 \tag{13}$$

$$E[\Delta\tau_{it}\Delta\tau_{i,t-2}] = \gamma\sigma_v^2. \tag{14}$$

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 (see Lee (2009) for more details). We can 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.

With the estimated regression parameters and the equation error variance we can simulate the first difference equation (4) if we add the assumption that the equation error ε_{it} has a normal distribution, i.e. we simulate

$$\Delta \ln C_{it}^* = \hat{\gamma}\Delta \ln C_{i,t-1}^* + \hat{\beta}'\Delta X_{it} + \Delta D_t + \Delta\varepsilon_{it} \qquad t \geq 2 \tag{15}$$

with

$$\varepsilon_{it} \sim N(0, \hat{\sigma}_\varepsilon^2). \tag{16}$$

In the simulation we ignore the sampling variation in the parameter estimates.

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

$$\ln C_{it}^* = \ln C_{i,t-1}^* + \Delta \ln C_{it}^*, \qquad t \geq 2. \tag{17}$$

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, (ii) no time-invariant measurement error and 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) 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}. Therefore the linear relations

$$\Delta \ln C^*_{i1} = \delta_0 + \beta_0 X_{i0} + \beta_1 X_{i1} + \zeta_{i0} \tag{18}$$

and

$$\ln C^*_{i1} = \delta_1 + \beta_2 X_{i0} + \beta_3 X_{i1} + \zeta_{i1}. \tag{19}$$

are linear projections of these relations on X_{i0}, X_{i1}. Here, ζ_{i0} and ζ_{i1} are the projection errors[17]. 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

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

and

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

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 quite strong assumption and if it fails the coefficients in (21) are biased but not those in (20).

We make two further assumptions on the projection errors that can be relaxed. We

---

[17] 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).

assume that the errors are homoskedastic and that they are normally distributed

$$\begin{pmatrix} \zeta_{i0} \\ \zeta_{i1} \end{pmatrix} \sim N(0, \Sigma) \tag{22}$$

where Σ is the variance-covariance matrix of the projection errors that has to be estimated. 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.

To estimate Σ we assume that the projection errors are uncorrelated with the time-varying measurement error. This assumption is in line with the assumptions that we made earlier on the relation between the equation error and measurement error. Under this assumption we can identify the variance of ζ_0 from the variance of the error in (20)

$$Var(\zeta_0) = Var(v_{i1} - v_{i0} + \zeta_{i0}) - 2\sigma_\nu^2 \tag{23}$$

The covariance of ζ_{i0} and $\Delta\varepsilon_{i2}$ is identified by

$$\mathrm{E}(\zeta_{i0}\Delta\varepsilon_{i2}) = (1 + 2\gamma)\sigma_\nu^2 + Cov(\Delta\tau_{i2}, \Delta v_{i1} + \zeta_{i0}) \tag{24}$$

where the covariance on the right hand side can be estimated from the projection residuals and the residuals of the first differenced consumption equation.[18]

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

---

[18] By the Cauchy-Schwartz inequality

$$Var(\zeta_{i0}) \geq \frac{(\mathrm{E}(\zeta_{i0}\Delta\varepsilon_{i2}))^2}{2\sigma_\varepsilon^2}$$

which is used as a check on the estimate of $Var(\zeta_0)$.

period log consumption by

$$\omega = Var(e_i + v_{i1} + \zeta_{i1}) = Var(e_i) + \sigma_v^2 + Var(\zeta_{i1}) \tag{25}$$

then this can be solved for $Var(e_i) + Var(\zeta_{i1})$. 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. In particular, if we denote

$$\kappa = Var(\alpha_i + (1-\gamma)e_i). \tag{26}$$

then

$$Var\left(\frac{1}{T}\sum_{t=1}^{T}\tau_{it}\right) = \kappa + \left(\frac{1}{T} + \frac{1}{T}\gamma^2 - 2\gamma\frac{T-1}{T^2}\right)\sigma_\nu^2 + \frac{1}{T}\sigma_\epsilon^2. \tag{27}$$

so that κ can be expressed in terms of estimable quantities.[19]

Therefore, we have two equations in the three unknowns $Var(\zeta_{i1}), Var(\alpha_i), Var(e_i)$

$$\omega - \sigma_\nu^2 = Var(\zeta_1) + Var(e_i) \tag{28}$$

and

$$\kappa = Var(\alpha_i) + (1-\gamma)^2 Var(e_i). \tag{29}$$

Because the variances are nonegative we obtain a lower and upper bound on $Var(\zeta_{i1})$. Before we derive the bound we note that the covariance of ζ_{i1} and $\Delta\varepsilon_{i2}$ is identified by

$$\mathrm{E}(\zeta_{i1}\Delta\varepsilon_{i2}) = (1+\gamma)\sigma_\nu^2 + Cov(\Delta\tau_{i2}, e_i + v_{i1} + \zeta_{i1}). \tag{30}$$

---

[19] A minor complication is that the data are not a balanced panel. The equation (27) 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.

By the Cauchy-Schwartz inequality this implies that

$$Var(\zeta_1) \geq \frac{(\mathrm{E}(\zeta_{i1}\Delta\varepsilon_{i2}))^2}{2\sigma_\varepsilon^2} \tag{31}$$

The lower and upper bound on $Var(\zeta_{i1})$ are obtained if we make the variance of the time-constant measurement error as small and as large as possible. We find that

$$\max\left\{0, \omega_1 - \sigma_\nu^2 - \frac{\kappa}{(1-\gamma)^2}, \frac{(\mathrm{E}(\zeta_{i1}\Delta\varepsilon_{i2}))^2}{2\sigma_\varepsilon^2}\right\} \leq Var(\zeta_{i1}) \leq \omega_1 - \sigma_\nu^2. \tag{32}$$

Finally the covariance of ζ_{i0} and ζ_{i1} is identified by

$$Cov(\zeta_0, \zeta_1) = cov(v_{i1} - v_{i0} + \zeta_{i0}, e_i + v_{i1} + \zeta_{i1}) - \sigma_\nu^2. \tag{33}$$

In the simulation we draw from the joint normal distribution of $\zeta_{i0}, \zeta_{i1}, \Delta\varepsilon_{i2}, \ldots, \Delta\varepsilon_{iT}$. For the point identified parameters we substitute the point estimates and for the interval identified $Var(\zeta_{i1})$ we take the estimates of the upper and lower bound. We are aware that this procedure ignores sampling variability. Although it is possible to incorporate sampling uncertainty in our simulations it is of smaller order than the population uncertainty in the random errors of the model.

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

$$\ln C_{i1}^* = \frac{\beta' X_{i1}}{1-\gamma L} + \frac{D_1}{1-\gamma L} + \frac{\alpha_i}{1-\gamma} + \phi_{i1} \tag{34}$$

with

$$\phi_{i1} = \frac{\varepsilon_{i1}}{1-\gamma L} \tag{35}$$

To use this equation in simulation we assume that

$$X_{it} = X_{i0} \ , \qquad t \leq 0 \tag{36}$$

$$D_t = 0 \ , \qquad t \leq 0 \tag{37}$$

These assumptions are in line with the assumption of stationarity of the 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 is 0. Note that this corresponds to the case of maximum uncertainty in the distribution of the initial observation in (32), 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

$$\alpha_i = E[\tau_{it}] \tag{38}$$

with τ_{it} the error of the log consumption equation. Therefore an obvious estimator is

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T} \hat{\tau}_{it} \tag{39}$$

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

$$Var(\phi_{i1}) = Var\left(\frac{\varepsilon_{i1}}{1-\gamma L}\right) = \frac{\sigma_\varepsilon^2}{1-\gamma^2}. \tag{40}$$

Because we have an estimate of the household effect α_i we can simulate the true log consumption in levels, so that we only need an initial distribution.

