This paper is a draft submission to

Inequality—Measurement, trends, impacts, and policies

5–6 September 2014 Helsinki, Finland

UNU-WIDER Conference:
Inequality – Measurement, Trends, Impacts, and Policies
5-6 September 2014, Helsinki, Finland

Poverty and Inequality Dynamics

Ira N. Gang, Rutgers University (gang@economics.rutgers.edu)
Ksenia Gatskova, IOS–Regensburg (gatskova@ios-regensburg.de)
John Landon-Lane, Rutgers University (lane@economics.rutgers.edu)
Myeong-Su Yun, Tulane University, (msyun@tulane.edu)

August 2014

Abstract: We offer a rigorous methodology for examining the dynamics of income inequality and poverty. Our strategy is to estimate a transition probability matrix with the aim of identifying the vulnerability of households to poverty and the sources of inequality change. We can then estimate the limiting stationary income distribution enabling us to assess the probability of moving into poverty and allowing us to define characteristics that make households vulnerable to poverty. Also, we estimate the marginal effect of correlates on poverty and those vulnerable to poverty. Similarly for inequality; so we can estimate the marginal effect these correlates on the current income distribution, the mobility of income and the limiting income distribution. As well, we can calculate the standard measures of poverty and inequality using the initial income distribution and the limiting stationary income distribution. (135 words)

Keywords: mobility measurement, vulnerability, poverty, inequality, measurement, Tajikistan
JEL classification: J60, D63, I32

Acknowledgements: We thank Ilhom Abdulloev and Melanie Khamis for long conversations on this material, and IOS-Regensburg for help with the data.

1 Introduction

"…We are challenged to rid our nation and the world of poverty. Like a monstrous octopus, poverty spreads its nagging, prehensile tentacles into hamlets and villages all over our world. Two-thirds of the people of the world go to bed hungry tonight. They are ill-housed; they are ill-nourished; they are shabbily clad. I've seen it in Latin America; I've seen it in Africa; I've seen this poverty in Asia."

– Martin Luther King Jr, Speech on Poverty at the National Cathedral, March 31, 1968.

This paper is part of our effort to employ rigorous econometrics in establishing a link between formal policy analysis and the implications of mobility for poverty and inequality changes, providing insight into the existence, size and role of vulnerability to poverty in the economy. Our strategy is to estimate a Markov transition probability matrix with the aim of identifying the vulnerability of households to poverty and the sources of inequality change. Importantly we can use the estimated dynamics to calculate the limiting stationary income distribution. By doing so we are able to assess the short and long-run probability of moving into poverty from other income categories which allows us to define characteristics that make households and individuals vulnerable to poverty. We are also able to estimate the marginal effect of correlates on poverty and those vulnerable to poverty. Similarly for inequality we are able to include correlates in the estimation procedure and so are able to estimate the marginal effect these correlates have on the current income distribution, the mobility of income and the limiting income distribution. Thus we are also able to estimate the marginal effects of various correlates on measures of inequality. As well, we can calculate the standard measures of poverty and inequality using the initial income distribution and the limiting stationary income distribution.

There is a long line of research on estimating the income (or social) mobility of individuals (or countries/ states) using discrete-state Markov chains going back to the work of Champernowne (1953) and Prais (1955). These two papers apply a discrete-state Markov chain to the problem of labor-income mobility and social mobility, respectively. This paper follows the long line of research that was initiated by these seminal papers.

Conceptually, throughout the paper we intertwine and advance the study of mobility and vulnerability. Shorrocks (1978) introduced a mobility index that summarizes the mobility information contained in the probability transition matrix. Gang, Landon-Lane, and Yun (2004) show that where a natural ordering of states exists, such as in the case of modeling income mobility, one can decompose the Shorrocks' mobility measure into measures of upward and downward mobility, or in fact into any part of the transition matrix. Here we go further and show how we can relate standard multivariate analysis to subparts of the transition matrix in such a way that our results are consistent with the mobility measure at various levels of decomposition. This way we can include variables that capture correlates of mobility, poverty change and inequality.

For vulnerability our contribution is along distinct though related lines. Our data is organized in the same way as for mobility – an initial distribution of income, estimated Markov transition probability matrices and invariant income distributions. However, the aggregation rule for vulnerability is different from that for mobility, and with vulnerability we also must identify the poor. The data organization here allows us to examine changes in vulnerability over time, as well as its correlates.

We apply our method to the data from Tajikistan. Tajikistan is the world's most remittance dependent country (more than half of 2012 GDP) with 37% of the labor force working abroad. Internally, 50% of nonagricultural employment is informal and in 2006 the shadow economy reached 60.9% of GDP. We use the 2007 and 2009 Tajikistan Living Standards Measurement Survey and the 2011 Tajikistan Household Panel Survey to construct a panel for the three years. This three year panel allows us to look at the dynamics of poverty in two distinct transitions. The first is from 2007 to 2009 which coincides with the impact that the global financial crisis had on Tajikistan. The second transition from 2009 to 2011 coincides with Tajikistan's recovery from the global financial crisis. Thus we are able to use our methodology in a recession period and an expansion period. This allows us to show how our method is able to easily characterize the dynamics of poverty in both recession and recovery.

The outline of the paper is as follows: Section 2 introduces the underlying Markovian model used in this paper while Section 3 introduces the methodology used. Included in Section 3 is the measure of vulnerability to poverty that we introduce. In Section 4 we describe the estimation strategy and in Section 5 we report the results of our work using data from Tajikistan. Finally in Section 6 we conclude.

2 Modelling Household Expenditure as a Markov Process

The dynamics of poverty is studied in this paper using a first order discrete state Markov chain for household expenditure. While we use reported household expenditure to model the dynamics of poverty, the use of Markov-chain models to study income dynamics has a long history with notable contributions by Champernowne (1953) and Shorrocks (1976). Methods similar to the ones used in this paper have been used in Gang, Landon-Lane, and Yun (2002, 2009), Dimova, Gang, and Landon-Lane (2006), and Co, Landon-Lane, and Yun (2009).

One of the most appealing aspects of using a Markov-chain to model household expenditure dynamics is the ability to investigate issues such as short and long-run movements into and out of poverty. The Markov assumption is a natural way of thinking about household expenditure dynamics while imposing only minimal theoretical structure. Before elaborating on how we investigate movements into and out of poverty we briefly discuss the first order discrete state Markov model. A fuller discussion of this model can be found in Geweke (2005).

The model is as follows: Let there be a total of C expenditure classifications where C is a finite number. The researcher is free to define the expenditure classifications as they see fit. Champernowne (1953), for example, suggests using (for income) classifications that are equal in log length. The classifications should cover the expenditure distribution and in most cases should be defined so that all observations are not contained in any one classification. In our example it is natural to define the expenditure classifications based on the "poverty line". For example, let e^{ep} be the extreme poverty line and let e^{p} be the poverty line. Then it makes sense to define the lowest expenditure classification to be

$$\mathcal{C}_1 = (-\infty, e^{ep}].$$

The next expenditure classification would be

$$\mathcal{C}_2 = (e^{ep}, e^{p}].$$

Thus the first two expenditure classifications would include those households below the extreme poverty line and poverty line respectively (by convention someone exactly on the line is in the higher group). After defining the first two expenditure classifications we are free to define the rest of the expenditure classifications as we see fit. A natural approach would be to define the upper bounds of the expenditure classifications as multiples of the poverty line e^p. An example of a set of expenditure classifications is

$$
\begin{aligned}
\mathcal{C}_1 &= (-\infty, e^{xp}] \\
\mathcal{C}_2 &= (e^{xp}, e^p] \\
\mathcal{C}_3 &= (e^p, 1.5e^p] \\
\mathcal{C}_4 &= (1.5e^p, 2e^p] \\
&\vdots \\
\mathcal{C}_C &= (me^p, \infty).
\end{aligned}
\tag{1}
$$

The value of m and therefore C is chosen so that the expenditure classifications cover most of the expenditure distribution. The highest expenditure classification should be chosen so that a small proportion of the total population is covered. It would not make sense for the highest expenditure classification to be defined to contain all of the population nor would it make sense for the highest expenditure classification to include none of the population.

Once the expenditure classifications are defined it is possible to model the dynamics of the (discrete) expenditure distribution. Let π_{kt} be the proportion of the population households who have expenditures that fall into classification $\mathcal{C}_k$ in period t. Another way to think of π_{kt} is as the probability a randomly chosen household is a member of classification $\mathcal{C}_k$. That is

$$\pi_{kt} = Pr(h \in \mathcal{C}_k). \tag{2}$$

Let $\pi_t = (\pi_{1t}, \ldots, \pi_{Ct})'$ be the (column) vector of probabilities for each of the expenditure classifications at time t. Therefore the variable π_t defines the ``state" of the world at time t in terms of the (cross-sectional) expenditure distribution. The only structure that is imposed on π_t is the first order Markov assumption. This assumption implies that the state of the world today is only dependent on π_{t-1} and not on its past history beyond the most recent time period. That is,

$$P(\pi_t \mid \pi_{t-1}, \pi_{t-2}, \ldots, \pi_{t-j}) = P(\pi_t \mid \pi_{t-1}) \quad \forall \quad j = 2,3,\ldots, \tag{3}$$

where $P(.)$ represents the conditional (cross-sectional) probability distribution of π. This first-order Markov assumption was introduced in Champernowne (1953) and was further discussed in Shorrocks (1976).

