2019

IARIW-World Bank

Special IARIW-World Bank Conference "New Approaches to Defining and Measuring Poverty in a Growing World" Washington, DC, November 7-8, 2019

Investigating Welfare Dynamics with Repeated Cross Sections: A Copula Approach

Francois Bourguignon

Hai-Anh H. Dang

Paper Prepared for the IARIW-World Bank Conference

Washington, DC, November 7-8, 2019

Session 4B: Monitoring Poverty Over Time: Methods and Tools

Time:13:30 – 15:30, September 8

Investigating Welfare Dynamics with Repeated Cross Sections: A Copula Approach

Francois Bourguignon and Hai-Anh H. Dang*

November 2019

Abstract

Panel household surveys take a key role for tracking household welfare dynamics over time. Yet, the lack of such data has posed severe challenges to research and policy work concerning welfare dynamics in developing countries. Using copula functions, we propose a new approach to construct synthetic panels based on repeated cross sections that can offer substitute estimates in the absence of actual panels. We validate estimates of various measures of poverty mobility and income mobility based on the synthetic panels against those of the actual panels using the Vietnam Household Living Standards Surveys. We find that the copula-based synthetic panels yield estimation results that are encouragingly close to those based on the actual panels.

Keywords: chronic poverty, income mobility, consumption, cross sections, synthetic panels, household surveys

JEL Codes: C53, D31, I32, O15

* Bourguignon (francois.bourguignon@psemail.eu) is with Paris School of Economics; Dang (hdang@worldbank.org; corresponding author) is with the World Bank, and is also affiliated with Indiana university, IZA, and Vietnam's Academy of Social Sciences. We are grateful to the UK Department of International Development for funding assistance through its Knowledge for Change (KCP) programs.

I. Introduction

Panel household surveys are indispensable for tracking household welfare dynamics over time. Yet, such surveys are rarely available, particularly for developing countries for various reasons. These can range from lack of financial resources (i.e. it is costly to implement panel surveys), and technical capacity (i.e., certain levels of technical expertise are required to maintain nationally representative panel surveys) to logistical challenges (i.e., in fragile and conflict contexts, it is difficult to implement surveys and/ or track households over time). Even where panel data are collected, such data do not often provide nationally representative data. For example, two middle-income countries, China and India, recently collected some panel data but these panel data are not commonly employed to provide poverty estimates. The surveys that are used in these countries for this purpose—the China Household Income Project (CHIP) survey and the National Sample Survey (NSS)—are both cross-sectional surveys.

This data shortage has, in fact, been the main obstacle that hinders research on poverty mobility in developing countries.[1] More generally, researchers and policy makers face the same data challenge when trying to better understand the dynamics of other welfare outcomes other than poverty such as income mobility. Recent statistical methods have been developed to overcome this data challenge, such that synthetic panels can be constructed using only two rounds of repeated cross sections (Dang et al., 2014; Dang and Lanjouw, 2013).[2] These synthetic panels have been validated against actual panel data and employed to study poverty transitions in a number of developing countries, including countries in Latin Americas (Ferreira *et al.,* 2013; Cruces *et al.*,

[1] Still, the availability of cross section surveys should not be taken for granted, especially for poorer countries. A recent survey by Beegle *et al.* (2016) points out that just more than half (i.e., 27) of the 48 countries in Sub-Saharan Africa had two or more comparable household surveys for the period between 1990 and 2012. Even worse, Serajuddin et al. (2015) find that, over the period 2002- 2011, more than one-third (i.e., 57) of the 155 countries for which the World Bank monitors poverty data using the WDI database have only one poverty data point or no data at all.

[2] Bourguignon et al. (2004) provide an early attempt to construct synthetic panels but using more rounds of survey data.

2015; Vakis *et al.,* 2015), Europe and Central Asia (Cancho *et al.,* 2015), the Middle East and North Africa, Sub-Saharan Africa, and India.[3] Most recently, Bourguignon, Moreno, and Dang (2019) further extend this method in various directions to study income mobility and apply their method to data from Mexico.

In this paper, we build on existing methods to construct synthetic panels in an alternative and more general way using copulas. Copulas require fewer (parametric) assumptions and have been widely used in other fields such as engineering or finance, and more recently in economics to provide a flexible estimate of the joint distributions of different variables each with its marginal distribution.[4] These copula-based synthetic panels allow us to significantly extend the capability analysis of synthetic panels to examine general income (consumption) mobility, rather than just poverty and vulnerability mobility. Furthermore, we can offer estimates of various other absolute and relative mobility measures and indexes, such as income movement, positional movement, and non-anonymous growth incidence curves (GIC). In terms of modelling techniques, we offer selection tests for several commonly used copulas. Our method are also straightforward to implement in other similar contexts.

We validate our proposed method using both actual panel data and repeated cross sections from the Vietnam Household Living Standards Survey (VHLSS). We find that the copula-based

[3] Researchers at international organizations including the UNDP, the Asian Development Bank, and the OECD have also applied these methods for analysis of welfare mobility (UNDP, 2016; Jha *et al.*, 2018; OECD, 2018); see also OECD (2015) for an application by the OECD to study labor transitions in richer countries. A recent validation study by Herault and Jenkins (2019) find that these synthetic panel may work less well for data from richer countries such as Australia and the UK, but other validation results by Dang and Lanjouw (2013) and Garces-Urzainqui (2017) find encouraging results for data from the US and Thailand. Dang, Jolliffe, and Carletto (2019) and Dang (forthcoming) offer recent reviews of studies that employ synthetic panel techniques.
[4] For example, Bonhomme and Robin (2009) and Chetty et al. (2014) study earnings mobility in France and intergenerational mobility in the US using actual panel data; Foster and Rothbaum (2015) analyze intergenerational mobility for Mexico with synthetic panels. Trivedi and Zimmerman (2005) and Fan and Patton (2014) offer reviews of other applications of copulas in economics.

synthetic panels yield encouraging estimation results that are close to those based on the actual panels.

This paper consists of five sections. We provide an overview of the synthetic panel technique before discussing the new features with the copulas in the next section. We subsequently describe data in Section III, discuss estimation results in Section IV, and finally conclude in Section V.