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. This means that we are not so much interested in the population true transition probabilities p_{jk}^* between consumption intervals. Therefore we draw a sample of the same size as our data sample from the true transition process. We then compare the observed transitions in this sample to those in our data sample that has measurement error in consumption. In section 5 several such comparisons are made. Any such comparison will be different if we draw other equation and projection errors from their assumed joint normal distribution. Because we average over the about 4,000 observations the simulation variability in the comparisons is small and we ignore it for now.

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.[20] Though there are fewer incentives to under-report survey consumption compared to income, substantial recall errors are assumed because of the lack of documented 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

---

[20] 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.

main summary statistics are reported in Table 1.

As explained in Section 3, The two-step GMM estimation of equation (5) 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. Equations (7) and (8) 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 shocks.

Sample Size

The estimation of equation (5) 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 A3. Total annual household expenditure, asked directly, is available from 1997 to 2005.[21] 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

---

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

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).[22]

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 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).[23] 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

---

[22] 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.

[23] The Ministry of Health and Welfare in Korea releases the MCL each year.

for three and four person households (household, not *percapita*) respectively before taking logarithms, and so 87.0 for a 3.43 person household.[24] 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 surveyed *percapita* expenditure (pce) in 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 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 (5), the autoregressive coefficient on log pce, is .472.[25] The variance of time-varying measurement error (.038) is almost as large as the

---

[24] This is about \$8 per day per capita in 2002.

[25] See Appendix Tables 4 and 5 for the other estimates of equations (5), (18) and (19).

variance of equation error (.043), which suggests that time-varying measurement error exists in reported 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 (20) and (21)), and the differenced equation error, $\Delta\varepsilon$, in equation (5) (see equations (24) and (30)). The projection errors, ζ_{i0} and ζ_{i1}, are also weakly correlated.

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, $Var(\zeta_0)$, is estimated as .103 and satisfies the Cauchy-Schwartz inequality so that it is consistent with the other estimates. The variance of the projection error for the level initial condition, $Var(\zeta_1)$ 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 4. $Var(\zeta_1)$ lies between .048 and .190. Note that when $Var(\zeta_1)$ is largest , the time-invariant measurement error variance is 0. $Var(\zeta_1)$ is smallest if the observed household effect is only due to time-constant measurement error. We can compare the bounds on $Var(\zeta_1)$ to the implied stationary equation error variance that is equal to .055 (Table 3). Note that in particular the upper bound on the projection error variance is much larger than the stationary equation error variance. This makes sense since the former is equal to the (conditional) cross-sectional variance and will include the variance of the household effect. Given this observation the lower bound on the projection error variance may be too low, but we will use it anyway as a comparison.

To impose stationarity, we need to estimate the household fixed effect α, in addition to the two-step GMM estimates of equation (5) and the estimate of the variance of ϕ from the MD estimation. The α_i is estimated from the residuals of the estimation of equation (5) (see equation (39)) for each household. The mean of α_i is 3.9 and its variance .264 (Table

3). The variance is just slightly higher than the variance of α_i under the highest projection error boundary in the projection methodology (Table 4).

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 Table 5. Based on the highest projection error, 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 is used. The density for the stationarity simulation shows a higher mean pce for 2001 and somewhat higher variance as well.

Simulated expenditures from 2002 to 2005 are sequentially constructed starting with simulated initial conditions and estimates of equation (5) and then using equations (15) and (17). Figure 2 displays the density functions of simulated consumption from 2002 to 2005 used to construct transition matrices and Table 5 shows their summary statistics. Simulated consumption using projection has comparable means to surveyed consumption data, but unlike consumption data, the variances increase little by little over time. 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 approaches the stationary one.[26] 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 means and 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 consumption has smaller variances than simulated consumption. One possible explanation is that surveyed consumption in reality consists not only of random measurement error but also of systematic measurement error. Simulated consumption, except for the initial consumption as a level, removes the time-invariant measurement error that partly captures systematic

---

[26] 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.

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.

Table 6 presents poverty headcount rates based on the minimum cost of living as a threshold, using expenditure data and simulated expenditure. Figure 3 describes headcount rates from the cumulative density function (CDF), or the poverty incidence curve. For this study measurement error results in underestimating headcount measures.[27] Here a comparison of the distribution of measured expenditure data with simulated data using projections of initial conditions shows again the smaller variance of the measured expenditure data. As noted, this is consistent with the Bound and Krueger observation of regression to the mean for US income self-reports. In our case, the smaller variance of the measured data together with the measured and simulated data having similar means, will tend to result in lower estimated poverty rates with the measured data, as we see here. In addition, however, the measured expenditure data show a decreasing trend in headcount rates, from 23.4% to 19.1% over time. However, the projection simulated expenditure estimates do not support this trend. Poverty rates calculated with the projection simulations stay at about 28% and 25% for Panels A1 and A2 respectively. If we were to decompose changes in poverty into growth and inequality effects (see Datt and Ravallion 1992), the rising variance of simulated expenditures would be equivalent to rising inequality, which we know raises poverty rates, ceterus paribus. While the growth effect of rising income lowers poverty, the net effect of both growth and inequality can be in either direction (Datt and Ravallion, 1992). Note that we control for growth in our simulations via time dummies used in the GMM estimation of equation (5). When we impose stationarity, simulated headcount start in 2001 almost the same as the simulation with the lowest projection error, but then decline slightly in 2002 and then more or less stay the same. Since the variance is not increasing over time when we

---

[27] In contrast, 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.

impose stationarity, there is no inequality effect tending to raise headcount estimates, unlike for the projection simulations.

5.2. Mobility

The main focus of this study is on poverty dynamics, or the movement into and out of poverty. Table 7 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. However, the simulated expenditure data shows a very different story.

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 24% are poor in only one or two years. Hence most of the poverty is transitory using these estimates. On the other hand, using our simulated, measurement error-free data based on the highest projection error, some 38% of households are estimated to be poor at least one of these four years, but nearly half of those, 18.5%, are poor each year and another 8% in three of the four years, while only 6.5% are poor in only one year and 12% poor in one or two years. Using the lowest projection errors, the results are quite similar, though the fraction in poverty in all years, is slightly lower, 15%. When we impose stationarity, a slightly lower fraction are ever in poverty, 33%, and of these 9.6% are in poverty all four years, while 10.1% are for three years. Thus, nearly 20% are in poverty 3 or 4 years, lower than the estimates from our projection simulations, but still considerably higher than the 12.1% estimate from the measured data (2/3 instead of 1/3 of those who are ever poor).

This study thus indicates that the number of 'always poor' households with these Korean data is downward biased by measurement error, and substantially so, which means that "chronic" poverty is understated when measurement error, both time-varying and time-invariant, is not corrected for. It is also worth noting that the number (or percentage) of households who never experience poverty is not noticeably affected by measurement error.

