What Happens Between the Waves? Estimating Inter-Wave Dynamics from Limited Survey Data with Application to Poverty Transitions in South Africa and Vietnam

Bob Baulch, Lindsay Chant & Sherman Robinson[1]

1 Introduction

In recent years, the number of household panel data sets available for developing countries has increased dramatically and, along with these, the number of studies on poverty dynamics and economic mobility in low and middle-income countries (see Dercon and Shapiro, 2007; Baulch, forthcoming). These studies confirm previous studies (Baulch and Hoddinott, 2000; CPRC, 2004) that movements in and out of poverty are ‘strikingly large’, whether poverty is conceptualised in absolute terms (as in the poverty dynamics literature) or in relative terms (as in many studies of economic mobility). It is however difficult to compare the results of these studies because of the different time periods they span and the different welfare measures and poverty lines they use. What conclusions about the magnitude of chronic poverty can one, for example, draw from two countries with panel data saves that are respectively two and five years apart and one of which measures poverty using adult equivalent incomes while the other uses per capita expenditures? Furthermore, while panel data studies are increasingly available, it is often extremely difficult for researchers to access the unit record (household level) data in the extant studies. A method for making comparisons of poverty dynamics and economic mobility using the limited information (typically in the form of transition matrices, poverty and inequality measures, and growth rates) contained in published studies and reports would therefore be extremely useful.

This paper aims to address these lacunae by outlining a procedure for estimating annual poverty transition matrices from periodic panel survey data using cross entropy methods. The aim of the study is to develop a flexible econometric method for estimating

[1] Bob Baulch is Coordinator of the Poverty Dynamics and Economic Mobility Theme at the Chronic Poverty Research Centre; Lindsay Chant is a Quantitative Economist at LEI Wageningen UR, The Netherlands; Sherman Robinson is Professor of Economics at the University of Sussex. This research was funded by the Chronic Poverty Research Centre, which is itself funded by UK Department for International Development’s Development.

poverty transition dynamics in the years between survey waves. We apply the method to panel data from two countries with three waves of panel data, South Africa and Vietnam. The next section describes the panel data used. Section 3 develops the cross-entropy estimation method we used to estimate inter-wave dynamics, first in the context of maximum entropy estimation, and then for cross entropy estimation with noise, together with how corresponding maximisation problem can be solved using GAMS. Section 4 outlines the results of our preliminary application of this methodology using panel data from South Africa and Vietnam. A number of poverty and mobility measures are calculated from the estimation output, along with the various diagnostic statistics. Section 5 conducts a sensitivity analysis, in which the annual poverty transition probabilities are estimated with different levels of observed and unobserved error, while Section 6 discusses some caveats and extensions to the methodology. Section 7 concludes with a summary of the empirical results for South Africa and Vietnam and an initial assessment of the estimation technique..

2 Data

When estimating monetary poverty, inequality and other welfare measures, expenditure is usually preferred over income as a measure of welfare for several reasons (Deaton, 1997). First, consumption expenditure is likely to be more regular and subject to less measurement error than income, especially where the income from agriculture and household enterprises are concerned. Second, households expenditure patterns tend to be less volatile than income, so expenditure is more likely to truly reflect households long-term living standards. Survey based estimates of income in developing countries also tend to be subject to more imputation problems and be lower (sometimes substantially) than corresponding the corresponding survey based estimates of expenditures. For all these reasons, we focus on expenditure as our main welfare measure in both KwaZulu-Natal in South Africa (for which both income and expenditure data are available) and Vietnam (for which only expenditure data are available).

The following sub-sections described the panel data and poverty lines used in our empirical applications.

2.1 South Africa

Data for South Africa are taken from the KwaZulu-Natal Income Dynamics Survey (KIDS). This is a three waves panel study conducted in 1993, 1998 and 2004. The first wave of Kids is derived from the KwaZulu and Natal portions of the Project for Statistics on Living Standard and Develoment (PSLSD), and contains information on 1558 households of all races located in 73 sampling clusters (May and Woolard, 2006). For the 1998 waves, which was restricted to the newly created province of KwaZulu-Natal, white and coloured households are excluded because of their limited samples in the 1993 survey. Of the 1354 eligible households, 1171 were tracked. Using the same core households as the 1993 survey, the 2004 wave of KIDS tracks 867 households, of whom 760 are core households and the remainder households which had split-off from their 'parent' households. Although, the modules of the three waves of KIDS contains some new modules, the all three modules used consistent information on household composition and demographics, and income and expenditure modules. It is this data, which is used to calculate the transition matrices reported below.

Following, Woolard and Klasen, the absolute poverty line for households in KwaZulu-Natal (KZN) is set at 422 Rand per month .[2] The 974 households for which we have data are then divided-up into seven categories based on multiples of the absolute poverty line defined in expenditure terms is shown in Table 1. Households in categories 1 and 2 have expenditure levels below the absolute poverty line and therefore are classified as 'poor' whereas households in categories 3 to 7 have monthly expenditure levels on or above the poverty line and are therefore classified as 'non-poor'.

[2] 40% of households are poor by construction: Woolard and May (2005) choose the absolute poverty line as the level of expenditure that makes 40% of households 'poor'.

Table 1 Transition Categories Defined as Multiples of the Absolute Poverty Line (KZN)

		Poverty Line	**R422**
Category	Rand	Lower limit	Upper limit
1	0-210	0.0	0.5
2	211-421	0.5	1.0
3	422-632	1.0	1.5
4	633-843	1.5	2.0
5	844-1054	2	2.5
6	1055-1265	2.5	3.0
7	1266+	3	3+

The numbers of households that move between the two bottom poor categories and the other five non poor categories between the 1993-1998 and 1998-2004 are shown in panel A of Table 2, along with the corresponding raw transition probabilities in panel B. Although the number of households transitioning between the two states is different in the first and second transitions, the poverty transition probabilities are similar for the two transitions with only a small increase in the poor becoming non poor in the second transition and a slightly larger increase in the probability of non poor households becoming poor.

Table 2 Transition Matrices for KwaZulu-Natal, 1993-1998 and 1998-2004

A

		1998	
		Poor	Non Poor
1993	Poor	293	94
	Non Poor	235	352

		2004	
		Poor	Non Poor
1998	Poor	396	132
	Non Poor	191	255

B

		1998	
		Poor	Non Poor
1993	Poor	0.76	0.24
	Non Poor	0.40	0.60

		2004	
		Poor	Non Poor
1998	Poor	0.75	0.25
	Non Poor	0.43	0.57

The sum of the probabilities of households remaining poor and becoming poor (the first column in panel B) is greater than the sum of the probabilities households becoming non-poor and remaining non-poor (the second column in panel B). This suggests that the number of households below the poverty line increases over the survey period. This is borne out by the data as 387 (40%) of surveyed households are poor in 1993, rising to 528 (54%) in 1998 and 587 (60%) in 2004.

Poverty transition matrices only examine the number of households who cross, and do not cross, an essentially arbitrary poverty line, while mobility may occur between all points of the distribution. To assess this, a number of mobility measures have been proposed of which the most popular is the Shorrocks'mobility index (Shorrocks, 1997).[3] The Shorrocks mobility index, usually denoted by M, is a two-stage index derived from the transition matrix which is simulations studies have shown to be fairly robust to measurement error (Cowell and Schulter, 1998). It is evaluated as:

$$M = \frac{n - Tr(P)}{n - 1}, \qquad (1)$$

where *n* is the number of welfare states and *Tr(P)* is the trace of the transition matrix (*P*). A zero mobility index means that all households remain in state and are immobile and a value of 1 corresponds to perfect mobility. Using the 7 x 7 transition matrices we have constructed for KwaZulu-Natal, Shorrocks'mobility index is 0.97 between 1993 and 1998, 0.92 between 1998 and 2004 and 0.92 for the entire survey period. Such values indicate a relatively high level of mobility for the households included in the KIDS panel.

2.2 Vietnam

Data for Vietnam are taken from the Vietnam Household Living Standards Survey (VHLSS), a biennial household survey program which began in 2002, with subsequent waves in 2004 and 2006.[4] The VHLSS is a rotating core-and-module survey, in which common set of core modules (covering household composition, employment, incomes, expenditures and housing) are combined with specialist modules (for education and health, agricultural and non-farm enterprises, etc) in different years. The VHLSS expenditure module is administered to a sub-sample of total household sample survey,

[3] See Chapter 6 of Fields (2001) for an excellent summary of the extant mobility measures.
[4] The 2008 Vietnam Household Living Standard Survey is currently in the field.

which has varied from 30,000 households in 2002 to 9,189 households in 2004 and 9, 190 households in 2006 (Pham and Nguyen, 2006, World Bank, 2007).[5]

Poverty is measured using a cost-of-basic needs poverty line based on estimating the cost of a person acquiring 2,100 KCals per day plus a modest allowance for non-food expenditures estimated (Glewwe, Agrawal and Dollar, 2004; World Bank, 2007). This poverty line is applied to a per capita expenditure aggregate which is adjusted for regional and temporal (intra-annual) price differences. The per capita expenditure poverty line is set at 1,920,000Vietnam Dollars (VND) in 2002, VND 2,077,210 in 2003 and VND 2599,850 in 2006.

While the VHLSS is designed to be nationally representative, and utilised a master-sample design (Petersson, 2003), its panel sub-component—50% of which is replaced in each survey wave—may not be. In particular, under condition of rapid (if unofficial) migration to the main urban centres, concerns have been expressed about the representativeness of the VHLSS urban sub-sample (Pincus and Sender 2006; Nguyen and Hansen, 2007). Table 3 shows the transition matrices estimated from the VHLSS for its 2002-04 and 2004-06 sub-panels. There is also a transition matrix for 2002-2006, although because of the rotation of panel households this only covers 2,151 households.

Table 3 Transition Matrices for Vietnam, 2002-2004 and 2004-2006

A

		2004	
		Poor	Non Poor
2002	Poor	488	634
	Non Poor	320	2719

		2006	
		Poor	Non Poor
2004	Poor	384	517
	Non Poor	300	3488

B

		2004	
		Poor	Non Poor
2002	Poor	0.43	0.57
	Non Poor	0.11	0.89

		2006	
		Poor	Non Poor
2004	Poor	0.43	0.57
	Non Poor	0.08	0.92

[5] The total sample sizes of the successive rounds of VHLSS were 75,00 households in 2002, 45,00 households in 2004 and >??? Households In 2006.

As in KwaZulu-Natal, extended transition matrices based on multiples of the poverty line are calculated, in this case using eight categories.[6] The Shorrocks Mobility Index for Vietnam is 0.73 for the first wave between 2002 and 2004, 0.53 between 2004 and 2006, and 0.71 for 2002 to 2006.