The first order assumption is made operational by defining the *Markov transition matrix* $\mathbf{P}$. Define the probability of transiting from classification i in period $t-1$ to classification j in period t to be $P(\pi_t = j \mid \pi_{t-1} = i) \equiv p_{ij}$ so that the Markov transition matrix, $\mathbf{P}$, can be defined as $\mathbf{P} = [\,p_{ij}\,]$. Then the first order discrete-state Markov chain model can be written as

$$\pi_t' = \pi_{t-1}'\mathbf{P}. \tag{4}$$

At this point it will be useful to show the usefulness of the Markov model described in (4) using a simple illustrative example. Consider breaking the expenditure into three classifications, $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$. Further suppose that members of these three classifications could be thought of as low expenditure households, middle expenditure households, and high expenditure households respectively. The Markov transition matrix in this case would be

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}.$$

The first row of $\mathbf{P}$ reports the probabilities of a low expenditure household staying a low expenditure household (p_{11}), moving up to be a middle expenditure household (p_{12}), and moving up to be a high expenditure household (p_{13}) respectively. Because this exhausts all possibilities for the movements of low expenditure households between classifications the sum of the first row of $\mathbf{P}$ must equal 1. The same is true for the second and third rows of $\mathbf{P}$ as well. Thus each row of $\mathbf{P}$ sums to 1 which means that $\mathbf{P}$ is a row stochastic matrix. This property that the rows of $\mathbf{P}$ must sum to 1 is exploited later when we estimate the model.

There are many elements or functions of $\mathbf{P}$ that are interesting to us. The diagonal elements of $\mathbf{P}$ reflect the respective probabilities of staying in an expenditure classification. Thus $(1-p_{ii})$ is the probability of moving out of expenditure classification i. The function

$$\frac{1}{1-p_{ii}}$$

is the expected length of stay in expenditure classification i, and the probabilities p_{21} and p_{31} reflect the probability of falling back to the low expenditure classification from $\mathcal{C}_2$ and $\mathcal{C}_3$ respectively.

Information obtained from $\mathbf{P}$ is not the only important information we can get from the model in (4). We are also able to extract information about the dynamics of the cross-sectional distribution. Let the initial expenditure probability distribution be π_0. Then by (4) it follows that

$$\pi_1' = \pi_0'\mathbf{P},$$

and

$$\pi_2' = \pi_1'\mathbf{P} = \pi_0'\mathbf{PP} = \pi_0'\mathbf{P}^2.$$

Thus it is simple to show that

$$\pi_t' = \pi_0'\mathbf{P}^t.$$

The $m-period$ transition probability matrix is given by $\mathbf{P}^m$. The invariant or limiting income distribution, $\bar{\pi}$, is any distribution that satisfies

$$\bar{\pi}' = \bar{\pi}'\mathbf{P}, \tag{5}$$

and is equivalent to

$$\bar{\pi}' = lim_{t\to\infty}\pi_0'\mathbf{P}^t.$$

The invariant distribution is unique if there is only one eigenvalue of $\mathbf{P}$ with modulus one.[1] Thus we can characterize both the short-run dynamics, via $\mathbf{P}$, and the long-run dynamics of the expenditure distribution via $\bar{\pi}$. Note here that $\bar{\pi}$ is a non-linear function of $\mathbf{P}$ as it is the left eigenvector of $\mathbf{P}$ associated with the eigenvalue of $\mathbf{P}$ equal to 1. We therefore need to estimate the parameters of the model given in (4) as well estimate non-linear functions of those parameters.

3 Measuring Poverty and Income Dynamics and Vulnerability

There is a long history of using Markov transition matrices to measure income inequality and income mobility. Original work by Champernowne (1953) and Prais (1955) looked at income and social mobility respectively using Markov models. In the 1970's and 1980's this effort at using Markov probability matrices to measure income mobility was furthered by the work of Shorrocks (1976, 1978a, 1978b) and by Geweke, Marshall and Zarkin (1986a, 1986b).

3.1 Measuring Mobility

A highly cited and commonly used measure of mobility is the mobility measure due to Shorrocks (1978). This measure is defined as

$$\mathcal{M}_S(\mathbf{P}) = \frac{N - Tr(\mathbf{P})}{N-1} \tag{6}$$

This measure of mobility is related to the average length of stay in any one class, $\mathcal{C}_i$, which Prais (1955) showed to be

[1] Implicitly we are assuming that the eigenvalues have been ordered from highest to lowest in terms of magnitude. As $\mathbf{P}$ is row stochastic we know that the highest eigenvalue, in terms of magnitude, is 1. If the magnitude of the second eigenvalue is strictly less than 1 then we know that the invariant distribution is unique (Geweke, Marshall, and Zarkin (1986b)).

$$\mathcal{M}_{pi} = \frac{1}{1 - p_{ii}}. \tag{7}$$

In particular, Shorrocks' measure of mobility is the inverse of the harmonic mean of $\mathcal{M}_{pi}$ scaled by the ratio $\frac{N}{N-1}$.

Geweke (1986a) describes the various properties that a mobility measure can have. Consider the following criteria.

Criteria M: Monotonicity

A mobility measure, $\mathcal{M}(\mathbf{P})$, satisfies the *monotonicity* criterion if $\mathcal{M}(\mathbf{P}) > \mathcal{M}(\mathbf{P}^*)$ where $p_{ij} \geq p_{ij}^*$ for all $i \neq j$, and $p_{ij} > p_{ij}^*$ for some $i \neq j$.

Criteria I: Immobility

A mobility measure, $\mathcal{M}(\mathbf{P})$, satisfies the *immobility* criterion if $\mathcal{M}(\mathbf{P}) \geq 0$.

Criteria SI: Strict Immobility

A mobility measure, $\mathcal{M}(\mathbf{P})$, satisfies the *strict immobility* criterion if $\mathcal{M}(\mathbf{P}) > 0$ unless $\mathbf{P} = \mathbf{I}$.

These three criteria are grouped together by Geweke (1986a) as ***persistence criteria***. The Shorrocks measure, (6), satisfies all three criteria and so is internally consistent with respect to the persistence criteria.

Gang, Landon-Lane, and Yun (2004) show that when the C classifications in the Markov model can be ordered then directional mobility measures can be defined. Gang et. al. (2004) define the upward and downward mobility measures

$$\mathcal{M}_U(\mathbf{P}) = \frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} p_{ij}}{N-1}, \tag{8}$$

and

$$\mathcal{M}_D(\mathbf{P}) = \frac{\sum_{i=2}^{N}\sum_{j=1}^{i-1} p_{ij}}{N-1}. \tag{9}$$

Gang et. al. (2004) show that these directional mobility measures are internally consistent with a modified persistence criteria and moreover are related to the Shorrocks' mobility measure as

$$\mathcal{M}_s(\mathbf{P}) = \mathcal{M}_D(\mathbf{P}) + \mathcal{M}_U(\mathbf{P}). \tag{10}$$

The directional mobility measures allow one to discriminate between different movements within the income or expenditure distribution.

Consider the following example. For illustration purposes, suppose that there are only two states. We will later extend the analysis to cases with more than two states. An individual can be in one of two classifications, $\mathcal{C}_1$ or $\mathcal{C}_2$, with $\mathcal{C}_2$ being "better" than $\mathcal{C}_1$. Think of $\mathcal{C}_1$ as poor and $\mathcal{C}_2$ as non-poor. Note in practice we would use a finer set of states. Let the Markov transition matrices P_1 and P_2 be defined as follows:

$$P_1 = \begin{bmatrix} 0.6 & 0.4 \\ 0.7 & 0.3 \end{bmatrix} \qquad P_2 = \begin{bmatrix} 0.3 & 0.7 \\ 0.4 & 0.6 \end{bmatrix}.$$

Here, P_1 and P_2 could refer to transition probability matrices for two countries or transition probability matrices for one country in two separate periods. The Shorrocks' measure is identical for both matrices, but the matrices convey vastly different information about poverty persistence and inequality. In the first transition matrix, the probability of moving into the poverty state from the non-poor state is high while the probability of staying in the poverty state is relatively high. The opposite is true for the second transition matrix. The inferences one can make about the income dynamics is also different for the two transition matrices. The limiting steady state distribution for P_1 is 63.6% in the poor category and 36.4% in the non-poor category. For P_2 we have in the limit 26.4% in the poor category and 63.6% in the non-poor category.