Tables 8 and 9 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.[28]

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 up to around 28 percentage points for the simulations using projections for initial conditions.[29] The consumption data with measurement error show that 44% of households initially in poverty at time $t-1$ move out of poverty at time t. However, when measurement error is removed through a model-based simulation using projections, the rate of moving out of poverty is only between 16 and 18%. The difference between Panels A1 and A2 based on the highest and lowest projection errors is relatively small. When we simulate imposing stationarity, mobility out of poverty is also distinctly lower than the measured data show, 28.6%, but higher than the simulations using projections.[30]

In Table 9, when we examine 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 large downward bias for the probability of staying in the bottom quintile.

On the other hand, conditional on being non-poor in the initial year, the odds of

---

[28] 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 A6 and A7. Major differences across the pairs of years are not observed.

[29] McGarry (1995) also suggests a downward bias in the probabilities that households stay at the bottom, but the difference (74-63=11%) is smaller. 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.

[30] Why the differences between the simulation methods will be examined in future work.

moving into poverty is low, around 7 to 8% and hardly differs when measurement error is accounted for, although the lowest estimates are for the simulation assuming stationarity, which is different from the pattern for exiting the bottom quintile. While the probability that households stay as non-poor does not seem to be much biased by measurement error in the 2×2 matrix, the probability that households stay at the top is also biased, downwards, by up to 18 percentage points, in the 5×5 matrix, when projection simulations are used and 7.5 percentage points when we impose stationarity. Being in the top quintile at time $t-1$ the odds of staying there are 85 - 88%, once measurement error is corrected using projections, and 76.5% 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.

Among the probabilities of immobility for households remaining 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 is the probability of staying at the top.

Finally, Table 10 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 one quintile for households at the bottom is twice as large using survey data than using simulated consumption by projection. The difference is much smaller compared to the simulation that imposes stationarity. The difference due to measurement error in the percentage of movements by two or more quintiles for those at the bottom is even larger. For all diagonals in the transition matrices based on consumption data, the probability of moving by one quintile is also overstated when measurement error is not corrected, though the bias appears to be smaller than for those who start at the bottom. Interestingly, we see that the highest likelihood of staying on the diagonals in the 5×5 transition matrix, when all diagonals are considered is when we impose stationarity. Here the difference between the measured data results and all of the simulations is very large.

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, linear projection, and by imposing stationarity and using the parameter estimates from the dynamic 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 account for 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 quite different results from the literature.

This study has shown that with the KLIPS data from Korea, measurement error substantially magnifies economic mobility into and out of poverty. The probability of remaining poor is strongly downward biased by measurement error, though the degree of the bias depends some on which model is used. The difference of these probabilities is at least 15 percentage points and may be as high as 30 percentage points. Even the lower estimate of the impact of measurement error shows a large effect. The number of years spent in poverty is also substantially 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 largely 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 more important than transitory poverty, suggesting that policies to lower poverty should be focused on factors like education and health that can permanently raise people out of poverty, more than on social safety nets. 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 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] Fields, Gary. 2006. "The Many Facets of Economic Mobility", in M.McGillivray (ed.), Inequality, Poverty and Well-Being. New York: Palgrave Macmillan.

[16] 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.

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

[18] 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.

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

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

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

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

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

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

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

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

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

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

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

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

[31] 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.

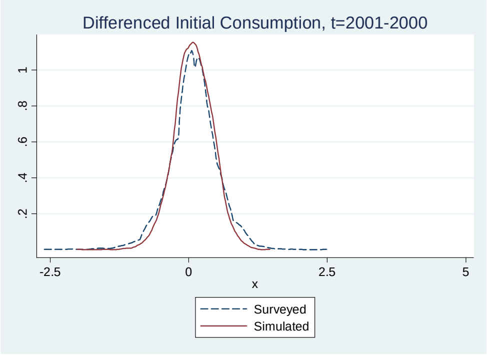
[32] Villanger, Espen. 2003. "The Effects of Disasters on Income Mobility: Bootstrap Inference and Measurement Simulations." Mimeo.

[33] 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 from Projection

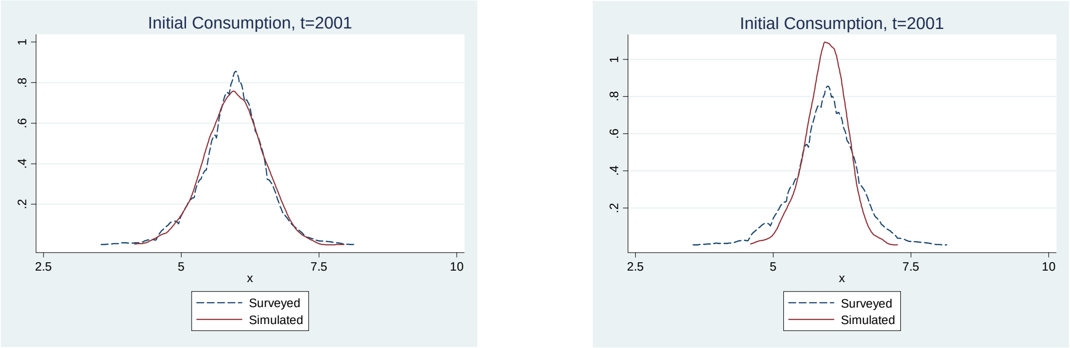
Panel A. Initial Observations by Projections

Differenced Initial Expenditure between 2001- 2000



Level Initial Expenditure in 2001

A1. Based on the highest projection error A2. Based on the lowest projection error



Panel B. Initial Observation from Stationary Distribution

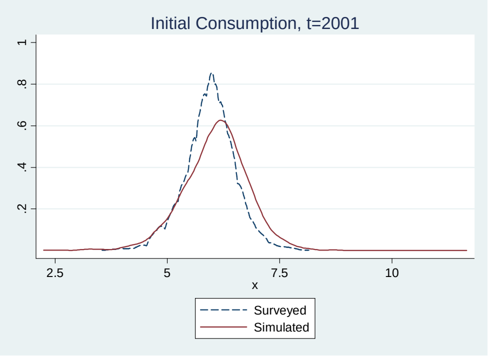
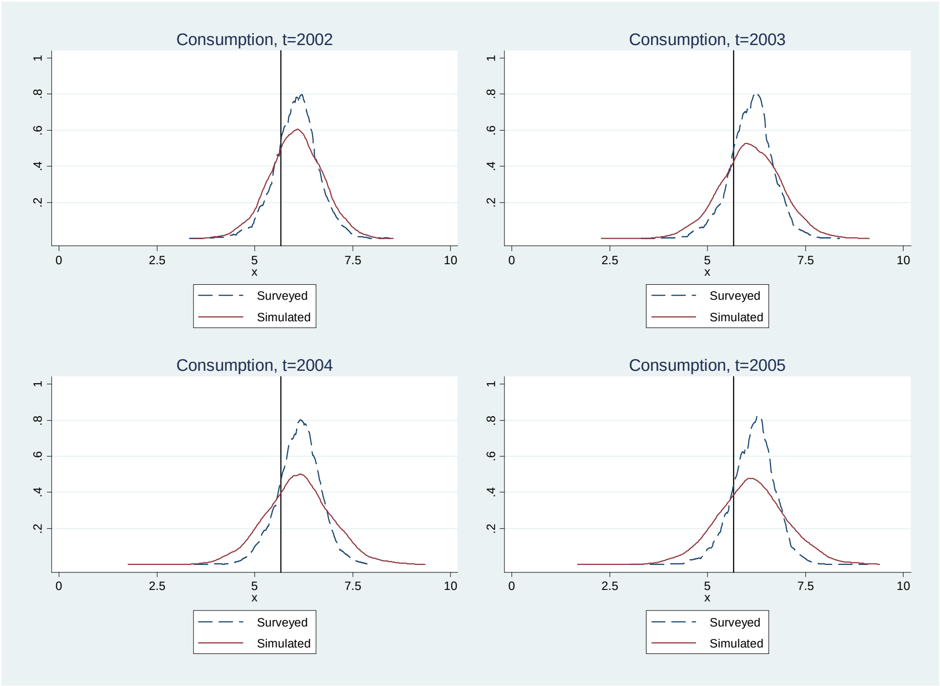