II. Analytical Framework

Let y_{ij} represent household consumption or income in survey round j for household i, where $i= 1,\ldots, N$, and $j= 1$ or 2. Let x_{ij} be a vector of time-invariant household characteristics that are observed in both survey rounds. Subject to data availability, these characteristics can include such variables as sex, ethnicity, religion, language, place of birth, and parental education as well as variables that can be converted into time-invariant versions based, for example, on information about household heads' age and education. The vector x_{ij} can also include time-varying household characteristics if retrospective questions about the round-1 values of such characteristics are asked in the second round survey.

Consider the following projection of household consumption (or income) on household characteristics for survey round j

$$y_{(i)j} = \beta_j' x_{(i)j} + \varepsilon_{(i)j} \tag{1}$$

where subscript i is placed inside parentheses to emphasize that we can only estimate Equation (1) with repeated cross-sectional data. In that equation β_j' stands for the "return" to household characteristics in $x_{(i)j}$ whereas $\varepsilon_{(i)j}$, which will be called the "error term" in what follows, stands for unobserved time invariant characteristics as well as time variant ones including a stochastic element. To further operationalize the framework, we make the following two assumptions.

Assumption 1: *The underlying population sampled is the same in survey round 1 and survey round 2.*

Assumption 1 ensures that the distributions of the time-invariant household characteristics in the two survey rounds would be the same. As such, these time-invariant household characteristics can be employed as the connectors of household consumption between the two periods (i.e., $x_{(i)1} \equiv x_{(i)2}$). Coupled with Equation (1), this assumption implies that households in period 2 with identical characteristics to those of households in period 1 would have achieved the same consumption levels in period 1 and vice versa (given the same error term). Assumption 1 will be violated if the underlying population changes due to major events as births, deaths, or migration; these events can be caused by natural disasters or economic crises or simply because the two survey rounds are too far apart. We can thus test this assumption by examining whether the observable time-invariant characteristics of the population of interest change significantly from one survey round to the next. Since we have to work with cross sectional data, we also need to choose a survey round, either the first or the second, as the base year for analysis (i.e., using survey j in equation (1)).

Since only the repeated cross sections are available, we could just estimate Equation (1) separately using data from each survey round j. One key challenge is how to specify a functional form $F(\varepsilon_{i1}, \varepsilon_{i2})$ that best models the connection of the error terms across the survey rounds. Dang and Lanjouw (2013) and Dang et al. (2014) propose to use the bivariate normal function for this purpose (which entails the assumption that the error terms have a bivariate normal distribution), whereas Bourguignon et al. (2019) use a different approach. However, in all cases, the proxy used for the joint distribution of the error terms relies on the estimation of the correlation coefficient between the error terms, ρ_p, through pseudo panel techniques applied to the successive cross-sections.

We propose an alternative and more flexible way that employs copulas to connect these error terms. In particular, we make the following assumption.

Assumption 2: *$\boldsymbol{\varepsilon_{i1}}$ and $\boldsymbol{\varepsilon_{i2}}$ can be linked together using a specific copula function C.*

The copula function offers a convenient way to link the two marginal distributions $\boldsymbol{\varepsilon_{i1}}$ and $\boldsymbol{\varepsilon_{i2}}$. Since these error terms are continuous variables, the identified copula function is unique (Sklar, 1973). This in turns results in unique predicted values for household consumption. Different from Assumption 1, testing for Assumption 2 would require actual panels. We provide a more detailed overview of copulas in Appendix 1, Part A.

Given Assumptions 1 and 2 and the predicted parameters from Equation (1), the synthetic panels are constructed as follows

$$\hat{y}_{ik} = \hat{\beta}_k^{'} x_{ij} + \tilde{\varepsilon}_{ik} \tag{2}$$

where *k*= *1* or 2.

A couple remarks are in order for Equation (2). First, the *x* characteristics are from the survey round *j* (or the base survey), while the predicted $\hat{\beta}_k$ are obtained from survey round *k*, and the error term $\tilde{\varepsilon}_{ik}$ is simulated from the copula function $C(\varepsilon_{(i)1}, \varepsilon_{(i)2}, \theta)$. We also remove the parentheses around subscript *i* in Equation (2) to indicate that we produce the predicted consumption for both survey rounds for each household. In other words, we simulate the consumption of households observed in both periods, rather than using observed values in the initial or final period as with previous methods. If our model assumptions are correct, the simulated marginal distributions of consumption would be identical to the observed cross-sectional distributions.

These predicted consumptions can be used to obtain estimates for various measures of poverty or consumption mobility. We employ several popular copula functions in our analysis, which include the Gaussian copula, the Clayton copula, the Frank copula, the Farlie-Gumbel-Morgenstern (FGM) copula, and the Gumbel copula. Since the Gaussian copula allows for equal degrees of positive and negative dependence, it is most appropriate for modelling symmetric distributions. The Frank and FGM copulas also display symmetric distributions, but the Frank copula has a weaker tail dependence than the Gaussian copula and the FGM can only accommodate moderate dependence between the two marginals. The Clayton copula exhibits strong left tail dependence but weak right tail dependence, while the Gumbel has the opposite properties of weak left tail dependence but strong right tail dependence. Second, since the error term $\tilde{\varepsilon}_{ik}$ is simulated from a copula function, we need to use multiple simulations to obtain the distribution of the estimates of these mobility measures. We use 400 simulations in our analysis.

Finally, to compare the degree of dependence across copulas, we can use the Spearman's rank correlation coefficient (ρ_s). This coefficient can be converted from the Pearson correlation coefficient (ρ_p) estimated from the repeated cross sections using the pseudo-panel techniques in Dang and Lanjouw (2013)[5]. We use the following conversion formula given in Kendall and Gibbons (1990)

$$\rho_s = \frac{6}{\pi(n+1)}\left(sin^{-1}\rho_p\right) + (n-2)sin^{-1}\frac{\rho_p}{2} \quad (3)$$

where *n* represents the sample size.

We validate the performance of the synthetic panels against the actual panels for Vietnam in the empirical analysis. But in the absence of any actual panel for validation purposes, it can be useful to compare the marginal distributions (i.e., the predicted consumption in each year) of the synthetic panels against those of the original cross sections. While passing this test does not

[5] Bourguignon et al. (2019) used a different approach still inspired by pseudo-panel estimates.

guarantee that the joint distribution of the copula-based marginal distributions compares well with that of the actual panels, it is useful as a general check on the model fit.