3.2 Measuring Inequality

Inequality tells us about the dispersion of a distribution, how scattered are the observations. There are many measures of inequality and a generally agreed upon set of properties that it is desirable for these measures to possess (Litchfield, 1999). Here we examine the cross-sectional income or expenditure distribution π_t which is a discrete representation of the continuous distribution underlying the population. Measures of inequality are functions of π_t. Co, Landon-Lane and Yun (2006) used Markov models to test for convergence of cross-sectional distributions. This work used, and we will initially use here, the standard deviation of the cross-sectional distribution, π_t, as a measure of the concentration of the cross-sectional distribution. That is define $\sigma(\pi)$ as

$$\sigma(\pi) = \sqrt{Var(\pi)} = \sqrt{\sum_{k=1}^{C} \left(k - E(\pi)\right)^2 \pi_k}, \tag{11}$$

where

$$E(\pi) = \sum_{k=1}^{C} k\pi_k. \tag{12}$$

Rising inequality in household income or expenditures would result in the "build up" of probability in the extremes of the distribution. This would lead to an increase in the variance of the cross-sectional income distribution. Measuring inequality here amounts to measuring whether the variance of the cross-sectional distribution is increasing over time. A natural way to do this is to check whether the standard deviation of the limiting income distribution is less concentrated than the initial distribution. Under the Markov assumption the cross-sectional distribution is operated on period by period by the Markov transition matrix $\mathbf{P}$.

Using (11) we find evidence of rising inequality if

$$\sigma(\bar{\pi}) > \sigma(\pi_0). \tag{13}$$

Thus, a function of interest that is useful for testing for rising inequality is

$$g(\pi_0, P) = \sigma(\bar{\pi}) - \sigma(\pi_0), \tag{14}$$

where $\bar{\pi}$ is the limiting distribution of π_t as $t \to \infty$.

3.3 Measuring Vulnerability

The Markov model is also an appropriate model to answer questions about vulnerability to poverty. Here vulnerability is the situation where a household is above the poverty line but there is concern that it may fall back into poverty. There is a large literature and no consensus as to what constitutes vulnerability. Dutta, Foster and Mishra (2011) in developing their formal measure provide a nice, terse summary of the discussion. As a starting point we can intuitively think of an income/expenditure less than twice the poverty line.

Consider the following example. Suppose there are five classifications with

$$
\begin{aligned}
\mathcal{C}_1 &= (-\infty, e^p] \\
\mathcal{C}_2 &= (e^p, 1.5e^p] \\
\mathcal{C}_3 &= (1.5e^p, 2e^p]\,. \\
\mathcal{C}_4 &= (2e^p, 2.5e^p] \\
\mathcal{C}_5 &= (2.5e^p, 3e^p]
\end{aligned}
\tag{15}
$$

Here the first classification includes all households whose income or expenditure is below the poverty line whilst the second and third classifications include households with incomes or expenditures above the poverty line but below twice the poverty line. These households would be considered to be vulnerable to poverty. Using the notation developed earlier then π_{1t} would be the proportion of the population who are below the poverty line while $\pi_{2t} + \pi_{3t}$ would be the proportion of the population that are vulnerable.

Because of the dynamic nature of the Markov model we are able to estimate the Markov transition matrix and hence estimate the limiting cross-sectional distribution, $\bar{\pi}$. Thus the proportion of the population that will be in poverty in the limit would be $\bar{\pi}_1$ and the proportion of the population that would be vulnerable to poverty in the limit would be $\bar{\pi}_2 + \bar{\pi}_3$. We could therefore detect if the proportions of the population in poverty or vulnerable to poverty was increasing or decreasing.

Another statistic of interest is a measure of vulnerability that can be used to compare two different populations. Apart from the cross-sectional distribution, π_t, we also estimate the Markov transition matrix, $\mathbf{P}$. For this example the Markov transition matrix looks like

$$
\mathbf{P} = \left[p_{ij} \right]_{i=1,\ldots 5,\, j=1,\ldots,5}, \tag{16}
$$

where p_{k1} is the probability that a household that starts in classification k in period $t-1$ falls into classification 1 (i.e. poverty) in period t. Consider the following measure which is the weighted average of the probabilities of falling back into poverty in one period from a classification above poverty,

$$\frac{\pi_{2t-1}p_{21}+\ldots+\pi_{5t-1}p_{51}}{\pi_{2t-1}+\ldots+\pi_{5t-1}}. \tag{17}$$

This measure is the weighted probability of moving back into poverty in one period. This measure is the instantaneous or short run measure of vulnerability for this population. Note that the numerator of (17) is just the contribution of π_{1t}, the proportion of the population in poverty in period t, from those households that were above the poverty line in period $t-1$ but fell back into poverty in period t.

What about multi-period vulnerability? According to the Markov assumption

$$\pi'_{t+2}=\pi'_t\mathbf{P}^2. \tag{18}$$

Let

$$\mathbf{P2}=[\,p2_{ij}\,]_{i=1,\ldots,C,\,j=1,\ldots,C}\equiv\mathbf{P}^2. \tag{19}$$

Then $\mathbf{P2}$ is also row stochastic and $p2_{ij}$ is the probability that a household that is in classification i in period t moves to classification j in period $t+2$. Thus a two period measure of vulnerability for this example would be

$$\frac{\pi_{2t-2}p2_{21}+\ldots+\pi_{5t-2}p2_{51}}{\pi_{2t-2}+\ldots+\pi_{5t-2}}. \tag{20}$$

This measure is the weighted probability that a household will fall back into poverty after two periods. Clearly such a measure can be defined for any period thus allowing us to measure short and long-run vulnerability.

This motivating example allows us to now formally define our measure of vulnerability. Suppose that there are C classifications and that these classifications are ordered so that classifications $\mathcal{C}_1,\ldots,\mathcal{C}_p$ are classifications in which households are below the poverty line and classifications $\mathcal{C}_{p+1},\ldots,\mathcal{C}_C$ are above the poverty line. Define the k-period measure of vulnerability as

$$V^k(\mathbf{P}) = \frac{\sum_{m=p+1}^{C} \pi_{mt} \left(\sum_{j=1}^{p} p^k{}_{mj} \right)}{\sum_{m=p+1}^{C} \pi_{mt}}, \tag{21}$$

where p_{ij}^k is the *ij*-th element of the matrix $\mathbf{P}^k$.

This measure of vulnerability is the weighted probability that a household will fall into a classification that is below the poverty line from a classification above the poverty line in within k periods. This measure is non-negative and is only equal to 0 if all the probabilities of moving into poverty are 0. Also this measure satisfies the monotonicity criteria so this measure, like the Shorrocks mobility measure, is consistent with the persistence criteria. Deaton (1997, p.122) suggests something similar in discussing the proper setting of a poverty line – establish the poverty line high enough so that everyone is poor, but the further income is from the poverty line (the greater the income shortfall) the greater the weight a la Sen's (1976) suggestion of using ordinal ranked weights

Finally, one last important consequence of using the Markov model is that we are able to endogenously determine the classifications that make up the vulnerable classification. By endogenous here we mean that the class of households that can be considered to be vulnerable to poverty can be determined by the resulting estimates rather than by some preordained criteria. We could easily define a household to be k-period vulnerable if the household is a member of a classification that has more than 20% probability of falling into poverty after k-periods. This motivates the following definition.

Definition *k*-period Vulnerability

Suppose that there are C classifications and that these classifications are ordered so that classifications $\mathcal{C}_1,\ldots,\mathcal{C}_p$ are classifications in which households are below the poverty line and classifications $\mathcal{C}_{p+1},\ldots,\mathcal{C}_C$ are above the poverty line. Suppose further that the underlying Markov model has a Markov transition matrix equal to $\mathbf{P}$. Households in classification *j* are *k*-period vulnerable of size $0 < q < 1$ if

$$\sum_{m=1}^{p} p^{k}{}_{jm} > q,$$

where $p^{k}{}_{jm}$ is the jm-th element of $\mathbf{P}^k$.

4 Exact Inference in Markov Model

To estimate this we use Bayesian methods and make inferences from the Markov chain model outlined above. One important consequence of using Bayesian methods is that it is simple to characterize the distribution of any function of the primal parameters, π_0 and $\mathbf{P}$, of the model. For example, we are able to characterize the distribution of the invariant distribution, $\bar{\pi}$; a highly non-linear function of $\mathbf{P}$. Other functions of $\mathbf{P}$ that are of interest include various mobility indices and poverty risk measures that we define later.