Figure 2: Density of Simulated Expenditure from 2002 to 2005

Panel A. Initial Observations by Projections

Panel A1. Based on the highest projection error



Panel A2. Based on the lowest projection error

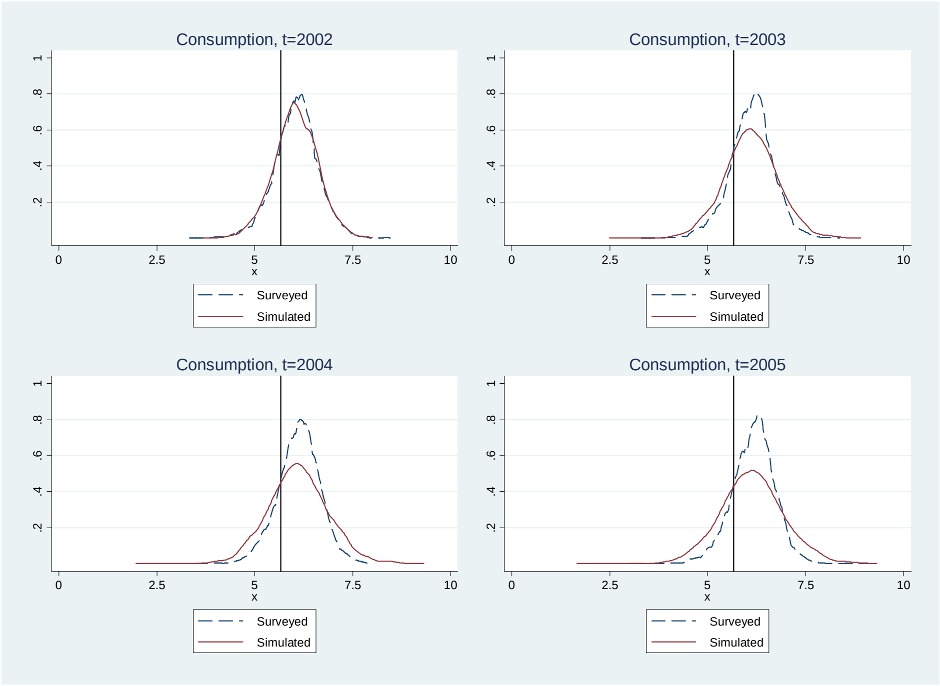
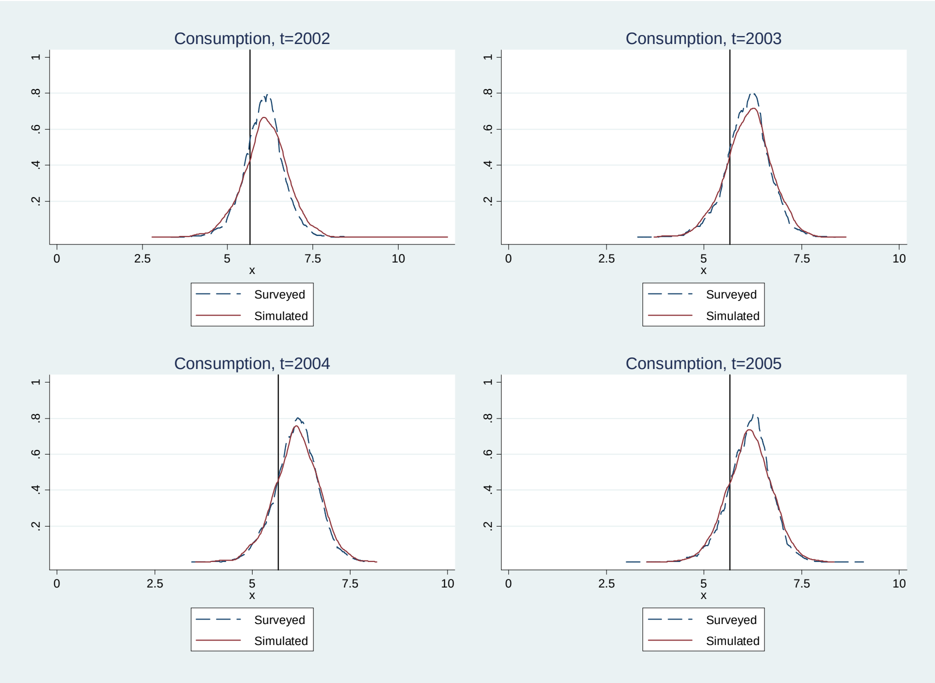


Figure 2: Continued

Panel B. Initial Observation from Stationary Distribution

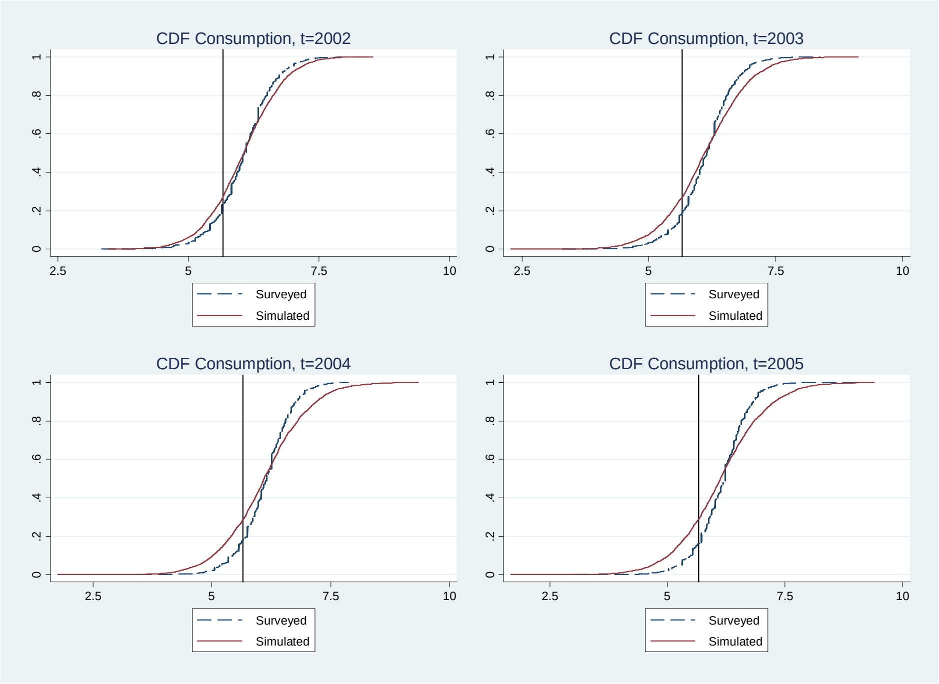


Vertical lines indicate the minimum cost of living (MCL).