3 Estimating Annual Poverty Transitions

Estimating annual poverty transition probabilities from periodic survey data is an under-defined problem. The number of annual transition probabilities exceeds the number of known data points as typically only the numbers of individuals/households in each welfare group are known and only for the survey years. This type of problem cannot be estimated using traditional econometric techniques due to the lack of data; however, maximum/cross entropy estimation methods are ideally suited to situations of limited data.

The entropy concept was first introduced in statistical mechanics (Shannon, 1948, p.11) and extended by Jaynes (1957) to estimate unknown probabilities. Golan et al. (1996) apply entropy estimation to situations of aggregated or limited data in which the number of unknowns is greater than the number of available data points. Entropy refers to the amount of uncertainty associated with a variable. The entropy metric provides a criterion for selecting transition probabilities from the multitude of consistent values,

$$x_{j,t} = x_{i,t-1} p_{ij} ,$$

without the additional assumptions that would be required by traditional methods. In maximum entropy estimation, maximising the entropy metric selects the set of probabilities that have the greatest degree of uncertainty associated with them whilst still being consistent with the data. The underlying principle is that maximum entropy estimation leads to a set of probabilities that differ from the uniform distribution (reflecting maximum uncertainty) only by the information signal contained within the data. As such, entropy estimation is ‘maximally non-committal’. Cross entropy estimation uses a specified prior probability distribution rather than the uniform distribution. This allows for prior beliefs about the likely probabilities to be incorporated

[6] The categories are <0.5z, 0.5-1z, 1-1.5z, 1.5-2z, 2-2.5z, 2.5-3z, 3-3.5z, and >3z, where z is the poverty line.

into the estimation procedure. Minimising the cross entropy metric selects the set of probabilities that differ from the prior distribution by the information signal contained within the data. The resultant probability distribution will be the same as the prior distribution if the data contain no additional information over that contained in the prior probability distribution.

Entropy estimation involves maximising the entropy metric, *H*, subject to two constraints, the data consistency constraint and the adding up constraint. The data consistency constraint requires that the estimated probabilities are calculated such that the values of the observed means (*y*) hold, given the values of the possible outcomes (*X*). The adding up constraint ensures that the probabilities for each outcome sum to 1. The maximum entropy estimation method following Golan et al. (1996) is,

Max.

$$H(p) = -\sum_k p_k \log p_k \quad .$$

Subject to,

$$y = Xp_k$$

and

$$\sum_k p_k = 1 \quad .$$

The maximum entropy estimation method selects the set of probabilities with the highest entropy or uncertainty subject to the data that is known. Therefore the estimation process is ‘maximally non-committal’, no assumptions about the probabilities are imposed and the probabilities only differ from the uniform distribution by the information content of the data.

The maximum entropy approach outlined in the previous section applies only to pure inverse problems and does not include prior information about the likely values of the transition probabilities. Furthermore, the data are likely to be measured with error, and therefore the estimation cannot be defined as a pure inverse problem. So, although complete transition count data are not available, prior beliefs about the likely values of the transition probabilities and additional data are available which, if incorporated into the estimation procedure, should contribute to more accurate estimates. Cross entropy estimation with noise is a more general form of the maximum entropy method, which allows for both noisy data and for prior beliefs and additional information to be included in the estimation method.

3.1 Cross-Entropy Estimation

3.1.1 Maximum Entropy Estimation

Cross entropy estimation allows additional information and prior beliefs about the likely values of the probabilities to be incorporated into the estimation procedure. Whereas the maximum entropy method selects the probabilities which are the most uncertain, given the observed data, the cross entropy method selects transition probabilities which are as close as possible to the specified prior values (*q*) given the observed data. Therefore whilst maximum entropy estimation selects the distribution of probabilities closest to the uniform distribution given the data, the cross- entropy method selects the distribution of probabilities closest to the prior distribution, given the data. Formally the cross entropy estimation method without noise is,

minimise,

$$I\left(p_{i,j,t}\right) = -\sum_{j,t} p_{i,j,t} \log\left(\frac{p_{i,j,t}}{q_{i,j,t}}\right) ,$$

subject to,

$$y_{j,t+1} = x_{i,t} p_{i,j}$$

and

$$\sum_{j} p_{i,j,t} = 1 \ .$$

The prior probabilities are a defined as a transition matrix of likely probabilities specified from additional data or from prior beliefs. When the prior probabilities are specified as a set of uniform distributions for each Markov state, the cross-entropy formulation collapses to the maximum entropy formulation. As with the maximum entropy estimation, the cross-entropy estimation yields the transition probabilities of moving from state *i* to state *j* in each period, such that the sum of the probabilities across each initial state is equal to one and all members of the initial state are accounted for.

3.1.2 Cross-Entropy Method with Noise

The observed number of people in each Markov state is likely to be measured with error so estimating the transition probabilities is not a pure inverse problem and an error term must be included. Golan et al. (1996, p.111) present an extended (cross) entropy estimation method with noise. They define the error term as a discrete random variable, *e*, comprising of two components: error supports, *v*, and error weights, *w*,

$$e_{j,t} = \sum_{d} v_{d,j,t} * w_{d,j,t} ,$$

and incorporate the error term into the cross-entropy formulation thus,

minimise,

$$I\left(p_{i,j,t}, w_{d,j,t}\right) = -\sum_{j,t} p_{i,j,t} \log\left(\frac{p_{i,j,t}}{q_{i,j,t}}\right) + -\sum_{d,j,t} w_{d,j,t} \log\left(\frac{w_{d,j,t}}{u_{d,j,t}}\right)$$

subject to:

$$y_{j,t+1} = x_{i,t} p_{i,j,t} + v_{d,j,t} w_{d,j,t}$$

and

$$\sum_{j} p_{i,j,t} = 1, \qquad \sum_{d} w_{d,j,t} = 1,$$

where *u* are the prior values of the weights of the error term. As with the transition probabilities, the estimation procedure selects the error weights which are closest to the prior weights but consistent with the data. Cross-entropy estimation with noise yields estimates for both the transition probabilities and the weights of the error term. The prior values of the error weights can be specified as uniform without the estimation taking the maximum entropy form if the prior transition probabilities are non-uniform. If the prior values of the probabilities *and* weights are specified as uniform, the formulation collapses to the maximum entropy with noise specification,

maximise,

$$H\left(p_{i,j,t}, w_{d,j,t}\right) = -\sum_{j,t} p_{i,j,t} \log p_{i,j,t} + \sum_{d,j,t} w_{d,j,t} \log w_{d,j,t}$$

subject to:

$$y_{j,t+1} = x_{i,t} p_{i,j,t} + v_{d,j,t} w_{d,j,t}$$

and

$$\sum_j p_{i,j,t} = 1, \qquad \sum_d w_{d,j,t} = 1.$$

Note that the cross entropy and maximum entropy estimation specifications estimate transition probabilities for each period, rather than a set of stationary transition probabilities which apply to all periods. This is desirable unless it is believed that the observed transition probabilities are representative of the ergodic distribution.

3.2 Estimating Annual Transition Probabilities from Periodic Survey Data

The cross entropy method with noise described above is estimated using a GAMS program modified from Chant (2008). This program estimates annual transition probabilities from known periodic survey data by minimising the entropy distance between the prior probabilities and the transition probabilities.[7] Sufficient information to estimate the annual transition probabilities is provided by assuming a linear trend between the number of households in each category in the survey years. Other trends can be applied without any loss of functionality. The equations of the model are grouped into six blocks: the entropy objective, system dynamics, knowledge, error definition equations, summation constraints, and accumulation functions. The equations of the model are detailed below and the notation follows the convention that variable names are

[7] The program is available from the authors on request and, after further testing and extensions (see conclusion), will be made available on the CPRC website (www.chronicpoverty.org) by the end of the year.

written in uppercase and parameter names in lowercase. The values of variables in the base period are given as the variable name suffixed with a zero.

3.2.1 Entropy Objective

The entropy objective equation gives the value of the cross-entropy metric, *CENTROPY*, as a function of the distance between the estimated transition matrices and their prior values and the estimated error weights and their prior values. The cross entropy metric minimises the distance between the annual transition probabilities, *PV* and the prior annual probability values, *PV2*, the distance between the estimated survey period transition matrix, *PVKK* and the observed survey period transition matrix, *PVKK0*, and the error weights, *W*, *W2*, and *W3* and their prior values, *u*, *u2* and *u3*.

$$
\begin{aligned}
CENTROPY = & \sum_{i,j,t} PV_{i,j,t} * \log\left(\frac{PV_{t,i,j}}{PV2_{i,j}}\right) \\
& + \sum_{i,j} PVKK_{i,j} * \log\left(\frac{PVKK_{i,j}}{PVKK0_{i,j}}\right) \\
& + \sum_{d,j,t} W_{d,j,t} * \log\left(\frac{W_{d,j,t}}{u_{d,j,t}}\right) \\
& + \sum_{d,j,t} W2_{d,j,t} * \log\left(\frac{W2_{d,j,t}}{u2_{d,j,t}}\right)
\end{aligned}
\qquad (1)
$$

3.2.2 System Dynamics

The number of households in each category of the transition matrix in the next period, *XV*, is equal to the number of households in each category in the present period multiplied by the annual transition probability matrix, *PV*, where the time subscript on the transition matrix corresponds to the first year of the transition between *t* and *t+1*.

$$XV_{j,t+1} = \sum_{i} XV_{i,t} * PV_{t,i,j} \qquad (2)$$

$$TFLW_{t,i,j} = XV_{i,t} * PV_{t,i,j} \qquad (3)$$

The annual transition flow matrix, *TFLW*, is the number of people who transition between in category in each year; given by the multiplication of the category totals, *XV*, and the transition matrix, *PV*.

3.2.3 Knowledge

The three equations in the knowledge section of the model that reflect the amount of observed data available to the modeller. Equation (4) defines the number of households in each category as equal to the observed number of households, x, plus an error term pertaining to the category totals, *EHAT*.

$$XV_{j,t} = x_{j,t} * e^{EHAT_{j,t}} \qquad (4)$$

The specification of equation (4) allows the possibility that the numbers of households in each category, where known, may be measured with error. The total number of households may also be measured with error, *EHAT2*,

$$\sum_j XV_{j,t} = \sum_j x_{j,t} * e^{EHAT2_{j,t}} . \qquad (5)$$

The model also allows for the possibility that the cell frequencies of the observed survey transition matrix, *TFLWK0*, are measured with error, *EHAT3*,

$$TFLWK_{i,j} = TFLWK0_{i,j} * e^{EHAT3_{i,j}} . \qquad (6)$$

3.2.4 Error Definitions

The error term on the number of households in each category, *EHAT*, the error term on the total number of households, *EHAT2*, and the error associated with the cell frequencies are defined as the endogenously measured weights, *W*, *W2* and *W3*, multiplied by the support sets, *v*, *v2* and *V3*,

$$EHAT_{j,t} = \sum_d W_{d,j,t} * v_{d,j,t} , \qquad (7)$$

$$EHAT2_t = \sum_d W2_{d,t} * v2_{d,t} , \qquad (8)$$

$$EHAT3_{i,j} = \sum_d W3_{d,i,j} * v3_{d,i,j} . \qquad (9)$$

3.2.5 Summation Constraints

The optimisation of the entropy model is subject to constraints regarding the transition probabilities and the weights on the error terms. The estimated transition probabilities, *PV*, and estimated prior probabilities, *PV2*, must sum to one across the columns of the transition matrix, for all time periods,

$$\sum_j PV2_{i,j} = 1 \qquad \forall i\,. \tag{10}$$

$$\sum_j PV_{t,i,j} = 1 \qquad \forall i,t \tag{11}$$

Note that the matrix of prior probabilities, *PV2*, is endogenous and therefore determined by the model but is not defined over time, t. Under cross entropy optimisation, the stationary nature of the prior values allows the estimated transition probabilities, *PV*, to differ from the constant prior values, *PV2*, to the extent that such deviations are justified by the data. The specification of a constant prior reflects the intuition that households are unlikely to make significant transitions between expenditure categories on an annual basis.