We observe N individuals over T time periods and place them into C classifications. Let $i \in \{1, 2, \ldots, C\}$, $n \in \{1, 2, \ldots, N\}$, and let $t \in \{1, 2, \ldots, T\}$. For each individual, n, define

$$\delta_{nit} = \begin{cases} 1 & \text{if individual } n \text{ is in classification } i \text{ for time period } t \\ 0 & \text{else} \end{cases}.$$

For each individual, n, and for each time period t we observe the individuals' classification $s_{nt} \in \{1, 2, 3, \ldots, C\}$. Therefore $S_{NT} = \{\{s_{nt}\}_{n=1}^{N}\}_{t=1}^{T}$ is the relevant information set at time T. Define $k_{j0} = \sum_{n=1}^{N} \delta_{nj0}$ as the number of individuals that are in classification j in the initial period and define $k_{ij} = \sum_{n=1}^{N} \sum_{t=1}^{T} \delta_{ni(t-1)} \delta_{njt}$ as the total number of transitions from classification i in time period $t-1$ to class j in time period t across all time periods. The matrix $\mathbf{K} = [k_{ij}]$ will be referred to as the data transition matrix. Note that if T is greater than two it is implicitly assumed that $\mathbf{P}$ is the same for all $T-1$ transition periods.

The data density, or likelihood function, for the model defined in (4) is

$$p(S_{NT} \mid \pi_0, \mathbf{P}) \propto \prod_{i=1}^{C} \pi_{i0}^{k_{i0}} \prod_{j=1}^{C} p_{ij}^{k_{ij}} \tag{22}$$

which is the kernel of the product of two independent multivariate Dirichlet (Beta) distributions. Natural conjugate priors for π_0 and $\mathbf{P}$ are also independent Dirichlet distributions defined as

$$p(\pi_0) = \left[\frac{\Gamma(\sum_{i=1}^{C} a_{i0})}{\prod_{i=1}^{C} \Gamma(a_{i0})} \right] \prod_{i=1}^{C} \pi_{i0}^{(a_{i0}-1)}$$

and

$$p(\mathbf{P}) = \prod_{i=1}^{C} \left[\frac{\Gamma(\sum_{j=1}^{C} a_{ij})}{\prod_{j=1}^{C} \Gamma(a_{ij})} \right] \prod_{j=1}^{C} \pi_{ij}^{(a_{ij}-1)}.$$

Here the priors are parameterized by the vector $a_0 = (a_{10}, \ldots, a_{C0})'$ and $\mathbf{A} = [a_{ij}]$. The priors have a notional sample interpretation. We can think of $a_{i0} - 1$ as the number of households in the i^{th} class of the initial expenditure distribution of a notional sample and $a_{ij} - 1$ can be interpreted as the total number of transitions from classification i in period $t-1$ to classification j in period t for the notional sample. Assuming that the priors are independent then the posterior distribution for (4) is

$$p(\pi_0, \mathbf{P} \mid S_{NT}) \propto \left[\frac{\Gamma(\sum_{i=1}^{C} a_{i0})}{\prod_{i=1}^{C} \Gamma(a_{i0})} \right] \prod_{i=1}^{C} \pi_{i0}^{(k_{i0}+a_{i0}-1)} \prod_{i=1}^{C} \left\{ \left[\frac{\Gamma(\sum_{j=1}^{C} a_{ij})}{\prod_{j=1}^{C} \Gamma(a_{ij})} \right] \prod_{j=1}^{C} \pi_{ij}^{(k_{ij}+a_{ij}-1)} \right\}. \tag{23}$$

The joint posterior density kernel in (23) is the kernel for the product of two Dirichlet distributions. The posterior distribution for π_0, the initial income distribution, is Dirichlet with parameters $(k_{10}+a_{10},\ldots,k_{C0}+a_{C0})'$. The posterior distribution for $\mathbf{P}$ is the product of C independent Dirichlet distributions with parameters $(k_{i1}+a_{i1},\ldots,k_{iC}+a_{iC})'$ for $i=1,\ldots,C$ (Geweke, 2005).

We are interested in calculating conditional expectations, $E(g(\pi_0,\mathbf{P})\,|\,S_{NT})$, for any well-defined function, $g(\pi_0,\mathbf{P})$, of the parameters of the Markov chain. Examples include the invariant income distribution and measures of poverty risk, which was defined in (21). It is a simple matter to make identical and independent draws from these independent Dirichlet distributions using the method described in Devroye (1986). Let $\pi_0^M=\{\pi_0^{(m)}\}_{m=1}^M$, and $\mathbf{P^M}=\{\mathbf{P^{(m)}}\}_{\mathbf{m=1}}^{\mathbf{M}}$ be the i.i.d. samples from the posterior distribution, $p(\pi_0,\mathbf{P}\,|\,S_{NT})$. By the law of large numbers, we can approximate this conditional expectation by

$$\overline{g}_M = M^{-1}\sum_{m=1}^{M} g(\pi_0^{(m)},\mathbf{P}^{(m)}) \text{™}\ E(g(\pi_0,\mathbf{P}\,|\,S_{NT})).$$

Thus we are able to obtain posterior means and standard deviations of any function of interest that is based on the initial distribution, π_0, or the Markov transition matrix $\mathbf{P}$.

5 Measuring Poverty in Tajikistan: 2007-2011

Tajikistan, a poor Central Asian economy and former Soviet Republic underwent severe economic, social and political changes following its separation from the USSR. Independence in 1991 with its rupture of economic ties was followed by civil war among rival regional clans from 1992 to 1997 and then an initially tenuous peace. By the end of the war GDP had shrunk to 35% of its 1990 level and inflation was at 65.2% (World Bank, 2011).

New economic policies were implemented soon after the peace accord and formation of the joint government in 1997. Over the 2001-2010 period annual real GDP grew at an 8.8% average rate; average annual inflation was 20.7% (World Bank, 2013a). Despite these positive achievements, Tajikistan remains economically far behind other countries of the former USSR with the highest poverty rate and lowest GDP per capita. GDP per capita was US$820 in 2010 (for comparison, in the Russian Federation – US$10,481); poverty by the headcount ratio was 46.7% in 2009 (World Bank, 2013a). Average monthly wages were US$82.90 in 2010; about 8.5 times lower than those of the Russian Federation (Statistical Committee of CIS, 2011). In traditional sectors of economy – agriculture, forestry and fisheries, which together employ 50% of Tajikistan's working population, monthly wages were US$23.60, $39.10 and $41.60, respectively (Statistical Agency of Tajikistan, 2011).

With large income and wage differentials between Tajikistan and other former Soviet countries came significant emigration of one-fifth of its working population. In turn, Tajikistan became one of the world’s most remittance-dependent countries: remittances reached 52 percent of GDP in 2012 (World Bank, 2013b). Data from the 2007 World Bank Living Standard Measurement Survey on Tajikistan (TLSS 2007) show that most (95.3%) of its migrants go to Russia, are predominantly men (93.5%) from rural areas (76.4% of all migrants), are ethnically Tajiks (81.4% of all migrants), and have only secondary education (64.36% with no university or other post-secondary school training). 74.2% of Tajikistan's migrants remitted in cash only, 1% remitted in-kind only, and 6.6% remitted both in-kind and in cash. The share of households having no migrant (in the respective year) decreased from 85.2% in 2007 to 60.9% in 2011 (Danzer, Dietz, Gatskova, 2013a). The socio-demographic characteristics of migrants did not change much: Tajikistan Household Panel Survey (2011) shows that nearly all migrants go to Russia, almost two thirds chose Moscow. The proportion of women among migrants slightly increased (up to 10.6%). The average age was 31.6 years for those who returned back home and 28.9 years for those who were still living abroad at the time of survey in 2011. 99% of the returned migrants remitted money home, while among those still living abroad 78% remitted money.

Tajikistan provides a good case for studying mobility, poverty dynamics and inequality change. Since independence in 1991, large numbers of Tajiks have emigrated and returned. This has made Tajikistan the subject of several useful data collection efforts. We use the 2007 and 2009 TLSS and the 2011 Tajikistan Household Panel Survey (Danzer, Dietz and Gatskova, 2013b,). The three years of panel data permit analysis of two transitions. These surveys ask questions on migration, education, health, labor market, housing, transfer and social assistance, subjective poverty and food security, as well as detailed data for household's expenditure and income.

Using household level income and expenditure data, total income is computed as including total receipts from employment, net transfers from government agencies, remittances from household members living away from home, the market value of assets consumed (e.g. livestock, vegetables etc.), and the market value of labor services rendered for which payment was in kind. Total expenditure for a household includes total payments for good and services consumed, the market value of goods and services consumed where payment was made in kind, the market value of assets consumed, and the value of savings (or asset accumulation). All income and expenditure variables are converted to monthly equivalent for each household in our estimations. As in Abdulloev, Gang and Landon-Lane (2012) we impute informal sector activity for households by comparing reported income and reported expenditure. Households with reported expenditures more than twice reported income are assigned to informal sector activity.