As a rough way of validation, we then compare estimates using the synthetic panels against those based on the actual panels for poverty transitions, both unconditional and conditional, and the more general quintile transition matrix. We also examine several consumption mobility indexes including various versions of the Fields-Ox index, the share movement index, the growth rate of consumption for different consumption groups, as well as the GIC. We offer more details on these mobility measures in Appendix 1, Part B.

III. Data

To validate our method with real survey data, we analyze household panel survey data from the Vietnam Household Living Standards Survey (VHLSS) in 2006 and 2008. The number of households is roughly 9,189 households for each round of the VHLSSs. The VHLSSs are nationally representative surveys implemented by Vietnam's General Statistical Office with technical assistance from international organizations including the World Bank. This survey has been widely used in academic studies as well as in household welfare assessments undertaken by the government and the research community. As is the common practice with the VHLSSs, we use household consumption as a household welfare measure.

The VHLSSs have a rotating panel design, where around one half of the households in the first round are repeated in the next round for the VHLSSs. This combination of both cross-sectional data and panel data in one survey provides an appropriate setting for us to implement our procedures on the cross-sectional component, and then validate our estimates against the true mobility rates from the panel component. Hereafter we also refer to the cross-sectional and panel components respectively as sample A and sample B (where each survey round consists of both samples). We use population weights to provide estimates that are nationally representative.

Consistent with the literature on pseudo-panel data, we restrict household heads' age range to 25-55 for the first survey round and adjust this appropriately for later survey rounds to ensure stable household formation (e.g., looking at the age cohort 27-57 if the next survey round is two years later). While this age range can be extended to include older people, it may be ill-advised to include those who are younger, at least since most household heads tend to be older than 25 in most developing countries. The time-invariant variables that we use include the household head's age, years of schooling, ethnicity (i.e., whether belonging to ethnic majority groups), and whether the household resides in urban areas.

IV. Estimation Results

IV.1. Testing Model Assumptions

We show in Appendix 2, Table 1 the test results for the distributions of the time-invariant variables for the household heads across the two survey rounds. These variables include gender, completed years of schooling, whether the head belongs to ethnic majority groups, and residence areas (urban or rural). The t-test statistics are very close to 0, suggesting that Assumption 1 holds for the data.

The true tests for the synthetic panels are whether the synthetic panel estimates can track those based on the actual panels, which we will show in the next section. But in contexts where such actual panels are not available, we can examine a partial test for whether the predicted marginal distributions in each survey round track the original marginal distributions (which are samples A for both years). As discussed earlier, this is because we simulate the consumption of households observed in both periods, rather than using observed values in the initial or final period as with previous methods.

We examine five copulas: the Gaussian, Clayton, Frank, FGM, and Gumbel copulas. The Pearson correlation coefficient for the error terms $\varepsilon_{(i)j}$ is estimated at 0.62 using the repeated cross sections, which is very close to the corresponding figure of 0.61 using the actual panels.[6] Converting to the Spearman correlation coefficient, we obtain a value of 0.60. We use this value to fix the dependence parameters at 1.5, 4.6, and 1.8 respectively for the Clayton, Frank, and Gumbel copulas. Since the FGM does not accommodate strong dependence for the marginals, we fix its dependence parameter at 0.5, which results in a Spearman correlation coefficient of 0.34.

As an illustration, we plot in Figure 1 the predicted consumption using the first three copulas versus the actual consumption for 2006 (Panel A) and 2008 (Panel B). The copula-based distributions (dotted lines) appear similar to each other and can generally track the actual marginal (solid lines). However, Figure 1 represents just a single draw, and estimates should be based on multiple simulations from the specified copula function.[7]

We subsequently implement a more rigorous test where we divide each distribution into vintiles and test the hypothesis that the median of each of these vintiles for the predicted marginals is not statistically significantly different from that for the original marginals (i.e., with a p-value greater than 0.05; we do not use the means to avoid the influence of potential outliers). Specifically, for each copula, we implement 40 tests against the actual panels (20 tests for the vintiles multiplied by two years). Again, to save space we show in Table 1 the test results for three copulas, the Gaussian, the Clayton, and the Frank copulas, and the full results for all the five copulas in

[6] The Pearson correlation coefficient for household consumption in two years is larger at 0.78.

[7] We implement two non-parametric tests, the Kolmogorov-Smirnov and Wilcoxon rank-sum tests, to compare the copula-based distributions against the actual panels. However, estimation results (obtained using multiple simulations, available upon request) show that these tests do not pick up any differences between the copula-based distributions.

Appendix 2, Table 2.2. While the copula-based distributions, with the base year in 2008, perform reasonably well with replicating the original consumption distribution in this year (i.e., passing between 15 and 18 tests of 20 tests), they can pass at most five tests for 2006. Overall, the summary of the test results at the bottom of Table 1 suggests that the Gaussian copula performs best, passing 23 out of 40 tests. The Frank copula comes in second with passing 22 tests and the Clayton copula comes last with passing 19 tests.

Table 2.2 in Appendix 2 shows that both the FGM and Gumbel copulas perform worse than the Gaussian with passing 20 and 19 tests respectively. However, looking more closely into the full test results in this table, for almost three-fourths (29) of the tests, the Gaussian copula shows larger p-values than the Frank copula. These findings suggest that the Gaussian copula is most appropriate for constructing the synthetic panels. Consequently, we offer subsequent estimation results using the Gaussian copula, but we also offer some estimates using the other copulas for comparison.

IV.2 Estimation Results

We start first with showing estimates for joint (or unconditional) poverty transitions against those based on the actual panels—or the "true" estimates—in Table 2 for the two years 2006 and 2008 of the VHLSS. That is, we produce the synthetic panel estimates based on the cross-sectional component (Samples A) and validate them with the actual panel component (Samples B). The underlying estimated parameters are shown in Appendix 2, Table 2.3. The synthetic panel estimates encouragingly fall within the 95 percent confidence interval (CI) of the true estimates in three cases. For example, the chronic poverty rate using the synthetic panel is 8.4 percent, which is not statistically significantly different from the true chronic poverty rate of 9.9 percent with a standard error of 0.8 percent (Table 2, first row). In fact, the synthetic panel estimates even lie within one standard error of the true estimates for half of the cases (i.e., two out of four). The standard errors of the synthetic panel estimates are somewhat smaller than those of the actual panels, since the former are model-based estimates while the latter are design-based estimates (Matloff, 1981; Binder and Roberts, 2009).