The weights on the error terms must also sum to one across the number of dimensions, d,

$$\sum_d W_{d,j,t} = 1 \qquad \forall j,t\ , \tag{12}$$

$$\sum_d W2_{d,t} = 1 \qquad \forall t\,, \tag{13}$$

$$\sum_d W3_{d,i,j} = 1 \qquad \forall t\,. \tag{14}$$

3.2.6 Accumulation Functions

Five accumulation functions are specified within the model. The cumulative flow matrix for years between the survey points, *TFLWK*, is defined as the number of households in each category in the first year (ttstart) multiplied by the estimated transition probability matrix for the entire survey period, *TFLWK*,

$$TFLWK_{i,j} = \sum_{ttstart} XV_{i,tp} * PVKK_{i,j} \ . \qquad (15)$$

The cumulative probability transition matrix, *PVK*, is set equal to the annual transition matrix in the first year of the survey period (*ttstrt*),

$$PVK_{ttstrt,i,j} = PV_{ttstrt,i,j} \ . \qquad (16)$$

Equation (19) captures the accumulation of the annual transition probabilities through the survey period (*ttmid*),

$$PVK_{t+1,i,j} = \sum_{ip} PVK_{t,i,ip} * PV_{t+1,ip,j} \quad when\ t = ttmid(t)\ and\ not\ ttend(t) \ . \qquad (17)$$

The number of households in each category in the final year is given by multiplying the number of households in each category in the first year of the survey period by the survey transition matrix,

$$XVK_{j} = \sum_{i,ttstrt(t)} XV_{i,tp} * PVKK_{i,j} \ . \qquad (18)$$

Finally, the estimated survey transition matrix, *PVKK*, equals the accumulated annual transition matrix

$$PVKK_{i,j} = PVK_{ttend,i,j} \ . \qquad (19)$$

4 Application to South Africa and Vietnam

Having outlined the data and described the cross entropy method for estimating annual transition probabilities, we now apply it to three waves of panel data from KwaZulu-Natal in South Africa and from Vietnam. In the basic model used here, the annual transition matrices between the survey waves are estimated for each two-wave panel separately. This approach is used to allow for differences in transition patterns between periods and also allows for the estimation of two endogenous priors. The estimation of the transition model allows for a multiplicative error of 5% on the number of households in each transition category carries an error of 5% while the total number of households in the survey is assumed to be measured without error. A larger error of 30% is assumed for on the interpolated number of households in each transition category in the interim years.

We first present estimates of annual poverty transition matrices and associated poverty and mobility statistics for the KwaZulu-Natal in South Africa from 1993 to 2004, and then present comparable results for Vietnam between 2002 and 2006. A sensitivity analysis of the robustness of these results to the measurement errors assumed is conducted in Section 5.

4.1 South Africa

For KwaZulu-Natal we have three waves of survey data in 1993, 1998 and 2004. Estimates of the annual poverty transition probabilities for the years spanned by these waves are shown in Figure 1. The full annual transition matrices with the seven expenditure categories described in Table 1 are given in the appendix.

The five years spanned by the first and second waves of the KIDS panel are shown in the top part of Figure 1. Row 1 of the first poverty transition matrix indicates that 77% of poor households in KwaZulu-Natal remained poor while 23% of poor households in KwaZulu-Natal escaped poverty during 1993. Row 2 of the same matrix shows that 76% of non-poor households remained non-poor, while 24% of non poor households fell into poverty in this year. This period of mobility during the first year of the first panel (which corresponds to the ending of apartheid and lead-up to South Africa's first multi-racial elections survey wave was followed by two years in which there was very little movement between the poor and non poor states. Households that were poor in 1994 and 1995 remained poor in the next period while the vast majority (94% and 97%, respectively) of households that were non poor in these years non poor in the next period. There was some movement out of poverty again in 1996 and 1998 with little movement in 1997.

Figure 1 Annual Transition Probabilities, KwaZulu-Natal 1993-2004

1993	Poor	Non Poor
Poor	0.77	0.23
Non Poor	0.24	0.76

1994	Poor	Non Poor
Poor	1.00	0.00
Non Poor	0.06	0.94

1995	Poor	Non Poor
Poor	1.00	0.00
Non Poor	0.03	0.97

1996	Poor	Non Poor
Poor	0.80	0.20
Non Poor	0.23	0.77

1997	Poor	Non Poor
Poor	1.00	0.00
Non Poor	0.05	0.95

1998	Poor	Non Poor
Poor	0.86	0.14
Non Poor	0.19	0.81

1999	Poor	Non Poor
Poor	0.94	0.06
Non Poor	0.11	0.89

2000	Poor	Non Poor
Poor	0.96	0.04
Non Poor	0.05	0.95

2001	Poor	Non Poor
Poor	0.96	0.04
Non Poor	0.08	0.92

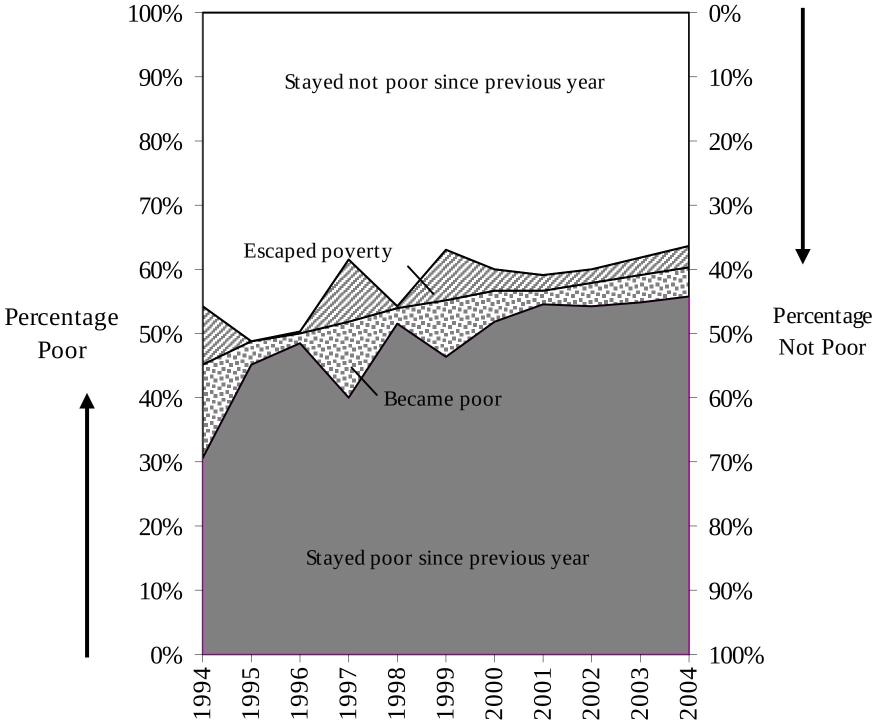
2002	Poor	Non Poor
Poor	0.95	0.05
Non Poor	0.10	0.90

2003	Poor	Non Poor
Poor	0.94	0.06
Non Poor	0.11	0.89

The six years spanned by the second and third survey waves are characterised by a similar degree of movements out of poverty of between 4% and 6%. Movements into poverty were also similar during these years, ranging from 5% to 11%. Only the transitions in the first year of the second panel period exhibits a greater degree of movement with 14% of households moving out of poverty and 19% of non poor households moving back into poverty.

The annual poverty dynamics of KwaZulu-Natal households between 1993 and 2004 as estimated in these poverty transition matrices can also be shown graphically (Figure 2). The dark shaded area towards the bottom of this figure shows the percentage of households who remained poor since the previous year, while the unshaded area at the top shows the percentage of households who stayed non poor since the previous year. The cross-hatched and spotted areas in the middle of the figure, correspond to the percentage of households who escaped from poverty or became poor in consecutive survey years. The central line shows that the percentage of households with expenditures below the poverty line steadily increased from 40% in 1993 to 60% in 2004.[8] The period is characterised by three peaks in poverty mobility in 1994, 1997 and 1999. Then from 2000 onwards, movements in and out of poverty are relatively constant.

Figure 2: Poverty Dynamics, KwaZulu-Natal, 1993-2004



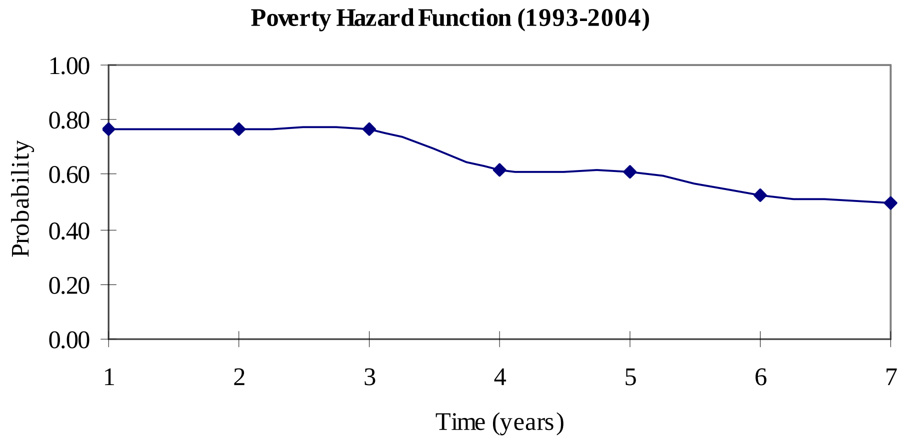
The CPRC's primary interest is in those who stay poor for extended periods of time (CPRC, 2004). One way to assess this is to examine the probabilities of whether an initially poor household will remain poor in successive periods. The probability of a poor household remaining poor in successive periods can be shown using a poverty hazard function, which compounds the annual probabilities of staying poor in the top left-hand corners of the poverty transition matrices in Figure 1[9]. Figure 3 shows the estimated poverty hazard function for KwaZulu-Natal between 1993 and 2004. Each point on the functions corresponding to the likelihood of a household remaining below the poverty

[8] Note that this refers only to households in the KIDS panel, not to KwaZulu-Natal as a whole, let alone the whole of South Africa!