Figure 3: Cumulative Density Function of Simulated Expenditure from 2002 to 2005

Panel A. Initial Observations by Projections

Panel A1. Based on the highest projection error



Panel A2. Based on the lowest projection error

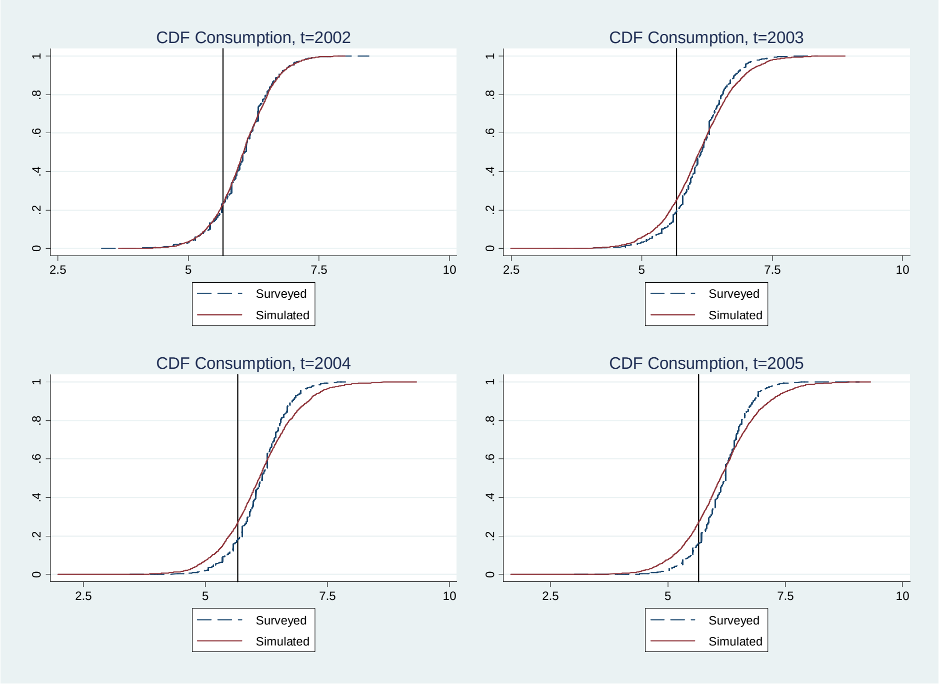
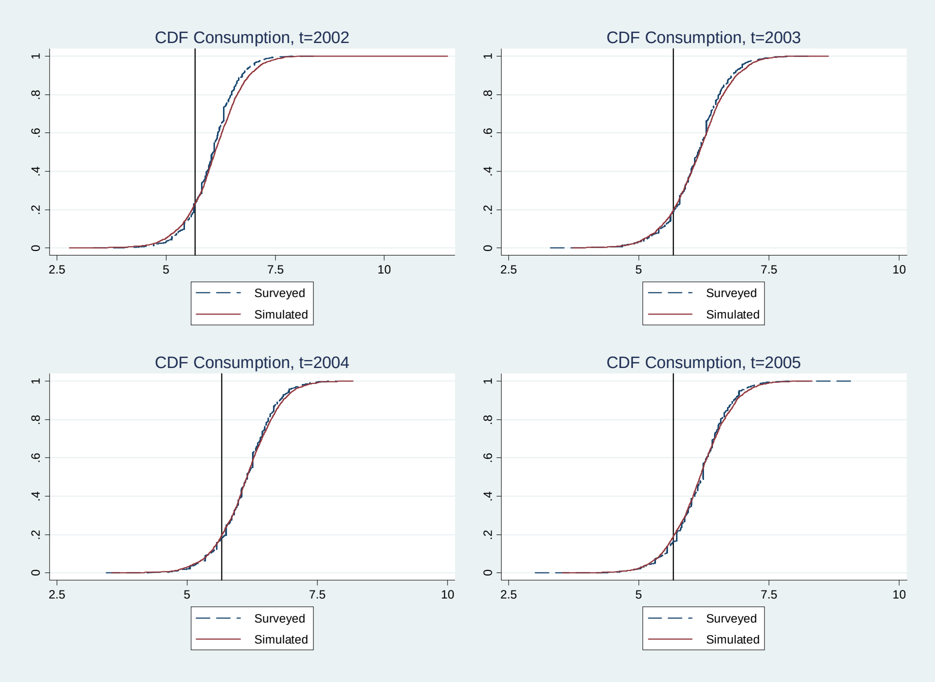


Figure 3: Continued

Panel B. Initial Observation from Stationary Distribution



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

Table 1: Summary Statistics - Mean and Standard Deviation

	Year					
Variable	2000	2001	2002	2003	2004	2005
Log (per capita expenditure)	5.87	5.96	6.06	6.11	6.13	6.16
	(0.56)	(0.57)	(0.56)	(0.56)	(0.53)	(0.54)
Household size	3.53	3.59	3.43	3.39	3.35	3.28
	(1.33)	(1.36)	(1.31)	(1.32)	(1.32)	(1.31)
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	0.20
	(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	10.22	10.27	10.17	10.17	10.15	10.16
	(4.43)	(4.41)	(4.45)	(4.45)	(4.43)	(4.44)
Seoul dummy	0.23	0.23	0.23	0.22	0.21	0.21
	(0.42)	(0.42)	(0.42)	(0.41)	(0.41)	(0.41)
Nonspouse dummy	0.19	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	49.55	50.17	51.43	52.86	53.82	54.76
	(12.88)	(12.85)	(12.83)	(12.78)	(12.67)	(12.55)
Obs #	3,062	3,230	3,073	2,955	2,855	2,788

Observations for households analyzed in this study
Standard deviations in parentheses

Table 2: Sample Size

	Estimation	Requirements	Simulation	Survey
Panel A	Differenced initial condition, 2001 - 2000	Household characteristics, X at 2000 and 2001	3,058	3,079
	Level initial condition in 2001	Household characteristics, X at 2000 and 2001	3,073	3,246
Panel B	Initial condition	Household characteristics, X at 2000	3,086	3,246
	Simulated expenditure in 2002	Differenced initial condition, 2001 - 2000 Level initial condition at 2001 Household characteristics, X at 2002 and 2001	3,073	3,362
	Simulated expenditure in 2003	Differenced simulated expenditure, 2002 - 2001 Simulated expenditure at 2002 Household characteristics, X at 2003 and 2002	2,995	3,248
	Simulated expenditure in 2004	Differenced simulated expenditure, 2003 - 2002 Simulated expenditure at 2003 Household characteristics, X at 2004 and 2003	2,855	3,151
	Simulated expenditure in 2005	Differenced simulated expenditure, 2004 - 2003 Simulated expenditure at 2004 Household characteristics, X at 2005 and 2004	2,778	3,111

The samples are restricted to the samples that are used to estimate the basic standard consumption dynamics model, equation (5).
For comparison between panel A and panel B, the samples size are further restricted to panel A.