For comparison, we also provide in Appendix 2, Table 2.4 similar estimates for the joint poverty transitions (shown in Table 2) but using the other copulas. While the FGM copula does not provide any estimates that fall within the 95 percent CIs of the true estimates, half of the estimates based on the Clayton, Frank, and Gumbel copulas do. Some estimates also fall within one standard error of the true estimates; for example, for the poor-nonpoor and nonpoor-nonpoor categories with the Frank copula. Yet, compared with the other copulas, the Gaussian still performs the best with scoring more cases that fall within the 95 percent CIs of the true estimates.

Table 3 examines the conditional poverty transitions. Since these mobility measures require an additional layer of estimation (for the denominators) compared to the unconditional transitions, they are often less accurate. Still, the synthetic panel estimates are not statistically significantly different from the actual panel estimates for half of the cases.

We go beyond the two by two poverty transitions in Tables 2 and 3 to further examine in Table 4 the more general five by five transition matrix for the consumption quintiles, where the quintile thresholds are defined separately for each year. More than four-fifths (i.e., 23 out of 25) of the synthetic panel estimates for the inner cells are not statistically significantly different from those based on the actual panels. These estimates are marked in bold letters. Furthermore, more than three-fifths (i.e., 18 out of 25) of the former fall within one standard error of the latter, which are marked with a star. We also provide estimates for two other variants of the quintile transitions, where we keep fixed the quintile thresholds either in the second period (Appendix 2, Table 2.5) or the first period (Appendix 2, Table 2.6). Estimation results become weaker for these cases, but are not statistically significantly different from those based on the actual panels for more than three-fifths of the inner cell transitions (i.e., 16 cases for Table 2.5 and 18 cases for Table 2.6).

Table 5 produces estimates for another index of consumption mobility, which is the median consumption growth rate over time for households in the four groups earlier defined in Table 2. The growth rates of consumption are more difficult to estimate than the poverty transitions in Table 2, because the former require more accuracy with households' exact levels of consumption rather than just scoring whether households fall below a given poverty line as with the latter. Yet, Table 5 provides encouraging estimation results with two of four estimates falling within the 95 percent CI the true estimates. One estimate (i.e., the consumption growth for the chronic poor in the first row of Table 5) even falls inside one standard error of the true estimates.

We generalize the estimation results in Table 5 and plot in Figure 2 the non-anonymous growth incidence curve (GIC) for Vietnam over the period 2006-08. The synthetic panel estimates (solid green line) mostly falling inside the 95 percent CI (gray area) around the true estimates (dotted red line), except for consumption deciles 5 and 8. But overall, the synthetic panel GIC appear to reasonably track the trend of the actual panel GIC.

We turn next to examining some mobility indexes in Table 6. Estimation results are rather encouraging with the Fields-Ok index, the absolute Fields-Ok index in logarithmic form, and the share movement index even falling within one standard error of the true estimates. Although the estimated log Fields-Ok index (0.24) is less accurate than the other estimates, it is reasonably close to the corresponding figure of 0.27 based on the actual panels.

V. Conclusion

We propose to use copula functions to extend existing techniques to construct synthetic panels using repeated cross sections. These copula-based synthetic panels allow us to study various measures of consumption mobility and other mobility indexes that are not available with current techniques. Validation results using both actual panels and repeated cross sections from Vietnam suggest that synthetic panels may not provide the perfect substitute for actual panels. But in contexts where actual panel data are not available, synthetic panels can offer a promising alternative.

References

Beegle, Kathleen, Luc Christiaensen, Andrew Dabalen, and Isis Gaddis. (2016). *Poverty in a Rising Africa*. Washington, DC: The World Bank.

Binder, David A. and Georgia Roberts. (2009). "Design- and Model-Based Inference for Model Parameters". In D. Pfeffermann and C.R. Rao. *Handbook of Statistics, Vol. 29B- Sample Surveys: Inference and Analysis.* North-Holland: Elsevier.

Bonhomme, Stéphane, and Jean-Marc Robin. (2009). "Assessing the equalizing force of mobility using short panels: France, 1990–2000." *Review of Economic Studies,* 76(1): 63-92.

Bourguignon, Francois, Chor-Ching Goh, and Dae Il Kim. (2004). "Estimating Individual Vulnerability to Poverty with Pseudo-Panel Data", *World Bank Policy Research Working Paper No. 3375.* Washington DC: The World Bank.

Bourguignon, Francois, Hector Moreno, and Hai-Anh Dang. (2019). "On the Construction of Synthetic Panels". *Working paper*.

Cancho, Author César, María E. Dávalos, Giorgia Demarchi, Moritz Meyer, and Carolina Sánchez Páramo. (2015). "Economic Mobility in Europe and Central Asia: Exploring Patterns and Uncovering Puzzles". *World Bank Policy Research Paper No. 7173*.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. (2014). "Is the United States still a land of opportunity? Recent trends in intergenerational mobility." *American Economic Review,* 104(5): 141-47.

Cruces, Guillermo, Peter Lanjouw, Leonardo Lucchetti, Elizaveta Perova, Renos Vakis, and Mariana Viollaz. (2015). "Estimating Poverty Transitions Repeated Cross-Sections: A Three-country Validation Exercise". *Journal of Economic Inequality*, 13:161–179.

Dang, Hai-Anh. (forthcoming). "To Impute or Not to Impute and How? A Review of Alternative Poverty Estimation Methods in the Context of Unavailable Consumption Data". *Development Policy Review.*

Dang, Hai-Anh and Peter Lanjouw. (2013). "Measuring Poverty Dynamics with Synthetic Panels Based on Cross-Sections". *World Bank Policy Research Working Paper No. 6504*, World Bank, Washington, DC.

Dang, Hai-Anh, Dean Jolliffe, and Calogero Carletto. (2019). "Data Gaps, Data Incomparability, and Data Imputation: A Review of Poverty Measurement Methods for Data-Scarce Environments". *Journal of Economic Surveys*, 33(3): 757-797.

Dang, Hai-Anh, Peter Lanjouw, Jill Luoto, and David McKenzie. (2014). "Using Repeated Cross-Sections to Explore Movements in and out of Poverty". *Journal of Development Economics*, 107: 112-128.