[9] Except in situations where there is no movements out of poverty, this sequence of probabilities will decline over time, with the steepness of the function providing a graphical representation of how quickly households (people) are moving out of poverty.

line (categories 1 and 2 of the extended transition matrix) in consecutive periods given that the household was poor in the previous period.

Figure 3: Poverty Hazard Functions for KwaZulu-Natal



The probability of a poor household in the first year of each survey wave still being poor in 1998 is 0.61, and this probability declines to 0.49 by 2003. If we compare the probability of a household remaining poor in the four years between 1993 and 1998 (0.61) with that for being persistently poor in the five years between 1998 and 2004 (0.66), this shows that it is slightly more difficult for a household to escape persistent poverty in the later than the earlier panel period.

Table 4 shows the poverty headcount, poverty gap and Shorrocks mobility index for the KIDS panel data between 1993 and 2004. The poverty headcount measure rises consistently from 40% of households in 1993 to 60% by 2004. The poverty gap measure, which shows the depth of poverty, also rises consistently across the period, from 0.13 in 1993 to 0.31 in 2004. These trends are broadly consistent with the poverty measures calculated from the 1995 and 2000 South African Income and Expenditure Surveys as reported in Aguero *et al.* (forthcoming).[10] The value of the Shorrocks Mobility Index varies greatly between 0.09 and 0.96 suggesting that households are highly mobile in some years (1993, 1996, 1998) but immobile in other years (1995-1996, 1999-2004).

[10] Note that the poverty line (known as the Household Subsistence Level) for the analysis of the Income and Expenditure Survey is an absolute one, while the poverty line we used for KIDS is a relative one (set equal to the cut-off between the 39 and 40th percentiles of the expenditure distribution.

Although this pattern of mobility is consistent with the poverty mobility shown in Figure 2, such variations in mobility are surprising and require further examination.

Table 4 Poverty & Mobility Measures, KwaZulu-Natal, 1993-2004

	Poverty Headcount	Poverty Gap	Mobility Index[11]
1993	40%	0.13	0.96
1994	45%	0.16	0.12
1995	49%	0.17	0.09
1996	50%	0.19	0.47
1997	52%	0.21	0.13
1998	54%	0.22	0.76
1999	55%	0.24	0.23
2000	57%	0.26	0.14
2001	57%	0.26	0.19
2002	58%	0.28	0.19
2003	59%	0.30	0.20
2004	60%	0.31	-

The average number of years for a household to move between the seven expenditure categories in the extended transition matrices is given by the mean passage time matrices in Table 5 and Table 6. The value in each cell shows the average number of years it takes a household to move from category *i* to category *j* during the panel period. Between 1993 and 1998 the mean first passage times suggest low upwards mobility. On average, a poor household in expenditure category 1 or 2 takes between 21 and 26 years to move to the lowest non poor category (category 3). With mean passage times between 23 and 125 years, movements from poverty to higher welfare categories take even longer. Such long mean passage times are consistent with the type structural poverty suggested by Carter and May (2001), and by May and Woolard (2005). As observed in the estimated annual poverty transition matrices, there are also movements into poverty with mean first passage times suggesting that households are more likely to move into the upper of the two poor categories (category 2). The average time for a household to move from being non poor to being poor is between 7 and 14 years for group nearer to the poverty line (category 2) and 22 to 31 years for the lower group (category 1). So once a household has escaped poverty, it is more likely that should they become poor again, they will not return to being extremely poor. In general, the mean first passage time matrix for the first

[11] Year corresponds to first year of transition e.g. 2002 is mobility between 2002 and 2003.

KIDS panel is characterised by low passage times between the same state, indicating low levels of general mobility.

Table 5 Mean First Passage Times (1993-1998)

	Mean First Passage Time (years)						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1	6	4	26	27	58	125	117
2	20	2	21	23	54	120	113
3	27	10	6	15	48	108	96
4	22	7	19	9	47	119	108
5	26	9	14	23	16	114	99
6	29	12	10	22	29	43	88
7	31	14	7	21	43	66	31

A similar pattern is observed in the mean first passage time matrix for the second survey wave period, 1998-2004, Table 6. There is a strong tendency to remain in state (as indicated by the low values on the leading diagonal) or to move to lower welfare categories (as indicated by the low values in the cells just below the leading diagonal). Average transition times out of poverty are greater in the second survey wave period, with an average time to move out of poverty of 31 years for households in extreme poverty years and 24 years for households in the higher poverty category. A lack of upwards mobility is evident, as the average time to move into the next welfare group is between 8 years for households in welfare category 1 and 157 years for households in category 5. This contrasts sharply with the average transition time from category 5 into the upper poverty category of 12 years. So, as in the first panel period, households may move relatively quickly from being non poor to the upper poverty category, but it takes between 20 and 28 years for non-poor households to become extremely poor.

Table 6 Mean First Passage Times (1998-2004)

	Mean First Passage Time (years)						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1	3	8	31	47	67	175	75
2	15	3	24	42	62	171	68
3	20	11	8	33	55	163	56
4	22	11	10	13	59	164	63
5	24	12	21	24	18	157	52
6	28	17	20	14	43	31	64
7	27	16	19	20	48	149	14

4.2 Vietnam

The panel data for Vietnam is used to compute poverty transition matrices between 2002 and 2004, and between 2004 and 2006. The cross-entropy estimation process yields four annual poverty transitions matrices for 2002-03, 2003-04, 2004-05 and 2005-06 (Figure 4).

Figure 4 Annual Poverty Transition Probabilities, Vietnam, 2002-2006

2002	Poor	Non poor
Poor	0.63	0.37
Non poor	0.08	0.92

2003	Poor	Non poor
Poor	0.63	0.37
Non poor	0.06	0.94

2004	Poor	Non poor
Poor	0.61	0.39
Non poor	0.06	0.94

2005	Poor	Non poor
Poor	0.69	0.31
Non poor	0.04	0.96

The annual transition probabilities show significant movements out of poverty between 2002 and 2006. Furthermore, the movements in and out of poverty are similar for each year of the estimation period: with between 31% and 39% of households escaping poverty and between 4% and 8% of households moving back into poverty each year between 2002 and 2006. The high probabilities of remaining in the non poor state indicates that once households escaped poverty during this period, there is only a small probability that they will return to being poor. These dynamics are in line with the observed substantial fall in the national poverty headcount in Vietnam from 29% in 2002 to 16% 2006 (World Bank, 2007).[12]

The movement of households in and out of poverty between 2002 and 2006 in Vietnam is shown in Figure 5. The percentage of households that escape poverty (the cross-hatched area) is consistently higher than the number that move back into poverty the dotted area) over the estimation period. As a consequence the percentage of households that has stayed poor from one year to the next (the dark shaded area) has declined to 11% , while the percentage of household staying non poor (the unshaded area) has increased to 80%.

[12] Note that the same absolute poverty line, based on a cost-of-basic-needs approach and adjusted for inflation, is used in both World Bank (2007) and the estimation underlying .

Figure 5 Poverty Dynamics, Vietnam, 2002-2006

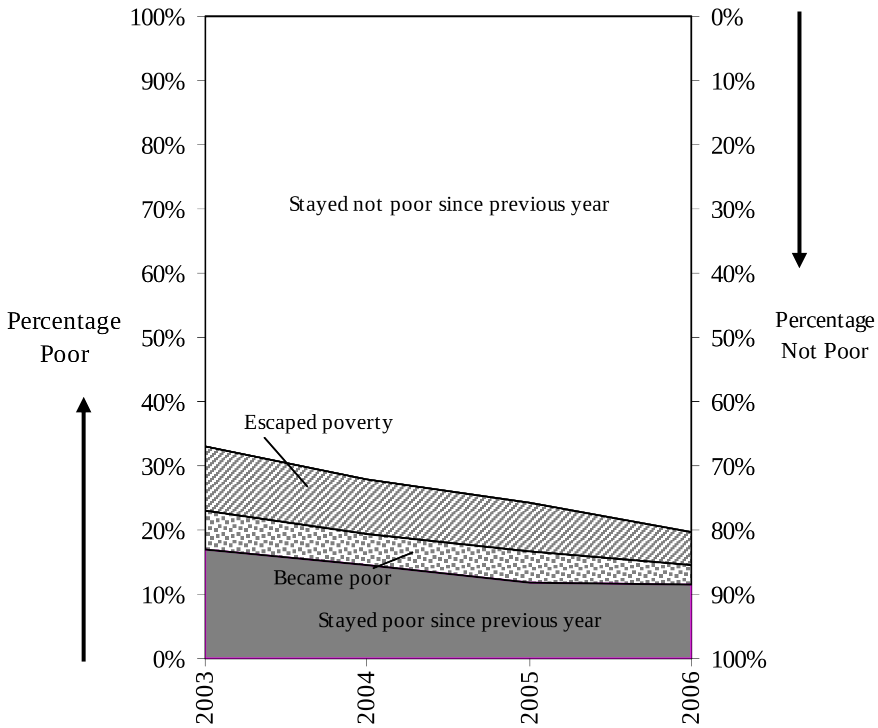


Figure 6 shows the poverty hazard functions for Vietnam between 2002 and 2006, shows the probability of remaining poor compounded over consecutive years. The slope of the poverty hazard function between 2002 and 2004 is slightly steeper than between 2004 to 2006 indicating that households were able to escape poverty more quickly during the period covered by the first two survey waves. The poverty hazard function for the entire survey period shows the probability of a poor household in 2002 still being poor in 2006 is just 0.17. While such rapid reduction in poverty, which is mirrored by the improvements in most other welfare indicators in Vietnam., it is important to recognise that there are still sections of the population, in particular the ethnic minority groups living in remote upland and mountainous areas who have not benefited to the same extent as the majority of the population (Baulch, Pham and Reilly, 2007).