Table 3: Basic Results

	Parameter	Estimate
	γ	.472
	σ_ϵ^2	.043
	σ_ν^2	.038
Panel A	$Var(v_{i1} - v_{i0} + \zeta_{i0})$	.180
	$Var(e_i + v_{i1} + \zeta_{i1}) = \omega$	.229
	$Var(\alpha_i + (1-\gamma)e_i) = \kappa$	.240
	$cov(\zeta_{i0}, \Delta\varepsilon_{i2})$	-.075
	$cov(\zeta_{i1}, \Delta\varepsilon_{i2})$	-.064
	$cov(\zeta_{i0}, \zeta_{i1})$	.053
Panel B	$Var(\phi_{i1})$	.055
	$\bar{\alpha}_i$	3.872
	$Var(\alpha_i)$	.264

Other estimates of equation (5) and their standard errors are in Appendix Table 1.
$\hat{\kappa}_1 = .259$, $\hat{\kappa}_2 = .211$, $\hat{\kappa}_3 = .235$, and $\hat{\kappa}_4 = .241$.

Table 4: Variances for Panel A

	A1. Upper bound	A2. Lower bound
$Var(\zeta_{i1})$	.190	.048
$Var(e_i)$	.000	.142
$Var(\alpha_i)$	.240	.200

Table 5: Statistics for Surveyed and Simulated Expenditure

Year	Obs #	Mean	Std. Dev.	Min	Max
(1) Observed expenditure					
2001-2000	3,058	.084	.441	-2.605	2.486
2001	3,073	5.954	.567	3.543	8.149
2002	3,073	6.057	.558	3.334	8.458
2003	2,955	6.114	.556	3.230	8.375
2004	2,855	6.125	.533	3.446	7.869
2005	2,778	6.163	.540	3.014	9.078
(2) Simulated true expenditure					
Panel A. Initial observations by projection					
Panel A1. Based on the highest projection error					
2001-2000	3,058	.085	.345	-2.017	1.462
2001	3,073	5.956	.532	4.153	7.958
2002	3,073	6.052	.676	3.457	8.532
2003	2,955	6.105	.784	2.280	9.121
2004	2,855	6.109	.856	1.766	9.345
2005	2,778	6.137	.899	1.671	9.392
Panel A2. Based on the lowest projection error					
2001-2000	3,058	.085	.345	-2.017	1.462
2001	3,073	5.955	.377	4.589	7.253
2002	3,073	6.051	.561	3.666	8.002
2003	2,955	6.105	.689	2.488	8.902
2004	2,855	6.108	.768	1.974	9.321
2005	2,778	6.136	.816	1.659	9.316
Panel B. Initial observation from stationary distribution					
2001	3,073	6.078	.733	2.249	11.662
2002	3,073	6.109	.663	2.787	11.454
2003	2,955	6.140	.596	3.721	8.638
2004	2,855	6.136	.573	3.548	8.187
2005	2,778	6.163	.578	3.528	8.323

Table 6: Poverty Rates from 2002 to 2005

Poor (below MCL)	2002	2003	2004	2005	Avg.
(1) Observed expenditure					
Obs #	719	553	511	447	558
%	23.4	18.71	17.90	16.09	19.12
(2) Simulated true expenditure					
Panel A. Initial observations by projection					
Panel A1. Based on the highest projection error					
Obs #	827	804	820	797	812
%	26.9	27.2	28.7	28.7	27.9
Panel A2. Based on the lowest projection error					
Obs #	706	735	765	752	740
%	23.0	24.9	26.8	27.1	25.4
Panel B. Initial observation from stationary distribution					
Obs #	710	576	550	525	590
%	23.10	19.49	19.26	18.90	20.2

Table 7: 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
%	63.47	14.62	9.82	6.03	6.06
(2) Simulated true expenditure					
Panel A. Initial observations by projection					
Panel A1. Based on the highest projection error					
Obs #	1,701	180	149	229	511
%	61.41	6.50	5.38	8.27	18.45
Panel A2. Based on the lowest projection error					
Obs #	1,744	189	189	231	417
%	62.96	6.82	6.82	8.34	15.05
Panel B. Initial observation from stationary distribution					
Obs #	1,861	164	200	279	266
%	67.18	5.92	7.22	10.07	9.60

The number of samples which exist in 2005 (and so all over the years) are 2,770.
% of HHs, for a certain number of years in which poor, is calculated based on these samples.

Table 8: 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	56.47	43.53
(2) Simulated expenditure, Panel A1	84.08	15.92
(2)' Simulated expenditure, Panel A2	82.42	17.58
(2)" Simulated expenditure, Panel B	71.36	28.64
Not in poverty		
(1) Observed expenditure	7.80	92.20
(2) Simulated expenditure, Panel A1	6.83	93.17
(2)' Simulated expenditure, Panel A2	7.58	92.42
(2)" Simulated expenditure, Panel B	5.62	94.37

Row percentages are presented.
Panel A1, A2 and B are based on the highest and lowest projection errors, and first observation from the stationary distribution, respectively.
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 9: 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	58.12	24.40	9.64	6.21	1.62
	(2) Simulated expenditure, Panel A1	84.22	11.91	3.04	0.61	.22
	(2)' Simulated expenditure, Panel A2	82.05	14.53	2.40	0.83	.20
	(2)" Simulated expenditure, Panel B	71.04	20.14	6.91	1.73	.18
2	(1) Observed expenditure	21.11	34.57	24.26	15.68	4.38
	(2) Simulated expenditure, Panel A1	26.93	40.59	22.69	8.50	1.29
	(2)' Simulated expenditure, Panel A2	26.37	41.10	23.43	8.03	1.07
	(2)" Simulated expenditure, Panel B	20.83	40.44	27.89	9.56	1.28
3	(1) Observed expenditure	7.83	18.74	33.16	30.78	9.49
	(2) Simulated expenditure, Panel A1	4.02	24.69	36.26	27.32	7.71
	(2)' Simulated expenditure, Panel A2	5.13	23.51	36.62	27.05	7.70
	(2)" Simulated expenditure, Panel B	4.25	23.08	38.55	28.15	5.96
4	(1) Observed expenditure	3.66	8.48	20.24	42.18	25.44
	(2) Simulated expenditure, Panel A1	.61	6.07	20.35	41.53	31.44
	(2)' Simulated expenditure, Panel A2	.69	6.05	19.38	39.79	34.09
	(2)" Simulated expenditure, Panel B	.51	6.90	23.46	44.63	24.49
5	(1) Observed expenditure	1.20	2.28	5.21	22.27	69.04
	(2)' Simulated expenditure, Panel A1	0	.33	1.65	10.23	87.79
	(2)' Simulated expenditure, Panel A2	.05	.28	2.33	12.22	85.12
	(2)" Simulated expenditure, Panel B	.23	.60	3.32	19.31	76.54

Row percentages are presented
Panel A1, A2 and B are based on the highest and lowest projection errors, and first observation from the stationary distribution, respectively.
Averaged probability for all years (2002-2005)
Expenditure classes are based on the 2002 quintile (1: poorest).