Fan, Yanqin, and Andrew J. Patton. (2014). "Copulas in econometrics." *Annual Review of Economics*, 6(1): 179-200.

Ferreira, Francisco H. G., Julian Messina, Jamele Rigolini, Luis-Felipe López-Calva, Luis Felipe López-Calva, and Renos Vakis. (2012). *Economic Mobility and the Rise of the Latin American Middle Class.* Washington DC: World Bank.

Foster, James E. and Jonathan Rothbaum. (2015). "Using Synthetic Panels to Estimate Intergenerational Mobility". *Working paper No. 013/2015*. Espinosa Yglesias Research Centre.

Garces-Urzainqui, David. (2017). Poverty Transitions Without Panel Data? An Appraisal of Synthetic Panel Methods. *Paper presented at the 7th Meeting of the Society for the Study of Economic Inequality*, New York city.

Herault, Nicolas and Stephen Jenkins. (2019). "How Valid are Synthetic Panel Estimates of Poverty Dynamics?" *Journal of Economic Inequality*, 17(1): 51-76.

Jantti, Markus and Stephen P. Jenkins. (2015). "Income Mobility", In A. B. Atkinson and F. Bourguignon, ed., *Handbook of Income Distribution*, 807-935, North Holland, Oxford.

Jha, S., A. Martinez, P. Quising, Z. Ardaniel, and L. Wang. (2018). "Natural Disasters, Public Spending, and Creative Destruction: A Case Study of the Philippines". *ADBI Working Paper 817*. Tokyo: Asian Development Bank Institute.

Kendall, M.G., & Gibbons, J.D. (1990). *Rank Correlation Methods* (5th ed.). New York: Oxford University Press.

Matloff, Norman S. (1981). "Use of Regression Functions for Improved Estimation of Means". *Biometrika*, 68(3): 685-689.

Nelsen, Roger B. (2006). *An Introduction to Copulas*. 2nd Edition. New York: Springer.

OECD. (2015). *OECD Employment Outlook 2015*. OECD Publishing, Paris. http://dx.doi.org/10.1787/empl_outlook-2015-en

---. (2018). *A Broken Social Elevator? How to Promote Social Mobility*. Paris: OECD Publishing. https://doi.org/10.1787/9789264301085-en.

Serajuddin, Umar, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew Dabalen. (2015). "Data deprivation: another deprivation to end." *World Bank Policy Research Paper no. 7252*, World Bank, Washington, DC.

Trivedi, Pravin K., and David M. Zimmer. (2007). "Copula modeling: an introduction for practitioners." *Foundations and Trends® in Econometrics*, 1(1): 1-111.

United Nations Development Programme (UNDP). (2016). *Multidimensional Progress: Well-being beyond Income*. New York: United Nations Development Programme.

Vakis, Renos, James Rigolini, and Leonardo Lucchetti. (2015). *Left Behind: Chronic Poverty in Latin America and the Caribbean*. Washington, DC: World Bank.

Table 1. Testing Copulas against the Actual Panels, Vietnam 2006-2008

Vintile	2006			2008		
	Gaussian	**Clayton**	**Frank**	**Gaussian**	**Clayton**	**Frank**
1	N	N	N	N	N	N
2	N	S	N	N	N	N
3	S	S	S	N	N	N
4	S	S	S	N	N	N
5	S	S	S	N	N	N
6	S	S	S	N	N	N
7	S	S	S	N	N	N
8	S	S	S	N	N	N
9	S	S	S	N	S	N
10	S	S	S	N	S	S
11	S	S	S	N	S	S
12	S	S	S	N	S	S
13	S	S	S	N	S	N
14	S	S	S	N	N	N
15	S	S	S	N	N	N
16	S	S	S	N	N	N
17	S	S	S	N	N	N
18	N	N	N	S	N	N
19	N	N	N	S	N	N
20	N	N	N	N	N	N
Summary						
Not significant	5	4	5	18	15	17
Significant	15	16	15	2	5	3
Total	20	20	20	20	20	20

Note: Each cell indicates whether the test of equality of median for the copula-based predicted consumption versus that of the actual panel for each year is statistically significant at the 5% level. "N" and "S" respectively stand for “not statistically significant” and “statistically significant” from the actual panels. This test is implemented for each vintile for each year (panel half). The second survey round is the base year. We use 400 simulations for each copula model.

Table 2. Unconditional Poverty Transitions Based on Synthetic Data for Two Periods, Vietnam 2006-2008 (Percentage)

Poverty Status		
First Period & Second Period	**Actual Panel**	**Synthetic Panel**
Poor, Poor	9.9	**8.4**
	(0.8)	(0.5)
Poor, Nonpoor	5.9	**6.1***
	(0.5)	(0.4)
Nonpoor, Poor	4.9	6.3
	(0.5)	(0.5)
Nonpoor, Nonpoor	79.3	**79.1***
	(1.0)	(0.7)
N	2723	3701

Note: Synthetic panels are constructed from cross sections for Vietnam using the Gaussian copula. The second survey round is the base year. Standard errors are obtained adjusting for complex survey design. All numbers are weighted using population weights. Poverty rates are in percent. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. Joint probabilities are shown. Estimates based on the synthetic panels that fall within the 95% CI and one standard error of those based on the actual panels are shown respectively in bold and in bold with a star "*". We use 400 simulations to obtain estimates.

Table 3. Conditional Poverty Transitions Based on Synthetic Data for Two Periods, Vietnam 2006-2008 (Percentage)

Poverty Status		
First Period & Second Period	**Actual Panel**	**Synthetic Panel**
Poor--> Poor	62.8	**57.9**
	(2.8)	(2.3)
Poor--> Nonpoor	37.2	**42.1**
	(2.8)	(2.3)
Nonpoor--> Poor	5.9	7.4
	(0.6)	(0.5)
Nonpoor--> Nonpoor	94.1	92.6
	(0.6)	(0.5)
N	2723	3701

Note: Synthetic panels are constructed from cross sections for Vietnam using the Gaussian copula. The second survey round is the base year. Standard errors are obtained adjusting for complex survey design. All numbers are weighted using population weights. Poverty rates are in percent. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. Joint probabilities are shown. Estimates based on the synthetic panels that fall within the 95% CI and one standard error of those based on the actual panels are shown respectively in bold and in bold with a star "*". We use 400 simulations to obtain estimates.