Figure 6 Poverty Hazard Functions, Vietnam, 2002-06, Welfare Categories 1 & 2 (Poor)

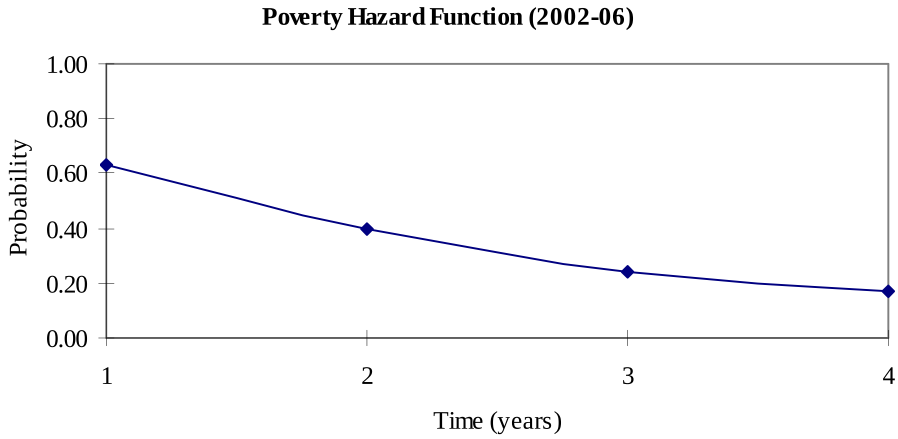


Table 7 shows the poverty headcount, poverty gap and mobility indices for 2002 to 2006. Headcount poverty falls from 27% of households in 2002 to 15% of households in 2006. The poverty gap falls from 0.08 to 0.04 over the same period.[13] Shorrocks' Mobility Index has generally high values, and shows that Vietnamese household were most mobile in 2002 and least mobile in 2003. This is consistent with the commodity boom experienced by several of the export crops, such as coffee and rice, produced by Vietnamese smallholders. The average annual mobility index for the period is 0.69.

Table 7 Poverty & Mobility Measures, Vietnam, 2002-2006

	Poverty Headcount	Poverty Gap	Mobility Index[14]
2002	27%	0.08	0.73
2003	23%	0.07	0.64
2004	19%	0.06	0.72
2005	17%	0.05	0.68
2006	15%	0.04	-

[13] Note that these poverty measures should not be expected to correspond to national level estimates, as they have been calculated using the sub-panel of the VHLSS which, by definition is not representative of Vietnam's population, Nonetheless, the estimated poverty headcounts from the panel are within 1 or 2 percentage points of the national poverty esitimates.

[14] Year corresponds to first year of transition e.g. 2002 is mobility between 2002 and 2003.

The average time taken for a household to move between each welfare category is shown by the mean first passage time matrices in Table 8 for 2002-2004 and Table 9 for 2004-2006. The mean passage time matrices for Vietnam suggest a high degree of upwards mobility. Table 8 shows that starting in 2002 a poor household (in expenditure categories 1 and 2) could be expected to move to a non-poor category (categories 3-8) in 5 to 8 years, and to reach the higher welfare categories (categories 6-8) in 10 to 12 years. Similar passage times are observed between 2004 and 2006, with poor households able to escape poverty in 6-8 years and able to reach the higher welfare categories in 11-13 years.

Table 8 Mean First Passage Times (years), Vietnam, 2002-2004

	Mean First Passage Time (Years)							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	107	7	8	11	12	18	21	20
2	256	9	5	8	10	16	21	19
3	264	18	5	7	8	14	20	17
4	268	21	7	6	8	12	18	15
5	269	23	9	8	8	10	16	13
6	269	22	10	8	8	10	16	12
7	269	24	11	10	4	11	15	11
8	272	25	13	11	10	11	15	5

Table 9 Mean First Passage Times (years), Vietnam, 2004-2006

	Mean First Passage Time (Years)							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	147	5	8	8	13	14	24	18
2	272	10	6	6	11	13	23	18
3	282	23	5	4	9	11	21	16
4	286	25	8	5	6	10	20	15
5	287	26	9	5	7	9	18	14
6	288	28	11	4	9	10	16	11
7	289	28	12	5	10	9	13	10
8	290	29	13	6	11	11	16	5

Figure 5 also shows low probabilities of moving back into poverty from the non poor states. This is corroborated by the high mean first passage times for households moving to the lowest expenditure group (category 1) in the passage time matrices. Households who in the ‘poorest’ non poor group (category 3) have an average return time to highest poor group (category 2) of 18 and 23 years. The mean first passage times for transitions into the poorest category 1 are even higher. Households who are non poor take at least 264 to 282 years to move into the poorest category. Overall, these matrices suggest that once households have escaped poverty in Vietnam they are unlikely to return to being poor.

5 Sensitivity Analysis

The annual poverty transition probabilities are estimated with different levels of error on the observed data and the unobserved data. The number of households in each category in the survey years and the frequencies of the survey transition matrix are observed data that are assumed to be measured with an error of 5%. The number of households in each category in non-survey years, generated using a linear time trend, is assumed to be measured with an error of 30%. The same errors are used for KwaZulu-Natal and Vietnam.

It is desirable to test the robustness of the annual transition probabilities estimates to the assumptions of the degree of measurement error in the observed and unobserved data. A systematic sensitivity analysis is conducted to quantify how much the annual transition probabilities of the poverty transition matrix vary under different error assumptions. Eight sensitivity analyses are conducted: the degree of error associated with the observed data is specified at 0%, 5% and 10% and the error associated with the unobserved data is specified at 0%, 30% and 60%. All error combinations are implemented which yields nine values for each cell of the poverty transition matrix of which the probability for the combination 5% observed error and 30% unobserved error is the main result against which the sensitivity results are compared. The full sensitivity results are included in the appendix and a summary measure, the Mean Absolute Percentage Error, is presented here to provide an overview of the sensitivity of the transition probability estimates to changes in the error assumptions.

The results of the sensitivity analysis for KwaZulu-Natal are shown in Figure 7 and Figure 8 for wave 1 (1993-1997)[15] and Figure 9 for wave 2 (1998-2004). The results

[15] The sensitivity analysis for 1993-1997 omits the combinations of 0% error in the unobserved data and 5% and 10% error in the observed data as the model cannot reach a feasible solution under these constraints.

for the off-diagonal probabilities are shown on a separate graph in wave 1 because of the small values of the initial probability estimates and hence the large changes in the probability values under sensitivity analysis. For a range of changes in the error on observed and unobserved data of between 0% and double the initial error value, the probabilities of remaining in state in KwaZulu-Natal during the first survey wave period vary by a maximum of 24% from the estimates presented in this paper. In contrast, the small base values in the off-diagonal probabilities lead to higher levels of variation in the probability values of up to 4449%. This mean absolute percentage error is for transition between Poor and Non Poor states in 1997 and is omitted from the graph to avoid scaling problems. The magnitude of this value is due to two factors: a near 0 transition probability value under the standard assumption of 5% observed data error and 30% unobserved data error and probability estimates of between 0.17 and 0.24 under an assumption of 0% observed data error and varying levels of unobserved data error.

Figure 7 Sensitivity Analysis, KwaZulu-Natal, Remaining in State, 1993-1998

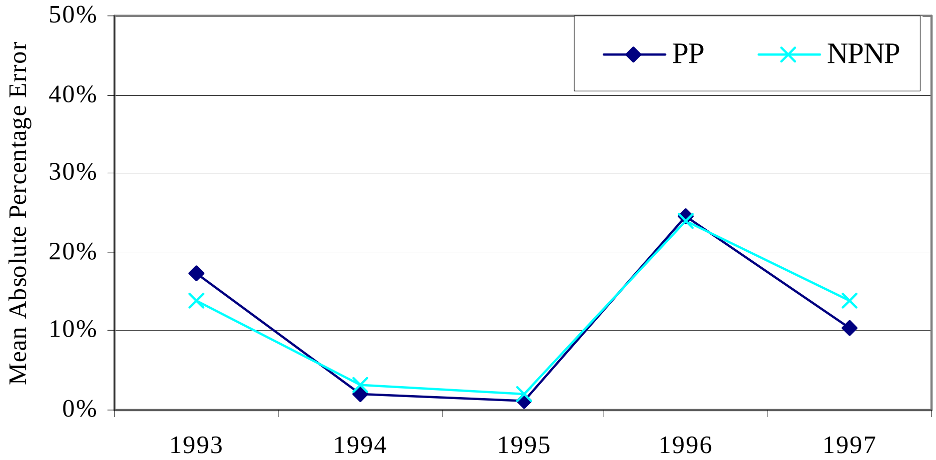
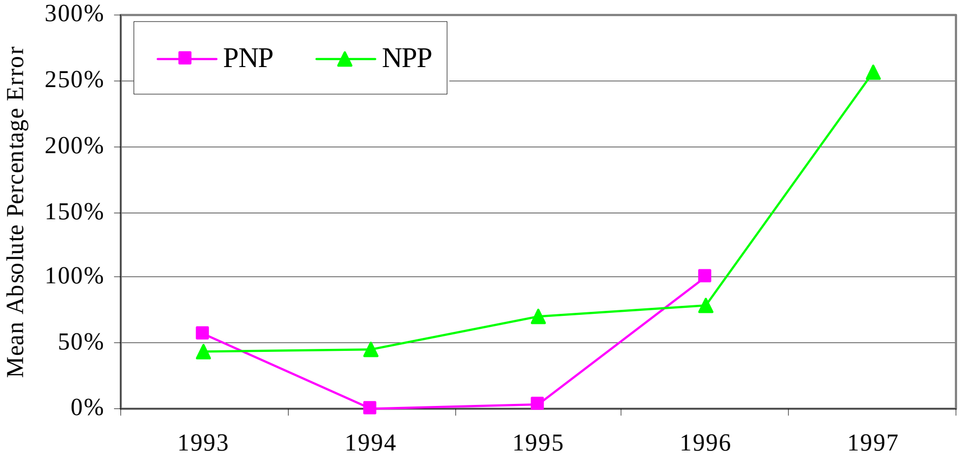
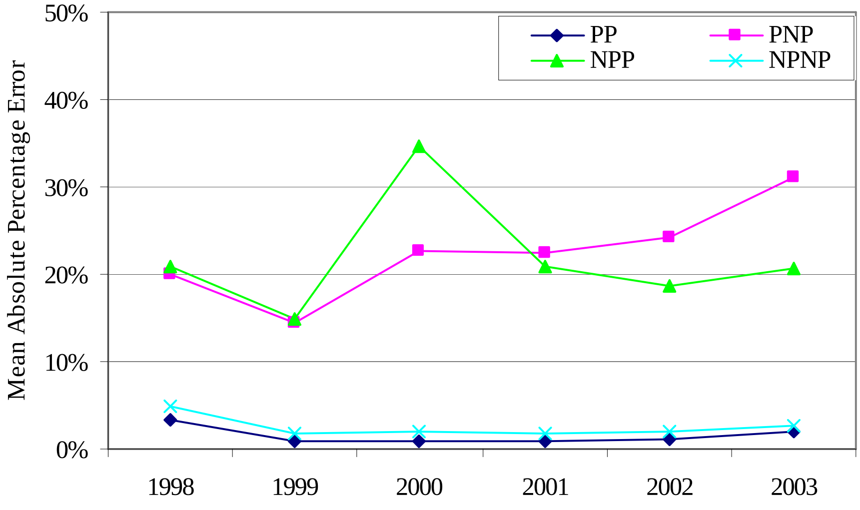