Table 10: Summary of Transition Matrices

	Percent of households that: Remain on diagonal	Move by one quintile	Move by two or more quintiles	Percent of households in bottom that: Remain in bottom	Move by one quintile	Move by two or more quintiles
(1) Observed expenditure						
2×2	85.02	14.98		56.47	43.53	
5×5	47.60	37.52	14.87	58.12	24.40	17.48
(2) Simulated true expenditure						
Panel A. Initial observations by projection						
Panel A1. Simulated expenditure with highest projection error for initial condition						
2×2	90.65	9.35		84.08	15.92	
5×5	65.16	29.05	5.79	84.22	11.91	3.87
Panel A2. Simulated expenditure with lowest projection error for initial condition						
2×2	89.93	10.07		82.42	17.58	
5×5	60.88	32.75	6.36	82.05	14.53	3.42
Panel B. Simulated expenditure with first observation from the stationary distribution						
2×2	89.63	10.37		71.36	28.64	
5×5	59.07	35.99	4.94	71.04	20.14	8.82

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)
- Other covariates at t	3,097	3,169	3,196	3,319	12,781
	(38)	(22)	(24)	(18)	(102)
- Other covariates at t-1	3,082	3,150	3,192	3,312	12,736
	(15)	(19)	(4)	(7)	(45)
- HH income satisfaction at t-1	3,056	3,125	3,161	3,242	12,584
	(26)	(25)	(31)	(70)	(152)
- HH income satisfaction at t-2	3,034	3,105	3,124	3,230	12,493
	(22)	(20)	(37)	(12)	(91)
- HH income satisfaction at t-3	2,939	2,918	2,983	2,996	11,836
	(95)	(187)	(141)	(234)	(657)
- Outliers	2,938	2,917	2,982	2,995	11,832
	(1)	(1)	(1)	(1)	(4)

Marginal loss of observations in parenthesis

Table A2: Household Expenditure Variables in KLIPS

	1998		1999		2000		2001		2002		2003		2004		2005	
		%		%		%		%		%		%		%		%
Total expenditure in a single question	454.3		452.9		505.4		555.6		498.1		527.0		527.0		550.8	
Total expenditure constructed	442.6		n/a		481.9		544.1		497.1		525.5		529.0		556.3	
1. Food eating in	157.3	35.5			142.7	29.6	134.0	24.6	116.1	23.3	114.8	21.8	108.4	20.5	106.1	19.1
2. Food eating out	14.2	3.2			17.9	3.7	21.2	3.9	18.8	3.8	17.0	3.2	16.4	3.1	17.5	3.1
3. Public education	42.4	9.6			44.3	9.2	39.3	7.2	30.3	6.1	27.4	5.2	32.9	6.2	34.1	6.1
4. Private education	37.9	8.6			43.8	9.1	47.3	8.7	40.8	8.2	36.5	6.9	34.8	6.6	37.7	6.8
5. Automobile maintenance	39.4	8.9			50.9	10.6	51.2	9.4	43.6	8.8	44.3	8.4	43.0	8.1	44.5	8.0
6. Housing maintenance	54.8	12.4			50.1	10.4	54.7	10.1	48.8	9.8	50.3	9.6	54.9	10.4	57.5	10.3
7. Expense for others	27.1	6.1			24.1	5.0	25.2	4.6	24.3	4.9	22.1	4.2	20.6	3.9	20.9	3.8
8. Health	23.0	5.2			21.6	4.5	21.8	4.0	23.3	4.7	22.9	4.4	22.9	4.3	26.2	4.7
9. Recreation	10.4	2.3			15.8	3.3	11.5	2.1	11.1	2.2	11.1	2.1	11.4	2.1	11.4	2.1
10. Durable goods	9.9	2.2			6.8	1.4	6.7	1.2	6.6	1.3	5.1	1.0	5.8	1.1	4.9	0.9
11. Others	26.3	6.0			28.0	5.8	14.4	2.6	12.3	2.5	11.0	2.1	5.1	1.0	0.5	0.1
12. Communication					36.1	7.5	39.6	7.3	35.0	7.0	35.5	6.8	34.2	6.5	34.4	6.2
13. Pocket money							77.1	14.2	69.5	14.0	61.6	11.7	67.1	12.7		
14. Clothing									16.6	3.3	15.2	2.9	16.3	3.1	17.3	3.1
15. Charity											9.2	1.7	8.5	1.6	9.2	1.7
16. Social security/ medical insurance											26.4	5.0	30.9	5.8	33.1	6.0
17. Public transportation											15.2	2.9	15.8	3.0	16.0	2.9
18. Pocket money (parents)															1.7	0.3
19. Pocket money (children)															28.4	5.1
20. Pocket money (others)															39.7	7.1
21. Basic necessities															15.1	2.7
Total expenditure excluding durables	432.7		n/a		432.7		537.3		490.5		520.4		523.2		551.4	

Annual per capita expenditure for observations analyzed

Table A3: 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 A4: Two-step GMM Estimates after First Differencing
Pooled over years, t=2002, 2003, 2004 and 2005

Dependent variable: △Log expenditure t	Coefficient
△Log expenditure $t-1$	0.4723***
	(0.072)
△Household size	-0.3038***
	(0.015)
△Male aged over 65	0.0559
	(0.099)
△Female aged over 55	0.1661**
	(0.072)
△Sex of head	0.0244
	(0.081)
△Education of head	0.0153
	(0.014)
△Seoul dummy	-0.0331
	(0.058)
△Nonspouse dummy	0.0945*
	(0.052)
△Age of head	0.0067
	(0.014)
△Square age of head	-0.0001
	(0.000)
△Year dummy (2003)	-0.0014
	(0.015)
△Year dummy (2004)	-0.0386**
	(0.019)
△Year dummy (2005)	-0.0317
	(0.022)
Hansen J statistics	10.82
(p value)	0.459
N	11,832

All covariates are first differenced (denoted by △)
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%, * significant at 10%,

Table A5: OLS estimates for Initial Conditions

Dependent variable:	Dependent Variable	
	ΔLog expenditure at 2001	Log expenditure at 2001
Household size at 2001	-0.3160***	-0.1906***
	(0.026)	(0.029)
Male aged over 65 at 2001	-0.0615	-0.045
	(0.161)	(0.182)
Female aged over 55 at 2001	0.1476	-0.137
	(0.139)	(0.157)
Sex of head at 2001	0.0213	-0.2056**
	(0.092)	(0.104)
Education of head at 2001	0.0087	0.0260**
	(0.010)	(0.011)
Seoul dummy at 2001	-0.0295	-0.0184
	(0.073)	(0.082)
Nonspouse dummy at 2001	-0.0148	-0.0667
	(0.075)	(0.085)
Age of head at 2001	-0.0169	0.0299*
	(0.014)	(0.016)
Square age of head at 2001	0.0002*	-0.0003
	(0.000)	(0.000)
Household size at 2000	0.3068***	0.0339
	(0.026)	(0.029)
Male aged over 65 at 2000	-0.1418	-0.5181***
	(0.162)	(0.182)
Female aged over 55 at 2000	-0.2351*	-0.1134
	(0.137)	(0.155)
Sex of head at 2000	0.0249	0.1228
	(0.093)	(0.105)
Education of head at 2000	-0.0093	0.0236**
	(0.010)	(0.011)
Seoul dummy at 2000	0.0089	0.1213
	(0.073)	(0.082)
Nonspouse dummy at 2000	-0.0393	0.0389
	(0.076)	(0.086)
Age of head at 2000	0.007	0.0038
	(0.014)	(0.016)
Square age of head at 2000	-0.0002	-0.0001
	(0.000)	(0.000)
Constant	0.4128***	5.3233***
	(0.135)	(0.152)
R-squared	0.073	0.288
N	3,058	3,073