Table 4. Consumption Dynamics for Two Periods, Vietnam 2006-2008 (Percentage)

			2008					
			Poorest	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Richest**	**Total**
Panel A: True Panels	**2006**	**Poorest**	12.7	4.7	1.7	0.6	0.2	20
			(0.8)	(0.4)	(0.3)	(0.2)	(0.1)	(0.9)
		Quintile 2	4.8	7.5	4.6	2.0	0.6	20
			(0.4)	(0.6)	(0.5)	(0.3)	(0.1)	(0.9)
		Quintile 3	1.8	5.2	6.9	4.6	1.5	20
			(0.3)	(0.5)	(0.5)	(0.5)	(0.2)	(0.9)
		Quintile 4	0.6	2.0	5.0	7.8	4.8	20
			(0.2)	(0.3)	(0.5)	(0.6)	(0.5)	(0.9)
		Richest	0.1	0.6	1.8	4.9	12.9	20
			(0.1)	(0.2)	(0.3)	(0.5)	(0.7)	(0.8)
		Total	20	20	20	20	20	100
			(1.0)	(0.9)	(0.9)	(0.9)	(0.9)	
			2008					
			Poorest	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Richest**	**Total**
Panel B: Synthetic Panels	**2006**	**Poorest**	**12.3***	**4.9***	**2.1**	**0.6***	**0.1***	20
			(0.3)	(0.3)	(0.3)	(0.2)	(0.0)	(0.0)
		Quintile 2	**5.0***	**6.8**	**5.1***	2.6	**0.6***	20
			(0.3)	(0.3)	(0.3)	(0.3)	(0.1)	(0.0)
		Quintile 3	**2.0***	**5.1***	**6.0**	**5.0***	**1.8**	20
			(0.2)	(0.3)	(0.3)	(0.3)	(0.2)	(0.0)
		Quintile 4	**0.6***	2.7	**4.9***	**6.9**	**4.9***	20
			(0.1)	(0.3)	(0.3)	(0.3)	(0.3)	(0.0)
		Richest	**0.1***	**0.6***	**1.8***	**4.9***	**12.6***	20
			(0.0)	(0.1)	(0.2)	(0.3)	(0.3)	(0.0)
		Total	20	20	20	20	20	100
			(0.0)	(0.1)	(0.1)	(0.1)	(0.0)	

Note: Synthetic panels are constructed from cross sections for Vietnam using the Gaussian copula. The second survey round is the base year. Standard errors are obtained adjusting for complex survey design. Transition rates are in percent and weighted using population weights. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. Joint probabilities are shown. Estimates based on the synthetic panels that fall within the 95% CI and one standard error of those based on the actual panels are shown respectively in bold and in bold with a star "*". We use 400 simulations to obtain estimates.

Table 5. Median Consumption Growth for Two Periods, Vietnam 2006-2008 (Percentage)

Poverty Status		
First Period & Second Period	**Actual Panel**	**Synthetic Panel**
Poor, Poor	3.6	**3.6***
	(0.3)	(0.3)
Poor, Nonpoor	9.1	**9.7**
	(0.4)	(0.3)
Nonpoor, Poor	-1.5	-2.5
	(0.3)	(0.3)
Nonpoor, Nonpoor	3.1	2.8
	(0.1)	(0.1)
N	2723	3701

Note: Synthetic panels are constructed from cross sections for Vietnam using the Gaussian copula. The second survey round is the base year. Standard errors are obtained adjusting for complex survey design. All numbers are weighted using population weights. Growth rates are in percent. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. Joint probabilities are shown. Estimates based on the synthetic panels that fall within the 95% CI and one standard error of those based on the actual panels are shown respectively in bold and in bold with a star "*". We use 400 simulations to obtain estimates.

Table 6. Mobility Indexes for Two Periods, Vietnam 2006-2008

First Period & Second Period	Actual Panel	Synthetic Panel
Fields-Ok index	1629.9	**1689.53***
	(72.7)	(44.1)
Fields-Ok index (log)	0.27	0.24
	(0.01)	(0.01)
Absolute Fields-Ok index (log)	0.32	**0.33***
	(0.01)	(0.01)
Share movement index	0.03	**0.03***
	(0.00)	(0.00)

Note: Synthetic panels are constructed from cross sections for Vietnam using the Gaussian copula. The second survey round is the base year. Standard errors are obtained adjusting for complex survey design. All numbers are weighted using population weights. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. Joint probabilities are shown. Estimates based on the synthetic panels that fall within the 95% CI and one standard error of those based on the actual panels are shown respectively in bold and in bold with a star "*". We use 400 simulations to obtain estimates.

Figure 1. Density Graphs for Actual Panels vs. Synthetic Panels, Vietnam 2006-2008