Figure 8 Sensitivity Analysis, KwaZulu-Natal, Moving Between States, 1993-1998



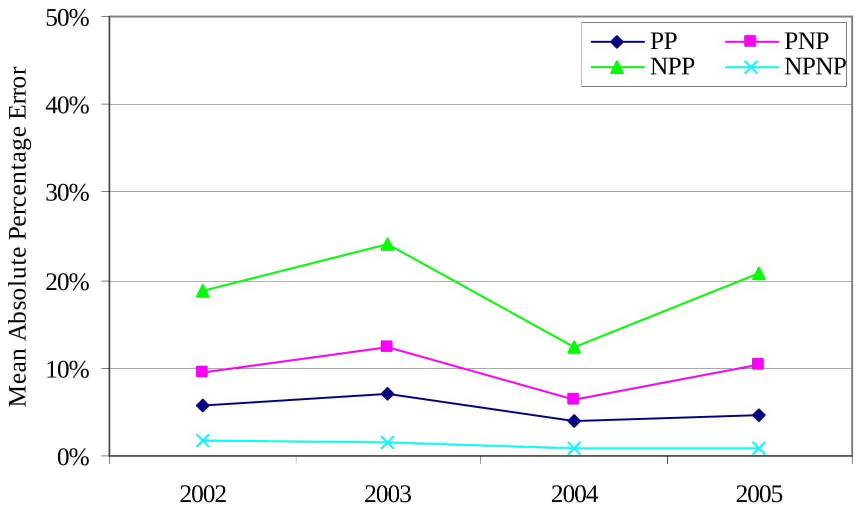
The sensitivity results for wave 2 are shown in . The probabilities of remaining in state vary by an average maximum of 5%. The probabilities of moving between Poor and Non Poor states are more sensitive to changes in the underlying error assumptions because of the small initial values. The off-diagonal probabilities vary by an average maximum of 35% for the period 1998-2004.

Figure 9 Sensitivity Analysis, KwaZulu-Natal, 1998-2003



The sensitivity results for both survey wave periods for Vietnam are shown in Figure 9. Again, the off-diagonal probabilities are more sensitive to changes in the underlying error assumptions than the probabilities of remaining in state. The probabilities of remaining in state vary by up to 7% over the survey period and the probabilities of moving between Poor and Non Poor states vary by up to 24%.

Figure 10 Sensitivity Analysis, Vietnam, 2002-2006



6 Caveats & Extensions

The annual poverty transition matrices estimated using the transition model presented in this paper are epistemic; they pertain only to the periods in question. Thus probability estimates cannot be extrapolated to other time periods unless the transition probabilities are believed to represent well-behaved ergodic distributions. Also, the poverty dynamics identified using the cross entropy estimation procedure can only be used to infer wider population poverty dynamics if the extended transition matrices on which they are based are truly representative. The transition probability estimates generated by the transition model should therefore be used to describe the households or individuals included in the survey, for the time period covered by the survey.

The assumption that the number of households in each welfare category follows a deterministic trend (with error) between the known survey year data is also an assumption that can be relaxed. An extension to the model presented in this paper is to use other (secondary) sources of information on poverty, inequality and/or growth as priors, which inform the number of households to be expected each transition category in the interim years. For example, the poverty headcount ratios from other poverty monitoring surveys could be incorporated for interim years, or estimates of inequality could be utilised to improve our estimates of poverty transitions.[16] A further extension would be to include the impact of key macroeconomic variables (e.g., growth rate) on poverty rates and the transition process using poverty-growth elasticities.[17] Such extensions do, however requires substantial additions to the transition modelling procedure, and will be the subject of further research

7 Conclusions

The number of panel surveys available to researchers is increasing. Such longitudinal data provide valuable information about how living standards evolve over time, and the transitions made between welfare groups for households and individuals during the period spanned by consecutive panel waves. It is desirable however to understand the dynamics that exist in the years between each survey wave and to be able to compare these measures across panels that use different welfare measures and which span different periods. A method for estimating annual transition matrices from such periodic survey data is presented in this paper. The method uses a cross entropy estimation technique with noise which is ideally suited to situations of limited data encountered in many developing countries.

The estimation technique was applied to panel data sets for KwaZulu-Natal in South Africa between 1993 and 2004 and to Vietnam from 2002 to 2006. In each case the annual transition matrices are estimated from 3 waves of survey data and poverty was estimated using an expenditure based welfare measure. The poverty transitions for South Africa during this period indicate strong rigidity in which it is difficult for households to escape poverty. In contrast, the estimated transition probabilities for Vietnam show a high level of upwards mobility in which many households are able to escape poverty. The poverty transition matrices for Vietnam also show that once households have escaped poverty, they are likely to remain non poor. These results are corroborated by mean first

[16]Note that the poverty headcount data used should be calculated using the same poverty line and welfare measures as the survey data on which the transition matrices are calculated. This is not the case with either the October Households Surveys in South Africa, or the administrative poverty monitoring system in Vietnam.

[17] Note that poverty-growth elasticities are themselves related the inequality in the region of the poverty line, and can be derived either analytically (by differentiating the cumulative distribution function at the poverty line) or econometrically.

passage times between states which indicate a high degree of upwards mobility. They are also consistent with the recent economic history of South Africa and Vietnam.

The estimation technique presented in this paper is flexible and can easily be extended to other data sets. In future work, we will be extending the method to incorporate additional information on poverty, inequality and growth, which should substantially improve its predictive power. Once this is done, we believe that this approach will provide a useful addition and powerful to current methods of analysing poverty dynamics and economic mobility using panel data.

References

Aguero, J., Carter, M,. and May, J, forthcoming, "Poverty and inequality in the first decade of South Africa's democracy: wht can be learnt from panel data from KwaZulu-Natal?", *Journal of Southern African Economies*

Baulch, B., forthcoming, "Understanding poverty dynamics and economic mobility' in Shepherd, A. and Moore, K. (eds*) Addressing Chronic Poverty: Concepts, Cross-Cutting Issues and Country Contexts*, Palgrave

Baulch, B. and Hoddinott, J., 2000, 'Economic mobility and poverty dynamics in developing countries', *Journal of Development Studies* 36(6): 1-24

Baulch, B., Pham, H. and Reilly, B., 2007, 'Ethnicity and household welfare in Vietnam: empirical evidence from 1993 to 2004', *Mimeo*, Institute of Development Studies, University of Sussex

Carter, M. and Barrett, C., 2005, 'The economics of poverty traps and persistent poverty: an asset based approach', *Journal of Development Studies* 42(2): 178-199

Carter, M., and May, J., 2001, 'One kind of freedom: poverty dynamics in post-apartheid South Africa', *World Development*, 29(12): 1987-2006

Cappellari, L., and Jenkins, S., 2002, 'Who stays poor? Who becomes poor? Evidence from the British Household Panel Survey', *Economic Journal* 112: C60-67.

Chant, L., 2008, 'The Macroeconomic Impact of HIV/AIDS in South Africa', *DPhil thesis*, Department of Economics, University of Sussex.

Cowell, F,. and Schulter, C., 1998, "Income mobility: a robust approach", *STICERD Discussion Paper 37*, London School of Economics

CPRC, 2004, *Chronic Poverty Report, 2004-05*, Manchester: Chronic Poverty Research Centre and Institute of Development Policy and Managment

Deaton, A., 1997, *The Analysis of Household Surveys: A Microeconometric Approach to Policy*, Baltimore: John Hopkins University Press.

Dercon, S. and Shapiro, 2007,'Moving on, staying behind, getting lost: lessons on poverty mobility from longitudinal data', GPRG Working Paper 75.

Fields, G., 2001, *Distribution and Development: A New Look at the Develong World*, New York: Russell Sage Foundation

Glewwe, P., Agrawal, N., and Dollar, D. (eds), 2004, *Economic Growth, Poverty and Household Welfare in Vietnam*, World Bank Regional and Sectoral Studies, Washington DC: World Bank.

Golan, A., Judge, G., and Miller, D., 1996, *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, Chichester: John Wiley & Sons.

Jaynes, E.T., 1957, *Information Theory and Statistical Mechanics*. The Physical Review, 106(4), 620-630.

Lee, T. C. and Judge, G., 1996, "Entropy and cross entropy procedures for estimating transition Pprobabilities from aggregate data" in Bayesian Conference in Statistics and Econometrics, *Essays in Honour of Arnold Zellner*, John Wiley: New York.

May, J. and Woolard, I.., 2007, 'Poverty traps and structural poverty in South Africa: reassessing the evidence from KwaZulu-Natal', *CPRC Working Paper 82*, Manchester; Chronic Poverty Research Centre

Morley, S., Robinson, S., and Harris, R., 1998, 'Estimating income mobility in Colombia using maximum entropy econometrics', Washington: *TMD Discussion Paper 26*, International Food Policy Research Institute.

Nguyen, T. and Hansen, H. (eds.), 2007, *Market, Policy and Poverty Reduction in Vietnam,* Hanoi: Vietnam Academy of Social Sciences/National Political Publishing House.

Petersson, H., 2003, 'Recommendations regarding the sampling design of the Vietnam Household Living Standards Surveys' Mimeo, Statistics Sweden

Phung, D.T. and Nguyen, P. 2007, 'Basic information on the Vietnam Household Living Standards Survey (VHLSS) 2002 and 2004' in Nguyen, T. and Hansen *op cit.*

Pincus, J. and Sender, J. 2006, 'Quantifying poverty in Viet Nam: who counts?' Mimeo, United Nations Development Program, Hanoi

Shannon, C.E., 1948. *The Mathematical Theory of Communication*. Bell System Technical Journal, 27(379-423), 623-656.

Shorrocks, A., 1997, 'The measurement of mobility', *Econometrica* 46(5): 1013-24

Stephens, A., 1999, Climbing out of poverty, falling back in: measuring the persistence of poverty over multiple spells', *Journal of Human Resources* 34(?):557-88

Woolard & Klasen, S., 2005, 'Determinants of income mobility and household poverty dynamics in South Africa', *Journal of Development Studies* 41(5): 865-897

Zellner, Arnold, 1988, "Optimal Information Processing and Bayes Theorem." *American Statistician* Vol. 42, pp. 278-284.

World Bank, 2007, *Vietnam Development Report 2008: Social Protection*, Hanoi: World Bank

Appendix 1: Diagnostic Statistics

Normalised entropy statistics provide a measure of the informational content of the data based on the distance of the posterior from the prior probability distributions for both the transition probabilities and the error terms. Golan et al. (1996) introduce a range of entropy measures including a normalised measure of entropy for each Markov state (S_i),

$$S_i = \frac{-\sum_{j,t}\left(PV_{t,i,j} * \log PV_{t,i,j}\right)}{-\sum_{j,t}\left(PV2_{i,j} * \log PV2_{i,j}\right)} \qquad PV, PV2 > 0 .$$

A normalised entropy measure for the system (S) including all states,

$$S = \frac{-\sum_{i,j,t}\left(PV_{t,i,j} * \log PV_{t,i,j}\right)}{-\sum_{i,j,t}\left(PV2_{i,j} * \log PV2_{i,j}\right)} \qquad PV, PV2 > 0 ,$$

where $PV_{i,j,t}$ is the estimated transition probabilities of moving from state i to state j in time t, and $PV2_{i,j,t}$ is the estimated prior transition probabilities.