Table A6: 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, Panel A1	80.73	19.27
(2)' Simulated expenditure, Panel A2	80.09	19.91
(2)" Simulated expenditure, Panel B	68.39	31.61
Non in poverty		
(1) Observed expenditure	7.96	92.04
(2) Simulated expenditure, Panel A1	7.56	92.44
(2)' Simulated expenditure, Panel A2	8.45	91.55
(2)" Simulated expenditure, Panel B	4.88	95.12
Investigated Years: from 2003 to 2004		
Poverty		
(1) Observed expenditure	58.99	41.01
(2) Simulated expenditure, Panel A1	86.76	13.24
(2)' Simulated expenditure, Panel A2	83.73	16.27
(2)" Simulated expenditure, Panel B	74.02	25.98
Non in poverty		
(1) Observed expenditure	8.45	91.55
(2) Simulated expenditure, Panel A1	6.99	93.01
(2)' Simulated expenditure, Panel A2	7.85	92.15
(2)" Simulated expenditure, Panel B	5.76	94.24
Investigated Years: from 2004 to 2005		
Poverty		
(1) Observed expenditure	57.63	42.37
(2) Simulated expenditure, Panel A1	84.79	15.21
(2)' Simulated expenditure, Panel A2	83.29	16.71
(2)" Simulated expenditure, Panel B	72.32	27.68
Non in poverty		
(1) Observed expenditure	6.98	93.02
(2) Simulated expenditure, Panel A1	5.88	94.12
(2)' Simulated expenditure, Panel A2	6.31	93.69
(2)" Simulated expenditure, Panel B	6.24	93.76

Row percentages are presented.

Panel A1, A2 and B are based on the highest and lowest projection errors, and first observation from the stationary distribution, respectively.

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 A7: 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, Panel A1	81.43	12.73	4.55	.91	.39
	(2)' Simulated expenditure, Panel A2	79.44	16.51	2.65	.93	.47
	(2)" Simulated expenditure, PanelB	69.06	22.19	7.5	1.25	0
2	(1) Observed expenditure	22.52	32.28	23.78	15.12	6.30
	(2) Simulated expenditure, Panel A1	26.27	40.98	21.96	9.22	1.57
	(2)' Simulated expenditure, Panel A2	26.72	38.32	25.38	8.40	1.18
	(2)" Simulated expenditure, Panel B	16.88	44.44	25.85	11.32	1.50
3	(1) Observed expenditure	7.46	21.56	31.67	27.03	12.27
	(2) Simulated expenditure, Panel A1	2.74	23.58	34.32	28.21	11.16
	(2)' Simulated expenditure, Panel A2	4.41	21.34	36.86	27.16	10.23
	(2)" Simulated expenditure, Panel B	4.05	21.58	37.76	31.21	5.40
4	(1) Observed expenditure	4.49	8.64	19.86	41.45	25.56
	(2) Simulated expenditure, Panel A1	1.06	4.67	17.80	38.77	37.71
	(2)' Simulated expenditure, Panel A2	.54	5.54	15.36	38.75	39.82
	(2)" Simulated expenditure, Panel B	.35	7.42	21.73	46.82	23.67
5	(1) Observed expenditure	1.53	3.74	5.60	23.09	66.04
	(2)' Simulated expenditure, Panel A1	0	0	.69	9.81	89.50
	(2)' Simulated expenditure, Panel A2	.17	.34	2.39	12.44	84.67
	(2)" Simulated expenditure, Panel B	.40	.13	2.90	21.64	74.93

Table A7: Continued

		Class at 2004				
Class at 2003		1	2	3	4	5
1	(1) Observed expenditure	59.92	22.96	8.95	7.20	.97
	(2) Simulated expenditure, Panel A1	86.85	10.76	2.12	.27	0
	(2)' Simulated expenditure, Panel A2	83.92	12.83	2.21	.88	.15
	(2)" Simulated expenditure, PanelB	72.90	17.94	7.29	1.87	0
2	(1) Observed expenditure	22.08	34.72	25.47	13.78	3.96
	(2) Simulated expenditure, Panel A1	27.31	40.05	23.61	7.41	1.62
	(2)' Simulated expenditure, Panel A2	29.36	41.06	21.70	7.23	.64
	(2)" Simulated expenditure, Panel B	22.04	40.54	29.11	7.28	1.04
3	(1) Observed expenditure	8.46	17.46	35.48	30.88	7.72
	(2) Simulated expenditure, Panel A1	5.74	27.42	33.94	27.42	5.48
	(2)' Simulated expenditure, Panel A2	5.81	27.74	37.20	23.66	5.59
	(2)" Simulated expenditure, Panel B	3.59	24.55	40.12	26.35	5.39
4	(1) Observed expenditure	3.12	9.50	22.12	40.97	24.30
	(2) Simulated expenditure, Panel A1	.23	7.21	23.72	43.72	25.12
	(2)' Simulated expenditure, Panel A2	.84	7.13	24.11	37.95	29.98
	(2)" Simulated expenditure, Panel B	.79	6.84	25.60	43.08	23.69
5	(1) Observed expenditure	1.28	1.93	5.62	22.79	68.38
	(2)' Simulated expenditure, Panel A1	0	.47	2.34	11.23	85.96
	(2)' Simulated expenditure, Panel A2	0	.13	2.23	12.98	84.67
	(2)" Simulated expenditure, Panel B	.14	.57	4.24	17.82	77.23

Table A7: Continued

Class at 2004		Class at 2005 1	2	3	4	5
1	(1) Observed expenditure	58.23	24.28	8.02	6.58	2.88
	(2) Simulated expenditure, Panel A1	84.45	12.21	2.44	.64	.26
	(2)' Simulated expenditure, Panel A2	82.60	14.36	2.35	.69	0
	(2)" Simulated expenditure, PanelB	71.57	19.88	5.77	2.19	.60
2	(1) Observed expenditure	18.02	37.58	23.52	18.68	2.20
	(2) Simulated expenditure, Panel A1	27.39	40.69	22.61	8.78	.53
	(2)' Simulated expenditure, Panel A2	22.61	44.99	22.61	8.39	1.40
	(2)" Simulated expenditure, Panel B	23.62	36.20	28.70	10.15	1.32
3	(1) Observed expenditure	7.60	16.88	32.47	34.88	8.16
	(2) Simulated expenditure, Panel A1	3.88	23.27	41.27	26.04	5.54
	(2)' Simulated expenditure, Panel A2	5.37	21.70	35.61	30.73	6.59
	(2)" Simulated expenditure, Panel B	5.03	23.16	37.88	26.93	7.00
4	(1) Observed expenditure	3.45	7.36	18.77	43.99	26.43
	(2) Simulated expenditure, Panel A1	.48	6.51	19.76	42.41	30.84
	(2)' Simulated expenditure, Panel A2	.72	5.50	19.38	43.30	31.10
	(2)" Simulated expenditure, Panel B	.36	6.46	22.80	44.17	26.21
5	(1) Observed expenditure	.79	1.27	4.45	20.99	72.50
	(2)' Simulated expenditure, Panel A1	0	.47	1.78	9.59	88.17
	(2)' Simulated expenditure, Panel A2	0	.38	2.39	11.34	85.89
	(2)" Simulated expenditure, Panel B	.14	1.13	2.84	18.30	77.59

Row percentages are presented
Panel A1, A2 and B are based on the highest and lowest projection errors, and first observation from the stationary distribution, respectively.
Averaged probability for all years (2002-2005)
Expenditure classes are based on the 2002 quintile (1: poorest).