The data on monthly household expenditures are used to determine whether a household is below or above the poverty line. Monthly expenditures are in current prices (local currency) and are reported as a total for the household. According to the World Bank (2009) in its report on poverty in Tajikistan in 2007 the per person poverty line, based on expenditures, is 139 Somoni. Using reported gross national expenditure indices for Tajikistan for 2009 and 2011 the poverty line for 2009 was computed to be 169 Somoni and for 2011 was computed to be 214 Somoni.[2] The variable of direct interest to us is then calculated as the relative household expenditure per member of the household relative to the relevant poverty line. This is the variable that we use in the estimation of the Markov transition matrix later in this section. In using expenditure and not income we are keeping with the standard approach in the development literature based partly on theory – income is more volatile than consumption and so the latter is a better indicator of

[2] Gross national expenditure index obtained from http://www.indexmundi.com/facts/tajikistan/gross-national-expenditure-deflator.

lifetime welfare (Deaton, 1997) – and partly on the practical matter that income is more difficult to measure than income (see the classic discussion on this matter in Deaton, 1997).

In this analysis we use a balanced panel of 1257 households. The panel was chosen so that each household had a reported value in each year for all variables of interest. Table 1 reports summary statistics and descriptions for variables of interest in our sample. The period of our sample straddles the global financial crisis and the data on income and expenditures reflect this. Monthly expenditures fall from 2007 to 2009 and then greatly increase in 2011 as the economy bounces back. Incomes rise during this time. Our measure of informal activity shows that about 73% of the households partake in informal sector activity in 2007 and this falls to 59% in 2009. However this increases back to 79% in 2011. We assign informal sector activity to those households who have reported expenditures greater than twice the reported incomes. The fall in the proportion of households being assigned to informal sector activity is consistent with the fall in expenditures. There is no clear explanation for this drop in informal sector activity during the international crisis period.

Consistent with the narrative above we do see an increasing number of households containing migrants. We also see an increase in the proportion of households receiving remittance income and we see an increase in the size of the remittances. The rest of the variables that we look at are demographic variables that we use to characterize the household. Roughly a third of our households live in urban areas and the average age of the head of the household is around 50 years old. An interesting feature is that the proportion of households with a female head increases from around 16% to 24% in 2011. This is consistent with the migrant data in that Tajikistan saw an increase in workers leaving for better paying jobs outside of Tajikistan. In most cases the person leaving was the male and often he was the head of the household. Almost 80% of the households are native Tajik with Uzbek the next highest percentage.

The final variable we look at is education. There are four categories: less than secondary, secondary, post–secondary (vocational) and higher education. The highest attained level is reported for each head of household and the highest proportion of the household heads highest level of education is secondary. Between 42% and 44% of the households in the sample have some form of post-secondary education and this proportion is fairly stable. There is a drop in

Table 1: Summary Statistics and Variable Definitions

Variable	Description	2007	2009	2011
Relative Expenditure	Per person household expenditure relative to poverty line	Mean: 1.88 Std Dev: 1.95	Mean: 1.58 Std Dev: 1.41	Mean: 3.30 Std Dev: 4.90
Expenditure	Monthly household total expenditure (current prices – Somoni)	Mean: 1632.73 Std Dev: 1954.18	Mean: 1601.51 Std Dev: 1152.55	Mean: 4459.04 Std Dev: 5347.73
Income	Monthly household total income (current prices – Somoni)	Mean: 489.30 Std Dev: 795.18	Mean: 847.59 Std Dev: 1034.86	Mean: 2504.22 Std Dev: 5620.55
Informal	Dummy variable =1 if reported expenditure > 2 * income	Mean: 0.73	Mean: 0.59	Mean: 0.77
Migrant	Dummy variable =1 if household has a current migrant or a recently returned migrant	Mean: 0.25	Mean: 0.35	Mean: 0.42
Remittances	Reported monthly remittances per household (Somoni)	Prop. of households reporting remittances: 0.14 Mean(overall): 47.79 (282.77) Mean(excl. 0s) 339.38 (686.4)	Prop. of households reporting remittances: 0.12 Mean(overall): 140.09 (564.50) Mean(excl. 0s) 1189.81 (1210.5)	Prop. of households reporting remittances: 0.24 Mean(overall): 1556.75 (5631.5) Mean(excl. 0s) 6522.78 (10036)
Age	Age of head of household	Mean: 51.12 Std Dev: 12.87	Mean: 52.57 Std Dev: 12.51	Mean: 54.14 Std Dev: 12.54
Gender	Gender of Head of Household. =1 if female	Mean: 0.17	Mean: 0.16	Mean: 0.24
Ethnicity	Ethnicity of head of household. Discrete variable. =1 if Tajik, =2 if Uzbek, =3 other	Prop Tajik: 0.77 Prop Uzbek: 0.21 Prop Other: 0.018	Prop Tajik: 0.76 Prop Uzbek: 0.22 Prop Other: 0.019	Prop Tajik: 0.78 Prop Uzbek: 0.21 Prop Other: 0.017
Urban	Location of Household. Dummy variable =1 if urban	Mean: 0.34	Mean: 0.34	Mean: 0.34
Education	Highest education level of head of household. =1 < secondary =2 secondary =3 > secondary =4 Higher	Prop < secondary 0.20 Prop secondary 0.37 Prop > secondary 0.23 Prop Higher 0.19	Prop < secondary 0.20 Prop secondary 0.36 Prop > secondary 0.25 Prop Higher 0.19	Prop < secondary 0.17 Prop secondary 0.42 Prop > secondary 0.22 Prop Higher 0.19
Children	Number of Children under age of 15 in household	Mean (overall): 2.18 (1.71) Mean (exl. 0’s) 2.73 (1.48) Prop. with Children 0.80	Mean (overall): 2.19 (1.69) Mean (exl. 0’s) 2.70 (1.46) Prop. with Children 0.81	Mean (overall): 2.09 (1.77) Mean (exl. 0’s) 2.70 (1.55) Prop. with Children 0.77
Old	Number of household members 65 and over	Mean (overall): 0.29 (0.57) Mean (exl. 0’s) 1.25 (0.46) Prop. with 65+ 0.23	Mean (overall): 0.27 (0.55) Mean (exl. 0’s) 1.24 (0.43) Prop. with 65+ 0.22	Mean (overall): 0.28 (0.55) Mean (exl. 0’s) 1.24 (0.43) Prop. with 65+ 0.22

2011 but again this can be explained by the change in official head of household that is caused by migration.

5.1 Estimating Mobility and Vulnerability for Tajikistan

We use Bayesian estimation methods to estimate the parameters of (4). The outline of our procedure is given in Section 4. The Markov model has two main parameters, the initial distribution π_0 and the Markov transition matrix $\mathbf{P}$. Before estimating the model we first need to define the expenditure classifications. As noted earlier the definition of the expenditure classifications should be done in such a way as to cover the expenditure distribution in an efficient way without having classifications that have no members.

Since we are interested in poverty we define the first expenditure classification to be those households who spend less than the poverty line each month. After that we define expenditure classifications shown in Table 2. Recall that the variable that we use to discretize the data is the expenditure per person in a household relative to the poverty line. Thus the first classification includes all households who spend less that 139 Somoni per person per month. We then divide the range from the poverty line to twice the poverty line up into five classifications. This range is the commonly accepted group of households that are considered to be at risk. We then divide the rest of the distribution into equally spaced classifications with the highest classification including every household who spend more than 6 times the poverty line per person in a one month period.

As we are using Bayesian methods we need to define priors for the two parameters of the model, π_0 and $\mathbf{P}$. We use naturally conjugate Dirichlet priors for π_0 and for each row of $\mathbf{P}$ and the parameters of these priors are reported in Table 3. The priors defined in this way have a notional sample interpretation. The prior for π_0 is parameterized by a_0 where $a_{0i}-1$ can be interpreted as the number of observations in classification i in the notional population. Thus the prior for π_0 places equal prior weight on each expenditure class. The same interpretation can be used for the prior for $\mathbf{P}$. The interpretation is slightly different in that in the notional population a household is equally likely to move from its current classification to any other classification. Note however that in the posterior the observations from the data are added to these notional prior values (see (23)) so that the prior is designed to have minimal impact on the posterior estimates. The priors are thus chosen to have minimal impact on the posterior estimates but are still informative.

Table 2: Expenditure Classifications

Classification	Lower Bound	Upper Bound
1	0	1
2	1	1.2
3	1.2	1.4
4	1.4	1.6
5	1.6	1.8
6	1.8	2
7	2	3
8	3	4
9	4	5
10	5	6
11	6	100

Table 3: Prior Parameters

a_0

1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1

$A_{i\bullet}$

1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1

Using these priors and the data we use the sampling procedure given in Geweke (2005) to make 10,000 independent and identically distributed draws from the posterior distribution. We do so for the 2007 to 2009 transition and the 2009 and 2011 transition separately and for both combined as well. The last set of results is based on the assumption that the Markov transition matrices are equal across both transitions. We report both mobility measures and vulnerability measures.

5.2 Results

Table 4 reports the mobility and vulnerability measures for Tajikistan for the 2007 to 2009 transition and for the 2009 to 2011 transition separately. Given the significant differences between the two transitions it does not make sense to pool the transitions together and report pooled estimates. The full results including estimates of the initial cross-sectional distribution, the Markov transition probability matrix and the limiting cross-sectional distribution for each transition can be found in the Appendix.