The normalised entropy measures take a value between 0 and 1, where a value of 1 means indicates that the data has no information content above that of the prior such that the posterior distribution equals the prior distribution. Conversely, a value of 0 means that the posterior consists entirely of information from the data and the prior values do not contribute to the values of the estimated probabilities or weights. In the case of maximum entropy estimation which does not make use of prior information, the denominator in each measure is replaced with the log of the uniform distribution ($\log\left(\frac{1}{K}\right)$) and the measures indicate the information content of the data over the uniform distribution.

Soofi (1992) defines a corresponding measure of normalised entropy, the Information Index (I),

$$I = 1 - S .$$

The Entropy Ratio (ER) statistic provides a further test of the information content of the data. Under the null hypothesis that the distribution of prior probabilities is equal to the distribution of estimated probabilities,

$$H_0 : PV_{tij} = PV2_{ij} \, ,$$

the entropy ratio is distributed as a χ^2_{K-1} distribution. Golan and Vogel (2000) define the entropy ratio statistic for the cross-entropy formulation as,

$$ER = 2\left(-\sum_{i,j} PV2_{i,j} \log PV2_{i,j} \right) * \left[1 - \left(\frac{-\sum_{i,j} PV_{i,j} \log PV_{i,j}}{-\sum_{i,j} PV2_{i,j} \log PV2_{i,j}} \right) \right].$$

The null hypothesis is rejected if the entropy ratio statistic is greater than the χ^2_{K-1} critical value, implying that the information content of the data is significant relative to the prior values or that the prior values are not consistent with the data.

Alternatively, a χ^2_{K-1} goodness of fit test can be used, defined as in Golan and Vogel (2000) as,

$$\chi^2_{k-1} = \sum_{i,j} \frac{1}{PV2_{i,j}} \left(PV_{i,j} - PV2_{i,j} \right)^2 \, ,$$

with the same null hypothesis as the entropy ratio test. If the calculated χ^2 statistic exceeds the critical value, the null hypothesis is rejected and the estimated probabilities differ significantly from the prior values at the given confidence level.

Appendix 2: Estimated Transition Matrices (KwaZulu-Natal)

1993

	c1	c2	c3	c4	c5	c6	c7
c1	0.00	1.00	0.00	0.00	0.00	0.00	0.00
c2	0.24	0.49	0.01	0.18	0.06	0.02	0.00
c3	0.00	0.39	0.22	0.31	0.00	0.02	0.05
c4	0.03	0.08	0.34	0.20	0.30	0.00	0.04
c5	0.02	0.21	0.51	0.00	0.17	0.00	0.10
c6	0.03	0.03	0.27	0.00	0.29	0.14	0.24
c7	0.00	0.00	0.44	0.00	0.00	0.55	0.02

1994

	c1	c2	c3	c4	c5	c6	c7
c1	0.95	0.05	0.00	0.00	0.00	0.00	0.00
c2	0.00	1.00	0.00	0.00	0.00	0.00	0.00
c3	0.00	0.00	0.95	0.00	0.01	0.00	0.04
c4	0.10	0.03	0.00	0.87	0.00	0.00	0.00
c5	0.05	0.09	0.03	0.00	0.82	0.00	0.00
c6	0.00	0.00	0.12	0.00	0.02	0.85	0.00
c7	0.00	0.00	0.12	0.00	0.00	0.03	0.85

1995

	c1	c2	c3	c4	c5	c6	c7
c1	0.98	0.02	0.00	0.00	0.00	0.00	0.00
c2	0.03	0.97	0.00	0.00	0.00	0.00	0.00
c3	0.00	0.00	0.95	0.00	0.01	0.00	0.04
c4	0.08	0.01	0.00	0.90	0.00	0.00	0.00
c5	0.00	0.00	0.04	0.00	0.95	0.00	0.01
c6	0.00	0.00	0.11	0.00	0.02	0.87	0.00
c7	0.00	0.00	0.11	0.00	0.00	0.03	0.86

1996

	c1	c2	c3	c4	c5	c6	c7
c1	0.87	0.13	0.00	0.00	0.00	0.00	0.00
c2	0.13	0.61	0.25	0.02	0.00	0.00	0.00
c3	0.00	0.01	0.48	0.30	0.06	0.06	0.09
c4	0.07	0.57	0.00	0.37	0.00	0.00	0.00
c5	0.00	0.33	0.04	0.00	0.63	0.00	0.01
c6	0.00	0.00	0.18	0.00	0.38	0.44	0.00
c7	0.00	0.00	0.21	0.00	0.00	0.04	0.75

1997

	c1	c2	c3	c4	c5	c6	c7
c1	0.98	0.02	0.00	0.00	0.00	0.00	0.00
c2	0.00	1.00	0.00	0.00	0.00	0.00	0.00
c3	0.00	0.01	0.91	0.08	0.00	0.00	0.00
c4	0.00	0.19	0.00	0.81	0.00	0.00	0.00
c5	0.00	0.01	0.07	0.00	0.91	0.00	0.01
c6	0.00	0.00	0.12	0.00	0.18	0.70	0.01
c7	0.00	0.00	0.06	0.00	0.00	0.04	0.89

1998

	c1	c2	c3	c4	c5	c6	c7
c1	0.46	0.53	0.00	0.00	0.00	0.00	0.00

c2	0.23	0.57	0.09	0.04	0.02	0.01	0.04
c3	0.16	0.08	0.50	0.06	0.18	0.01	0.01
c4	0.00	0.10	0.47	0.38	0.00	0.05	0.00
c5	0.00	0.34	0.00	0.08	0.17	0.11	0.30
c6	0.00	0.00	0.00	0.62	0.38	0.00	0.00
c7	0.02	0.03	0.05	0.28	0.14	0.10	0.37

1999

	c1	c2	c3	c4	c5	c6	c7
c1	0.86	0.14	0.00	0.00	0.00	0.00	0.00
c2	0.10	0.81	0.07	0.01	0.00	0.00	0.00
c3	0.16	0.07	0.64	0.03	0.01	0.05	0.05
c4	0.00	0.03	0.22	0.75	0.00	0.00	0.00
c5	0.00	0.03	0.00	0.09	0.84	0.00	0.04
c6	0.00	0.00	0.00	0.07	0.03	0.89	0.00
c7	0.00	0.02	0.02	0.06	0.03	0.05	0.82

2000

	c1	c2	c3	c4	c5	c6	c7
c1	0.97	0.03	0.00	0.00	0.00	0.00	0.00
c2	0.00	0.93	0.05	0.01	0.01	0.00	0.00
c3	0.00	0.10	0.78	0.04	0.01	0.00	0.07
c4	0.00	0.03	0.21	0.76	0.00	0.00	0.00
c5	0.00	0.03	0.00	0.09	0.85	0.00	0.03
c6	0.00	0.00	0.00	0.01	0.00	0.99	0.00
c7	0.00	0.02	0.02	0.06	0.04	0.00	0.87

2001

	c1	c2	c3	c4	c5	c6	c7
c1	0.93	0.05	0.00	0.00	0.00	0.01	0.00
c2	0.15	0.79	0.04	0.01	0.01	0.00	0.00
c3	0.01	0.11	0.75	0.04	0.01	0.00	0.07
c4	0.00	0.08	0.20	0.72	0.00	0.00	0.00
c5	0.00	0.06	0.00	0.10	0.81	0.00	0.03
c6	0.00	0.00	0.00	0.00	0.00	1.00	0.00
c7	0.00	0.02	0.01	0.06	0.04	0.00	0.87

2002

	c1	c2	c3	c4	c5	c6	c7
c1	0.97	0.02	0.00	0.00	0.01	0.00	0.00
c2	0.11	0.80	0.05	0.01	0.02	0.00	0.01
c3	0.01	0.13	0.73	0.04	0.01	0.00	0.08
c4	0.00	0.14	0.17	0.69	0.00	0.00	0.00
c5	0.00	0.08	0.00	0.12	0.77	0.00	0.03
c6	0.00	0.00	0.00	0.00	0.00	1.00	0.00
c7	0.00	0.02	0.01	0.05	0.03	0.00	0.88

2003

	c1	c2	c3	c4	c5	c6	c7

c1	0.98	0.01	0.00	0.00	0.01	0.00	0.00
c2	0.07	0.83	0.07	0.01	0.02	0.00	0.01
c3	0.01	0.16	0.71	0.04	0.01	0.00	0.08
c4	0.00	0.18	0.15	0.68	0.00	0.00	0.00
c5	0.00	0.10	0.00	0.15	0.72	0.00	0.03
c6	0.00	0.00	0.00	0.00	0.00	1.00	0.00
c7	0.00	0.01	0.01	0.05	0.03	0.00	0.90

Appendix 3: Estimated Transition Matrices (Vietnam)

2002

	c1	c2	c3	c4	c5	c6	c7	c8
c1	0.36	0.54	0.06	0.00	0.00	0.00	0.05	0.00
c2	0.04	0.57	0.20	0.16	0.02	0.00	0.00	0.01
c3	0.01	0.15	0.50	0.17	0.15	0.00	0.00	0.02
c4	0.00	0.04	0.31	0.33	0.10	0.09	0.09	0.05
c5	0.00	0.02	0.15	0.12	0.15	0.36	0.11	0.08
c6	0.00	0.04	0.11	0.22	0.16	0.16	0.08	0.22
c7	0.01	0.00	0.02	0.00	0.57	0.00	0.13	0.28
c8	0.00	0.02	0.01	0.05	0.02	0.11	0.09	0.71

2003

	c1	c2	c3	c4	c5	c6	c7	c8
c1	0.82	0.11	0.03	0.00	0.00	0.00	0.05	0.00
c2	0.00	0.61	0.32	0.04	0.02	0.00	0.00	0.00
c3	0.00	0.14	0.45	0.24	0.16	0.00	0.01	0.02
c4	0.00	0.03	0.18	0.43	0.12	0.14	0.04	0.06
c5	0.00	0.00	0.23	0.18	0.14	0.20	0.13	0.12
c6	0.00	0.07	0.07	0.15	0.16	0.26	0.16	0.14
c7	0.00	0.00	0.02	0.00	0.62	0.00	0.14	0.22
c8	0.00	0.01	0.01	0.05	0.02	0.09	0.10	0.72