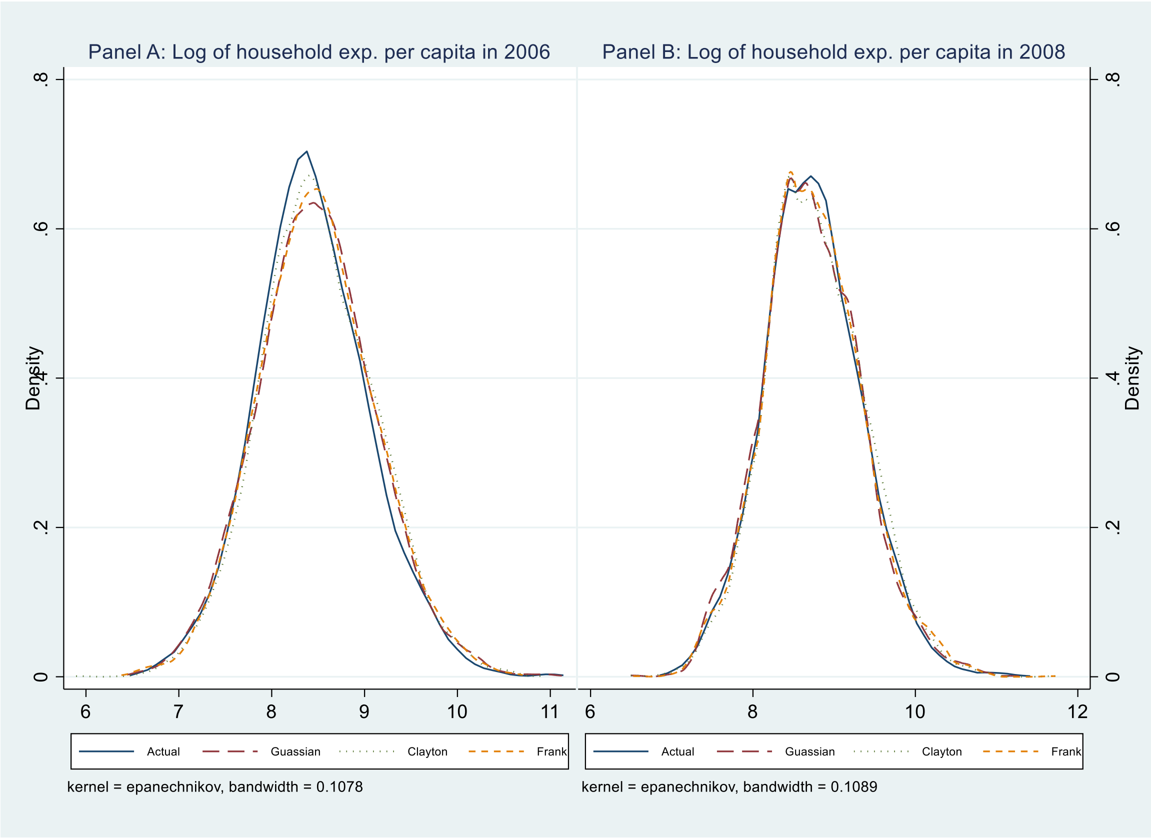
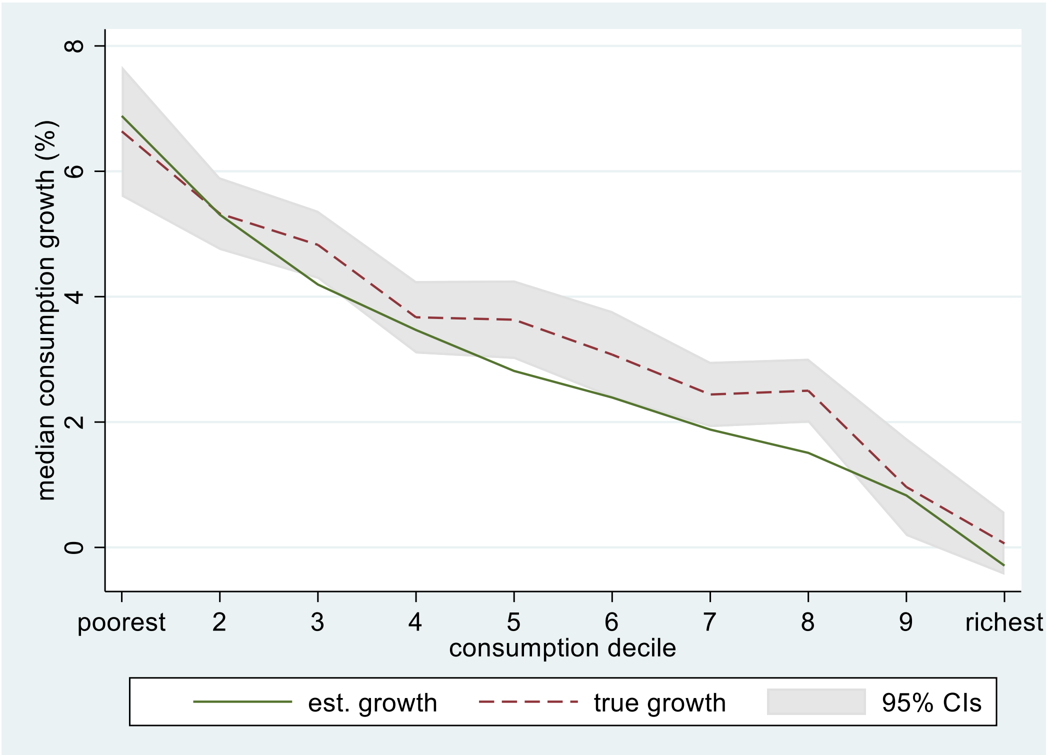


Figure 2. Non-anonymous Growth Incidence Curve, Vietnam 2006-2008



Appendix 1. Overview of Copulas and Mobility Measures

Part A. Overview of Copulas

We provide in this section a brief description of the most relevant features of copulas, their functional forms and properties for the bivariate case, which is mostly based on Trivedi and Zimmer (2005). A comprehensive textbook treatment is provided by Nelsen (2006).

Consider a bivariate continuous distribution function $F(y_1, y_2)$ with univariate marginal distributions $F(y_1)$ and $F(y_2)$ and inverse quantile function F_1^{-1} and F_2^{-1}. Then $y_j = F_j^{-1}(u_j)$, where u_j are uniformly distributed variables, for *j*= 1, 2. The copula $C(u_1, u_2)$ associated with the distribution function $F(y_1, y_2)$ is defined as follows

$$F(y_1, y_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right) = P(U_1 \leq u_1, U_2 \leq u_2) = C(u_1, u_2) \qquad (1.1)$$

The copula $C(u_1, u_2)$ is thus a two-dimensional distribution function with both the marginal distributions having a U(0,1) distribution. If the marginals in Equation (1.1) are continuous, the copula function is unique. Equation (1.1) is often written as follows

$$F(y_1, y_2) = C(F_1(y_1), F_1(y_1); \theta) \qquad (1.2)$$

which emphasizes the role of θ as the dependence parameter that measures the dependence between the two marginals. θ varies for each copula function and as such is often not comparable across different copulas. Some formulae to convert θ to the Spearman correlation coefficient and the Kendall correlation coefficients are provided in the cited texts above.

Table A.1 below lists several commonly used copulas that we examine in this paper. These include the Gaussian copula, the Clayton copula, the Frank copula, the FGM copula, and the Gumbel copula.