For the first transition, from 2007 to 2009, the results are as you would expect for a country suffering from the global financial crisis. The initial expenditure distribution is clumped at the lower end of the relative expenditure distribution with 32% of households being below the poverty line and a further 35% of households having a per person relative expenditure above the poverty line but less than twice the poverty line (see Table A1). In the first transition the Shorrocks mobility measure is 0.966 with approximately two thirds of the movements being movements down the expenditure distribution. This leads to a limiting distribution that is shifted to the left of the initial distribution.[3] Figure 1 depicts the initial and hypothetical limiting distribution with the expenditure classifications between a relative expenditure of 1 and 2 aggregated together.

[3] Recall that the limiting distribution is the distribution you would get after an infinite number of transitions under $\mathbf{P}$.

Table 4: Mobility and Vulnerability Results

Sample	$\mathcal{M}_S(\mathbf{P})$	$\mathcal{M}_U(\mathbf{P})$	$\mathcal{M}_D(\mathbf{P})$	$V^1(\mathbf{P})$	$V^2(\mathbf{P})$	$V^5(\mathbf{P})$
			Full Sample			
07-09	0.966	0.289	0.677	0.314	0.348	0.357
	(0.010)	(0.013)	(0.015)	(0.015)	(0.014)	(0.015)
09-11	1.002	0.636	0.366	0.019	0.023	0.024
	(0.012)	(0.014)	(0.016)	(0.004)	(0.007)	(0.008)
			Urban Households			
07-09	0.969	0.390	0.579	0.190	0.211	0.217
	(0.015)	(0.021)	(0.023)	(0.020)	(0.020)	(0.021)
09-11	1.003	0.632	0.371	0.028	0.033	0.035
	(0.015)	(0.019)	(0.020)	(0.008)	(0.011)	(0.011)
			Rural Households			
07-09	0.961	0.257	0.704	0.351	0.388	0.396
	(0.012)	(0.015)	(0.017)	(0.019)	(0.018)	(0.019)
09-11	1.007	0.607	0.400	0.024	0.033	0.033
	(0.014)	(0.017)	(0.019)	(0.006)	(0.010)	(0.010)
			Households with Informal Activity			
07-09	0.945	0.340	0.606	0.241	0.268	0.276
	(0.014)	(0.017)	(0.019)	(0.018)	(0.017)	(0.018)
09-11	0.999	0.628	0.372	0.022	0.029	0.031
	(0.014)	(0.016)	(0.017)	(0.005)	(0.009)	(0.010)
			Households without Informal Activity			
07-09	0.985	0.266	0.719	0.360	0.399	0.412
	(0.013)	(0.018)	(0.020)	(0.023)	(0.022)	(0.024)
09-11	1.008	0.605	0.403	0.035	0.038	0.040
	(0.017)	(0.025)	(0.024)	(0.011)	(0.012)	(0.012)
			Households with Remittances			
07-09	0.996	0.384	0.612	0.204	0.201	0.204
	(0.019)	(0.028)	(0.029)	(0.029)	(0.026)	(0.026)
09-11	1.010	0.598	0.413	0.035	0.039	0.040
	(0.017)	(0.022)	(0.024)	(0.010)	(0.012)	(0.012)
			Households without Remittances			
07-09	0.960	0.290	0.670	0.314	0.353	0.363
	(0.011)	(0.014)	(0.016)	(0.016)	(0.015)	(0.016)
09-11	0.999	0.634	0.365	0.022	0.028	0.030
	(0.013)	(0.015)	(0.017)	(0.005)	(0.009)	(0.009)
			Higher Education Households (More than secondary)			
07-09	0.971	0.361	0.610	0.245	0.254	0.259
	(0.014)	(0.019)	(0.020)	(0.019)	(0.018)	(0.019)
09-11	1.008	0.642	0.366	0.025	0.032	0.034
	(0.014)	(0.018)	(0.019)	(0.007)	(0.010)	(0.010)
			Lower Education Households (Secondary or less)			
07-09	0.962	0.262	0.700	0.336	0.380	0.393
	(0.013)	(0.016)	(0.019)	(0.020)	(0.019)	(0.020)
09-11	0.996	0.597	0.400	0.026	0.032	0.033
	(0.015)	(0.017)	(0.020)	(0.007)	(0.010)	(0.010)

Figure 1: Cross-Sectional Relative Expenditure Distributions for 2007-2009 Transition

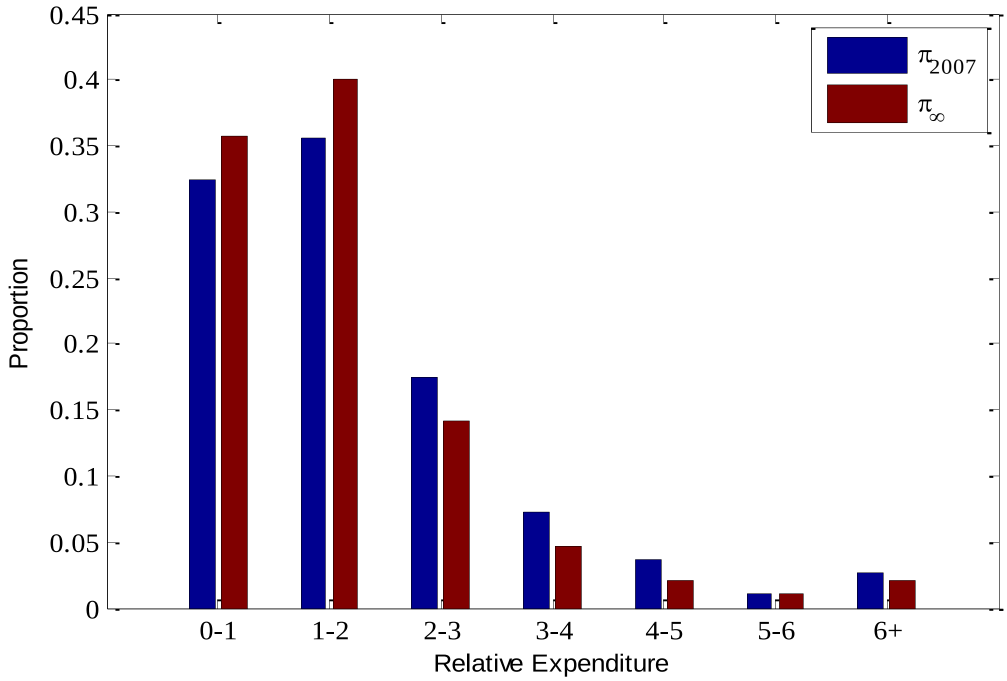
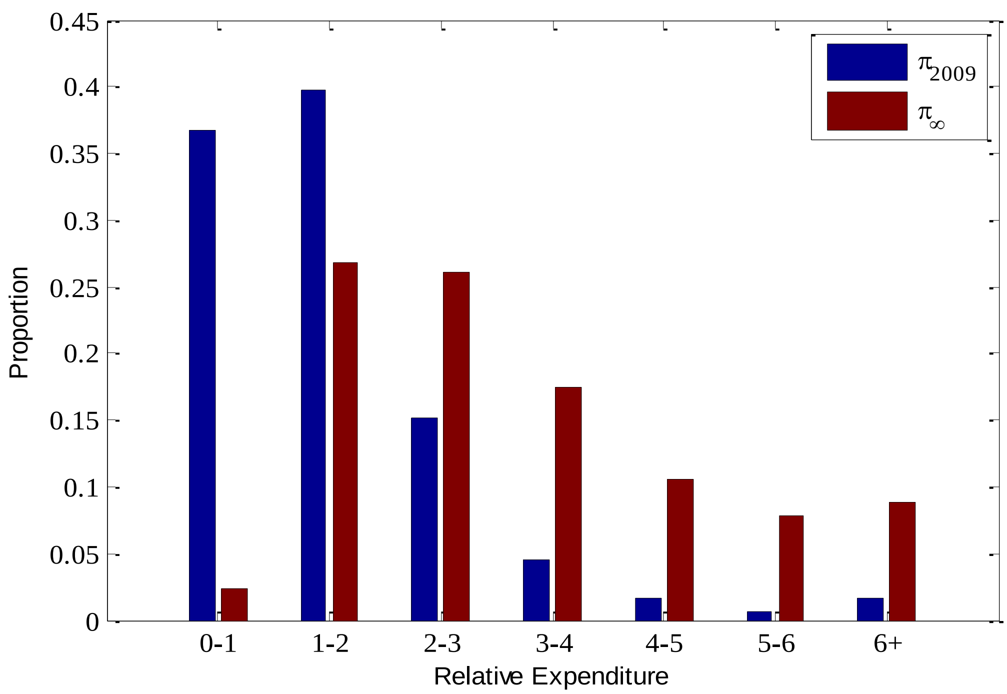


Figure 2: Cross-Sectional Relative Expenditure Distributions for 2009-2011 Transition



expenditure distribution. This leads to a limiting distribution that is shifted to the left of the initial distribution.[4] Figure 1 depicts the initial and hypothetical limiting distribution with the expenditure classifications between a relative expenditure of 1 and 2 aggregated together.