2004

	c1	c2	c3	c4	c5	c6	c7	c8
c1	0.30	0.61	0.00	0.00	0.00	0.04	0.00	0.04
c2	0.04	0.53	0.30	0.10	0.00	0.01	0.01	0.01
c3	0.01	0.11	0.50	0.20	0.03	0.12	0.01	0.01
c4	0.00	0.07	0.23	0.18	0.34	0.13	0.00	0.06
c5	0.00	0.04	0.14	0.18	0.35	0.08	0.11	0.10
c6	0.00	0.01	0.07	0.22	0.20	0.12	0.18	0.20
c7	0.00	0.01	0.02	0.28	0.00	0.16	0.29	0.24
c8	0.00	0.01	0.00	0.22	0.00	0.03	0.08	0.66

2005

	c1	c2	c3	c4	c5	c6	c7	c8
c1	0.64	0.32	0.00	0.00	0.00	0.02	0.00	0.02
c2	0.00	0.65	0.24	0.03	0.00	0.06	0.01	0.01
c3	0.00	0.07	0.51	0.34	0.03	0.03	0.01	0.01
c4	0.00	0.04	0.23	0.10	0.52	0.07	0.00	0.04
c5	0.00	0.03	0.13	0.24	0.24	0.24	0.05	0.08
c6	0.00	0.01	0.05	0.43	0.00	0.08	0.21	0.22
c7	0.00	0.01	0.03	0.18	0.00	0.25	0.30	0.23
c8	0.00	0.01	0.00	0.06	0.00	0.12	0.11	0.71

Appendix 4: Sensitivity Analysis

The results of the sensitivity analysis for Vietnam are shown below. The abbreviations PP, PNP, NPP and NPNP correspond to the poverty transition Poor-Poor, Poor-Non Poor, Non Poor-Poor and Non Poor-Non Poor. The measurement error associated with observed data is shown in the row heading of each matrix and ranges from 0%-10%. The measurement error associated with the unknown data is shown in the column headings and ranges from 0%-60%. In each matrix, the main result presented in the paper is shaded in grey for ease of comparison.

KwaZulu-Natal

1993

PP	0%	30%	60%
0%	0.99	0.94	1.00
5%	INFES	0.77	0.71
10%	INFES	0.71	0.72

PNP	0%	30%	60%
0%	0.01	0.06	0.00
5%	INFES	0.23	0.29
10%	INFES	0.29	0.28

NPP	0%	30%	60%
0%	0.05	0.10	0.04
5%	INFES	0.24	0.27
10%	INFES	0.28	0.27

NPNP	0%	30%	60%
0%	0.95	0.90	0.96
5%	INFES	0.76	0.73
10%	INFES	0.72	0.73

1994

PP	0%	30%	60%
0%	0.99	0.92	1.00
5%	INFES	1.00	0.99
10%	INFES	1.00	0.99

PNP	0%	30%	60%
0%	0.01	0.08	0.00
5%	INFES	0.00	0.01
10%	INFES	0.00	0.01

NPP	0%	30%	60%
0%	0.06	0.15	0.04
5%	INFES	0.06	0.10
10%	INFES	0.07	0.08

NPNP	0%	30%	60%
0%	0.94	0.85	0.96
5%	INFES	0.94	0.90
10%	INFES	0.93	0.92

1995

PP	0%	30%	60%
0%	0.98	0.95	1.00
5%	INFES	1.00	1.00
10%	INFES	1.00	1.00

PNP	0%	30%	60%
0%	0.02	0.05	0.00
5%	INFES	0.00	0.00
10%	INFES	0.00	0.00

NPP	0%	30%	60%
0%	0.07	0.04	0.04
5%	INFES	0.03	0.04
10%	INFES	0.00	0.05

NPNP	0%	30%	60%
0%	0.93	0.96	0.96
5%	INFES	0.97	0.96
10%	INFES	1.00	0.95

1996

PP	0%	30%	60%
0%	0.96	0.99	0.55
5%	INFES	0.80	1.00
10%	INFES	1.00	1.00

PNP	0%	30%	60%
0%	0.04	0.01	0.45
5%	INFES	0.20	0.00
10%	INFES	0.00	0.00

NPP	0%	30%	60%
0%	0.10	0.02	0.43
5%	INFES	0.23	0.03
10%	INFES	0.06	0.04

NPNP	0%	30%	60%
0%	0.90	0.98	0.57
5%	INFES	0.77	0.97
10%	INFES	0.94	0.96

1997

PP	0%	30%	60%
0%	0.80	0.83	0.76
5%	INFES	1.00	1.00
10%	INFES	1.00	1.00

PNP	0%	30%	60%
0%	0.20	0.17	0.24
5%	INFES	0.00	0.00
10%	INFES	0.00	0.00

NPP	0%	30%	60%
0%	0.27	0.27	0.34
5%	INFES	0.05	0.03
10%	INFES	0.04	0.03

NPNP	0%	30%	60%
0%	0.73	0.73	0.66
5%	INFES	0.95	0.97
10%	INFES	0.96	0.97

1998

PP	0%	30%	60%
0%	0.92	0.93	0.91
5%	0.83	0.86	0.85
10%	0.87	0.86	0.85

PNP	0%	30%	60%
0%	0.08	0.07	0.09
5%	0.17	0.14	0.15
10%	0.13	0.14	0.15

NPP	0%	30%	60%
0%	0.12	0.10	0.10
5%	0.22	0.19	0.20
10%	0.18	0.19	0.20

NPNP	0%	30%	60%
0%	0.88	0.90	0.90
5%	0.78	0.81	0.80
10%	0.82	0.81	0.80

1999

PP	0%	30%	60%
0%	0.94	0.94	0.93
5%	0.95	0.94	0.95
10%	0.96	0.94	0.95

PNP	0%	30%	60%
0%	0.06	0.06	0.07
5%	0.05	0.06	0.05
10%	0.04	0.06	0.05

NPP	0%	30%	60%
0%	0.10	0.09	0.10
5%	0.08	0.11	0.11
10%	0.07	0.10	0.09

NPNP	0%	30%	60%
0%	0.90	0.91	0.90
5%	0.92	0.89	0.89
10%	0.93	0.90	0.91

2000

PP	0%	30%	60%
0%	0.95	0.95	0.95
5%	0.98	0.96	0.96
10%	0.97	0.96	0.95

PNP	0%	30%	60%
0%	0.05	0.05	0.05
5%	0.02	0.04	0.04
10%	0.03	0.04	0.05

NPP	0%	30%	60%
0%	0.08	0.09	0.09
5%	0.05	0.05	0.05

NPNP	0%	30%	60%
0%	0.92	0.91	0.91
5%	0.95	0.95	0.95

	10%	0.06	0.07	0.07	10%	0.94	0.93	0.93
2001	PP	0%	30%	60%	PNP	0%	30%	60%
	0%	0.95	0.95	0.94	0%	0.05	0.05	0.06
	5%	0.98	0.96	0.95	5%	0.02	0.04	0.05
	10%	0.96	0.96	0.96	10%	0.04	0.04	0.04
	NPP	0%	30%	60%	NPNP	0%	30%	60%
	0%	0.10	0.09	0.14	0%	0.90	0.91	0.86
	5%	0.05	0.08	0.08	5%	0.95	0.92	0.92
	10%	0.07	0.08	0.08	10%	0.93	0.92	0.92
2002	PP	0%	30%	60%	PNP	0%	30%	60%
	0%	0.95	0.91	0.96	0%	0.05	0.09	0.04
	5%	0.98	0.95	0.95	5%	0.02	0.05	0.05
	10%	0.96	0.95	0.95	10%	0.04	0.05	0.05
	NPP	0%	30%	60%	NPNP	0%	30%	60%
	0%	0.09	0.16	0.10	0%	0.91	0.84	0.90
	5%	0.05	0.10	0.10	5%	0.95	0.90	0.90
	10%	0.07	0.10	0.09	10%	0.93	0.90	0.91
2003	PP	0%	30%	60%	PNP	0%	30%	60%
	0%	0.92	0.94	0.93	0%	0.08	0.06	0.07
	5%	0.90	0.94	0.95	5%	0.10	0.06	0.05
	10%	0.90	0.94	0.95	10%	0.10	0.06	0.05
	NPP	0%	30%	60%	NPNP	0%	30%	60%
	0%	0.15	0.11	0.11	0%	0.85	0.89	0.89
	5%	0.18	0.11	0.11	5%	0.82	0.89	0.89
	10%	0.18	0.11	0.10	10%	0.82	0.89	0.90

Vietnam

2002

PP	0%	30%	60%
0%	0.68	0.55	0.62
5%	0.61	0.63	0.58
10%	0.64	0.63	0.58

PNP	0%	30%	60%
0%	0.32	0.45	0.38
5%	0.39	0.37	0.42
10%	0.36	0.37	0.42

NPP	0%	30%	60%
0%	0.07	0.11	0.10
5%	0.09	0.08	0.11
10%	0.08	0.08	0.11

NPNP	0%	30%	60%
0%	0.93	0.89	0.90
5%	0.91	0.92	0.89
10%	0.92	0.92	0.89

2003

PP	0%	30%	60%
0%	0.59	0.75	0.67
5%	0.65	0.63	0.70
10%	0.61	0.63	0.70

PNP	0%	30%	60%
0%	0.41	0.25	0.33
5%	0.35	0.37	0.30
10%	0.39	0.37	0.30

NPP	0%	30%	60%
0%	0.07	0.03	0.04
5%	0.06	0.06	0.04
10%	0.07	0.06	0.04

NPNP	0%	30%	60%
0%	0.93	0.97	0.96
5%	0.94	0.94	0.96
10%	0.93	0.94	0.96

2004

PP	0%	30%	60%
0%	0.60	0.68	0.65
5%	0.58	0.61	0.62
10%	0.62	0.60	0.63

PNP	0%	30%	60%
0%	0.40	0.32	0.35
5%	0.42	0.39	0.38
10%	0.38	0.40	0.37

NPP	0%	30%	60%
0%	0.07	0.04	0.04
5%	0.07	0.06	0.06
10%	0.06	0.06	0.06

NPNP	0%	30%	60%
0%	0.93	0.96	0.96
5%	0.93	0.94	0.94
10%	0.94	0.94	0.94

2005

PP	0%	30%	60%
0%	0.71	0.61	0.63
5%	0.74	0.69	0.68
10%	0.67	0.71	0.67

PNP	0%	30%	60%
0%	0.29	0.39	0.37
5%	0.26	0.31	0.32
10%	0.33	0.29	0.33

NPP	0%	30%	60%
0%	0.03	0.05	0.05
5%	0.03	0.04	0.04
10%	0.04	0.03	0.04

NPNP	0%	30%	60%
0%	0.97	0.95	0.95
5%	0.97	0.96	0.96
10%	0.96	0.97	0.96