The vulnerability measures reported in Table 4 show the weighted probability of moving back below the poverty line after 1 year, 2 years and 5 years respectively. For the 2007-2009 transition these probabilities are high at 31.4%, 34.8%, and 35.7% respectively. Table A1 in the Appendix reports the individual probabilities of moving back below the poverty line for each expenditure classification. The probabilities of moving back to below the poverty line are all above 20% except for the highest expenditure classification. This classification includes all households with a per person expenditure greater than six times the poverty line. The probability of moving below the poverty line is higher for the classifications closer to the poverty line but it is clear that households with expenditures greater than twice the poverty line are at risk in this transition.

The next transition is from 2009-2011 and this was a transition where Tajikistan was recovering from the effects of the global economic crisis. This is evident in the results. The initial distribution for 2009 is similar to what we saw in 2007 with the largest proportions of households in the lowest few expenditure classifications. The estimates for the initial distribution in 2009 are reported in Table A2 and in Figure 2. The overall mobility score for the 2009-2011 transition is slightly higher than that of 2007-2009 but when we look at the directional mobility the contrast with 2007-2009 is clear. Instead of two-thirds of the mobility being downward mobility it is now the case that two-thirds of the mobility is upward mobility. This predominance of upward mobility then yields a vastly different limiting distribution as we see in Figure 2. The recovery period from 2009-2011 is clearly a period of rising incomes and expenditures. These are relatively good times for Tajikistan and this is evident in the vulnerability measures as well. The vulnerability measures for a one year transition, two years and five years are 0.019, 0.023 and 0.024 respectively. The individual probabilities in Table A2 also show that there is no expenditure classification with a probability of moving back below the poverty line above 5%. Therefore in this transition very few households are at risk or vulnerable and certainly

[4] Recall that the limiting distribution is the distribution you would get after an infinite number of transitions under $\mathbf{P}$.

households whose initial expenditures are less than twice the poverty line could not be considered vulnerable in this transitions.

The next part of Table 4 reports the mobility and vulnerability measures for both transitions for different subgroups of the population. Since the 2007-2009 is a transition where there is vulnerability we will discuss that transition in more detail. In the 2009-2011 transition all subgroups have similar low vulnerabilities. The first group we looked at was urban households. About a third of the population lives in urban areas and urban households appear to be significantly less vulnerable than rural households. The vulnerability measures for urban households are approximately equal to 0.2 whilst the vulnerability measures for rural households are between 0.35 and 0.40. This is clear evidence that rural households are more vulnerable to urban households in Tajikistan.

Following earlier work by Dimova et al. (2006) and Abdulleov et. al. (2012) we assign households to the informal sector if their reported expenditures are greater than twice their reported incomes. Households with this revealed informal activity are significantly less vulnerable to poverty in this transition than those who have not revealed any informal sector activity. Again the vulnerability measures range from 0.24 to 0.28 for informal sector households and from 0.36 to 0.41 for households without informal sector activity.

When we divide the population up by education we also get the expected results. Assigning households with their highest educational attainment being greater than secondary school we see these households with vulnerability measures approximately equal to 0.25. For households with low educational attainment the vulnerability measures range from 0.34 to 0.39.

The last sub-groups we looked at were households with migrants. We divided households up in two ways. First we divided households up into those with current or past migrants versus those without any migrants and secondly we divided households up in to households that received remittances and those that did not. The results here are interesting. Households that receive remittances are less vulnerable than those that did not receive remittances. Households that did receive remittances had vulnerability measures of approximately 0.20 whilst the households that did not receive remittances had vulnerability measures ranging from 0.31 to 0.36. The results for migrants are different in that households that had migrants had vulnerability measures ranging from 0.31 to 0.32 whilst households without migrants had vulnerability measures ranging from 0.28 to 0.33. Having a migrant is not enough to protect a household from poverty.

6 Discussion and Conclusion

In this paper we have shown how the literature on Markovian models of income can be utilized to model and measure the dynamics of income and poverty. We use a first order discrete state Markov model to model income and, in the case of poverty, expenditure dynamics. The appeal of the Markov assumption is that it imposes very little in terms of structure on the model. This allows us to measure income and poverty dynamics as functions of the underlying parameters of the Markov model. In particular we are able to investigate movements up and down the income distribution via the use of established mobility measures. We then show that if we discretize the income (expenditure) distribution appropriately – and by appropriately we mean in context to defined poverty lines – then we are able to measure movements into and out of poverty. We then introduce a measure of vulnerability to poverty as the weighted probability of a household falling back into poverty over a given horizon. Our measure enables us to look at short, medium and long-run vulnerability to poverty in a formal way.

We applied our methodology to a household panel from Tajikistan that covered the years starting in 2007 and ending in 2011. We observed expenditures for the years of 2007, 2009, and 2011 which enabled us to use our methodology in two distinct types of economic climates. The first transition from 2007 to 2009 was a period of recession in Tajikistan – coinciding in fact with the global financial crisis – while the second period from 2009 to 2011 coincided with a brisk recovery from recession.

We found that during the recession transition almost all households were vulnerable to poverty while during the recovery period almost no households were vulnerable to poverty. When we looked at the first transition in more detail we found separately that urban households were less vulnerable to poverty, more educated households were less vulnerable to poverty, households that were involved in the informal sector were less vulnerable to poverty and households receiving remittances from overseas were less vulnerable to poverty. Interestingly, households with a current or very recent migrant did not have a significantly lower measured vulnerability to poverty; only those households receiving remittances from migrants had a lower vulnerability to poverty.

What we attempted to do in this paper is to introduce a practical and formal way to measure vulnerability to poverty. We showed how to measure vulnerability to poverty and applied this to the case of Tajikistan. This application showed us the power of this method in measure vulnerability across sub-populations, across time and in identifying the subgroups that are vulnerable. We think that this is interesting and useful. Future versions of this paper will refine the analysis including the introduction of correlates to the analysis.

7 References

Abdulloev, I., Gang, I. N., & Landon-Lane, J. (2012). Migration as a Substitute for Informal Activities: Evidence from Tajikistan. *Research in Labor Economics*, *34*, 205-227.

Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, 318-351.

Co, C. Y., Landon-Lane, J. S., & Yun, M. S. (2006). Inter-state dynamics of invention activities, 1930–2000. *Journal of Applied Econometrics*, *21*(8), 1111-1134.

Danzer, A. M., B. Dietz, K. Gatskova (2013a). Tajikistan Household Panel Survey: Migration, Remittances and the Labor Market. Survey report. Regensburg. http://www.ios-regensburg.de/fileadmin/doc/VW_Project/Booklet-TJ-web.pdf, retrieved on 12.05.2014.

Danzer, A. M., B. Dietz, K. Gatskova (2013b). Migration and Remittances in Tajikistan: Survey Technical Report. http://ideas.repec.org/p/ost/wpaper/327.html, retrieved on 12.05.2014.

Deaton, A. (1997). The analysis of household surveys: a microeconometric approach to development policy. World Bank Publications.

Devroye, L. (2006). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.

Dimova, R., Gang, I. N., & Landon-Lane, J.S. (2006). The Informal Sector During Crisis and Transition in Guha-Khasnobis B. and R. Kanbur (eds) *Informal Labour Markets and Development*, pages 88—108, Palgrave-McMillan Press, September.

Dutta, I., Foster, J., & Mishra, A. (2011). On measuring vulnerability to poverty. *Social Choice and Welfare*, 37(4), 743-761.

Gang, I. N., Landon-Lane, J. S., & Yun, M. S. (2002). Gender differences in German upward income mobility. *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften*, **123**(1), p3-14.

Gang, I. N., Landon-Lane, J. S., & Yun, M. S. (2004). *Good and bad mobility: A note on directional mobility measures*. mimeo, Department of Economics, Rutgers University.

Gang, I. N., Landon-Lane, J., & Yun, M. S. (2009). Measuring Income Assimilation of Migrants to Germany. *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften*, *129*(2), 333-342.

Geweke, J., Marshall, R. C., & Zarkin, G. A. (1986a). Mobility indices in continuous time Markov chains. *Econometrica: Journal of the Econometric Society*, 1407-1423.

Geweke, J., Marshall, R. C., & Zarkin, G. A. (1986b). Exact inference for continuous time Markov chain models. *The Review of Economic Studies*, *53*(4), 653-669.

Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics* (Vol. 537). John Wiley & Sons.

Litchfield, J. A. (1999). Inequality: Methods and Tools. Internet site: http://www. worldbank. org/poverty/inequal/methods/litchfie. pdf (Accessed August 1, 2014).

Prais, S. (1955). Measuring social mobility. *Journal of the Royal Statistical Society. Series A (General)*, 56-66.

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, 219-231.

Shorrocks, A. F. (1976). Income mobility and the Markov assumption. *The Economic Journal*, 566-578.

Shorrocks, A. (1978a). Income inequality and income mobility. *Journal of Economic Theory*, *19*(2), 376-393.

Shorrocks, A. F. (1978b). The measurement of mobility. *Econometrica: Journal of the Econometric Society*, 1013-1024.

Statistical Agency of Tajikistan (2011) Database. Available at: http://stat.tj/ru/database/real-sector/ [Accessed 17 September 2011].

Statistical Committee of CIS (2011) Average Monthly Nominal Wage in the CIS Countries, in national currency. Available at: http://www.cisstat.com/index.html [Accessed 17 September 2011].

World Bank (2009) Republic of Tajikistan Poverty Assessment: Report No. 51341-TJ. Human Development Sector Unit, Central Asia Country Unit, Europe and Central Asia Region. Available at https://openknowledge.worldbank.org/bitstream/handle/10986/3159/513410ESW0P112101Official0Use0Only1.txt?sequence=2

World Bank (2011) Migration and Remittances. Factbook 2011. Second edition. Available at: http://siteresources.worldbank.org/INTLAC/Resources/Factbook2011-Ebook.pdf [Accessed 12 August 2011].

World Bank (2013a) World Data Bank. Available at: http://databank.worldbank.org/data/home.aspx [Accessed 17 May 2013].

World Bank (2013b). Annual Remittances Data. Available at: http://go.worldbank.org/092X1CHHD0, retrieved on 12.05.2014.

8 Appendix

Table A1: Estimation Results for 2007-2009 Transition

Initial Distribution: π_{2007}

0.324	0.077	0.077	0.076	0.076	0.05	0.174	0.073	0.036	0.011	0.026
(0.013)	(0.008)	(0.007)	(0.007)	(0.007)	(0.006)	(0.011)	(0.007)	(0.005)	(0.003)	(0.004)

Markov Transition Matrix: $\mathbf{P}$

0.418	0.133	0.086	0.088	0.057	0.036	0.135	0.029	0.007	0.007	0.005
(0.024)	(0.017)	(0.014)	(0.014)	(0.011)	(0.009)	(0.017)	(0.008)	(0.004)	(0.004)	(0.003)
0.368	0.083	0.102	0.056	0.056	0.074	0.157	0.038	0.029	0.01	0.028
(0.047)	(0.026)	(0.029)	(0.022)	(0.022)	(0.025)	(0.035)	(0.018)	(0.016)	(0.010)	(0.016)
0.380	0.168	0.149	0.047	0.065	0.029	0.047	0.075	0.019	0.01	0.01
(0.047)	(0.036)	(0.034)	(0.020)	(0.024)	(0.016)	(0.020)	(0.025)	(0.013)	(0.010)	(0.010)
0.308	0.114	0.066	0.085	0.085	0.067	0.132	0.066	0.029	0.01	0.038
(0.044)	(0.031)	(0.024)	(0.027)	(0.027)	(0.024)	(0.033)	(0.024)	(0.016)	(0.010)	(0.018)
0.352	0.066	0.047	0.085	0.102	0.038	0.195	0.047	0.019	0.019	0.029
(0.046)	(0.024)	(0.020)	(0.027)	(0.029)	(0.018)	(0.038)	(0.020)	(0.013)	(0.013)	(0.016)
0.258	0.109	0.136	0.069	0.056	0.109	0.123	0.042	0.056	0.015	0.028
(0.051)	(0.036)	(0.039)	(0.029)	(0.027)	(0.036)	(0.038)	(0.023)	(0.027)	(0.014)	(0.019)
0.358	0.082	0.091	0.096	0.052	0.048	0.177	0.057	0.022	0.005	0.013
(0.032)	(0.018)	(0.019)	(0.019)	(0.015)	(0.014)	(0.025)	(0.015)	(0.010)	(0.005)	(0.008)
0.204	0.079	0.069	0.088	0.088	0.069	0.176	0.108	0.02	0.03	0.069
(0.039)	(0.026)	(0.025)	(0.028)	(0.028)	(0.025)	(0.037)	(0.030)	(0.014)	(0.017)	(0.025)
0.212	0.089	0.107	0.16	0.089	0.072	0.159	0.019	0.019	0.037	0.037
(0.053)	(0.037)	(0.040)	(0.048)	(0.038)	(0.034)	(0.047)	(0.018)	(0.018)	(0.024)	(0.025)
0.204	0.084	0.123	0.083	0.044	0.084	0.084	0.044	0.084	0.044	0.123
(0.079)	(0.055)	(0.063)	(0.054)	(0.040)	(0.054)	(0.054)	(0.040)	(0.055)	(0.040)	(0.064)
0.184	0.07	0.138	0.07	0.07	0.071	0.206	0.048	0.07	0.025	0.048
(0.058)	(0.038)	(0.052)	(0.038)	(0.038)	(0.039)	(0.060)	(0.031)	(0.038)	(0.024)	(0.032)

Limiting Distribution: $\bar{\pi}$

0.357	0.111	0.094	0.081	0.064	0.051	0.141	0.047	0.021	0.011	0.021
(0.015)	(0.009)	(0.009)	(0.007)	(0.007)	(0.006)	(0.010)	(0.006)	(0.004)	(0.003)	(0.004)

Table A2: Estimation Results for 2009-2011 Transition

Initial Distribution: π_{2009}

0.367	0.108	0.093	0.084	0.064	0.049	0.151	0.045	0.016	0.006	0.017
(0.014)	(0.009)	(0.008)	(0.008)	(0.007)	(0.006)	(0.010)	(0.006)	(0.004)	(0.002)	(0.004)

Markov Transition Matrix: **P**

0.034	0.038	0.036	0.059	0.072	0.074	0.332	0.187	0.086	0.025	0.057
(0.008)	(0.009)	(0.009)	(0.011)	(0.012)	(0.012)	(0.021)	(0.018)	(0.013)	(0.007)	(0.011)
0.014	0.041	0.034	0.069	0.014	0.081	0.338	0.216	0.089	0.055	0.048
(0.010)	(0.016)	(0.015)	(0.021)	(0.010)	(0.022)	(0.039)	(0.034)	(0.023)	(0.019)	(0.017)
0.024	0.016	0.016	0.055	0.055	0.093	0.342	0.117	0.141	0.086	0.055
(0.013)	(0.011)	(0.011)	(0.020)	(0.020)	(0.025)	(0.042)	(0.028)	(0.030)	(0.024)	(0.020)
0.027	0.035	0.035	0.035	0.086	0.044	0.342	0.266	0.069	0.035	0.027
(0.015)	(0.017)	(0.017)	(0.017)	(0.026)	(0.019)	(0.043)	(0.040)	(0.023)	(0.017)	(0.015)
0.023	0.023	0.045	0.055	0.077	0.077	0.295	0.197	0.088	0.044	0.077
(0.015)	(0.016)	(0.021)	(0.024)	(0.028)	(0.028)	(0.047)	(0.042)	(0.030)	(0.021)	(0.028)
0.015	0.015	0.029	0.029	0.097	0.07	0.315	0.125	0.083	0.124	0.097
(0.014)	(0.014)	(0.020)	(0.020)	(0.035)	(0.030)	(0.054)	(0.038)	(0.032)	(0.039)	(0.034)
0.005	0.025	0.044	0.055	0.055	0.114	0.267	0.232	0.079	0.04	0.084
(0.005)	(0.011)	(0.014)	(0.016)	(0.016)	(0.022)	(0.031)	(0.029)	(0.019)	(0.014)	(0.019)
0.031	0.031	0.045	0.104	0.046	0.031	0.207	0.119	0.149	0.134	0.104
(0.021)	(0.021)	(0.025)	(0.037)	(0.025)	(0.021)	(0.049)	(0.039)	(0.043)	(0.041)	(0.037)
0.035	0.035	0.036	0.068	0.068	0.035	0.163	0.165	0.163	0.067	0.165
(0.032)	(0.032)	(0.033)	(0.044)	(0.044)	(0.033)	(0.065)	(0.066)	(0.066)	(0.044)	(0.065)
0.061	0.06	0.062	0.116	0.06	0.061	0.282	0.061	0.116	0.06	0.061
(0.054)	(0.054)	(0.055)	(0.074)	(0.055)	(0.055)	(0.103)	(0.055)	(0.074)	(0.055)	(0.055)
0.033	0.033	0.034	0.033	0.033	0.094	0.244	0.184	0.064	0.154	0.094
(0.031)	(0.031)	(0.031)	(0.031)	(0.030)	(0.050)	(0.073)	(0.067)	(0.042)	(0.062)	(0.050)

Limiting Distribution: $\bar{\pi}$

0.024	0.031	0.041	0.065	0.059	0.072	0.261	0.174	0.105	0.079	0.089
(0.008)	(0.008)	(0.009)	(0.011)	(0.010)	(0.010)	(0.018)	(0.015)	(0.015)	(0.012)	(0.013)