Table A.1. Commonly Used Copulas

Copula	Functional form	Domain for θ	Main Properties
Clayton	$\left(u_1{}^{-\theta} + u_2{}^{-\theta} - 1\right)^{-1/\theta}$	(0,∞)	Strong left tail dependence and weak right tail dependence
Frank	$-\frac{1}{\theta} log\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right)$	(-∞,∞)	Symmetric dependence, but with weak tail dependence
FGM	$u_1 u_2(1 + \theta(1 - u_1)(1 - u_2))$	[-1,1]	Symmetric dependence, but with weak magnitude of dependence
Gaussian	$\Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta)$	(-1,1)	Symmetric dependence
Gumbel	$e^{-\left((-log u_1)^{\theta} + (-log u_2)^{\theta}\right)^{1/\theta}}$	[1,∞)	Weak left tail dependence and strong right tail dependence

Note: Φ is the cdf of the standard normal distribution, and $\Phi_G(.)$ is the standard bivariate normal distribution with the dependence parameter θ.

Part B. Overview of Poverty and Consumption Mobility Measures

We provide a brief overview of the different poverty and mobility measures that we analyze in this paper. For a recent review of various income mobility indexes, see, e.g., Jantti and Jenkins (2015).

The unconditional poverty transitions are defined as

$$P(y_1 \sim z_1 \quad and \ y_2 \sim z_2) \tag{1.3}$$

and the conditional poverty transitions are defined as

$$P(y_1 \sim z_1 \quad | y_2 \sim z_2) \tag{1.4}$$

where y_j and z_j are respectively household consumption and the poverty line in period j, j= 1, 2. The relation sign (~) indicates either the larger sign (>) or smaller or equal sign (≤). For example, $P(y_{i2} > z_2 | y_{i1} \leq z_1)$ correspond to the percentage of the poor population in the first period that escape poverty in the second period.

The quintile transitions are defined more generally but in a similar way. The percentage of the population that move from consumption group l in period 1 to consumption group m in period 2 is defined as

$$P^{lm} = P(z_1^{l-1} < y_1 \leq z_1^l \ and \ z_2^{m-1} < y_2 \leq z_2^m) \tag{1.5}$$

where l, m= 1,..., 5, and the z_j are the thresholds that separate the different consumption groups, with $\mathrm{z}_\mathrm{j}^0 = -\infty$ and $\mathrm{z}_\mathrm{j}^5 = \infty$, for period j, j= 1, 2.

The Fields-Ok index is defined as

$$M^F = |y_2 - y_1| \tag{1.6}$$

the log Fields-Ok index is defined as

$$M_{log}^F = lny_2 - lny_1 \tag{1.7}$$

and the absolute log Fields-Ok index is defined as

$$M_{log}^F = |lny_2 - lny_1| \tag{1.8}$$

The share movement index is defined as

$$M^S = \left| \frac{lny_2}{\bar{y}_2} - \frac{lny_1}{\bar{y}_1} \right| \tag{1.9}$$

where $\bar{y}_j$ is the mean consumption in period j.

The non-anonymous growth incidence curve is defined as

$$g(p_1) = \frac{y_2(p_1) - y_1(p_1)}{y_1(p_1)} \tag{1.10}$$

which provides the consumption growth rate between period 1 and 2 of the population initially in position p_1 of the consumption distribution in period 1.

Appendix 2. Additional Tables and Figures

Table 2.1. Summary Statistics for the Cross Sections, Vietnam 2006-2008

	2006	2008	T-test
Log of expenditure per capita	8.45	8.73	
	(0.64)	(0.62)	
Age	42.93	43.76	
	(7.14)	(7.69)	
Female	0.17	0.17	0.0
	(0.38)	(0.38)	
Years of schooling	7.79	7.86	0.07
	(3.69)	(3.71)	
Ethnic majority group	0.85	0.86	0.01
	(0.36)	(0.35)	
Urban	0.26	0.28	0.01
	(0.44)	(0.45)	
N	3596	3701	

Note: *p<0 .1, **p<0.05, ***p<0.01. Standard errors are in parentheses. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. T-tests are obtained adjusting for complex survey design.

Table 2.2. Testing Copulas against the Actual Panels, Vietnam 2006-2008

Vintile	2006					2008				
	Gaussian	**Clayton**	**Frank**	**FGM**	**Gumbel**	**Gaussian**	**Clayton**	**Frank**	**FGM**	**Gumbel**
1	0.4268492	0.5806659	0.3536918	0.3076021	0.3911596	0.9574975	0.9555255	9.56E-01	0.9569172	9.58E-01
2	0.0995374	0.0444912	0.1108612	0.1787674	0.0931174	0.5247212	0.4558605	0.4927884	0.4920803	0.5311618
3	3.20E-07	2.54E-07	4.49E-08	0.00002	0.000014	0.2324239	0.1603651	0.1963073	0.1939605	0.2483295
4	1.44E-06	2.15E-09	2.17E-07	1.95E-06	4.16E-07	0.6985232	0.4498812	0.5762216	0.5370741	0.7643061
5	6.65E-22	3.47E-28	8.88E-23	1.38E-21	1.96E-19	0.8727584	0.6453078	0.8281443	0.7843267	0.8147781
6	1.87E-24	2.32E-25	5.95E-26	7.35E-22	1.20E-22	0.8617957	0.4438909	0.6749104	0.5954131	0.8935142
7	2.16E-25	4.48E-21	2.65E-30	2.69E-29	1.91E-25	0.8210865	0.3262995	0.5757789	0.4775852	0.8879467
8	2.92E-19	9.30E-18	3.56E-24	3.15E-24	2.48E-16	0.5987931	0.1355984	0.3219054	0.2431663	0.8202118
9	7.65E-27	1.10E-20	1.46E-21	9.88E-32	2.03E-18	0.1982027	0.0155856	0.0692422	0.0387992	0.3672082
10	8.49E-34	7.50E-25	1.12E-28	1.04E-27	2.22E-28	0.0725302	0.0018661	0.0153189	0.0055449	0.1809708
11	2.81E-15	5.95E-14	7.52E-16	4.87E-18	2.85E-15	0.1100235	0.0018514	0.0206688	0.0064861	0.3095073
12	8.93E-12	2.68E-09	1.23E-13	8.29E-15	1.55E-11	0.1687596	0.0018403	0.0256304	0.0057981	0.5344458
13	3.80E-08	2.25E-06	6.66E-09	2.58E-09	2.31E-09	0.3567234	4.21E-03	0.0634771	0.0119926	0.8160282
14	1.03E-04	1.08E-04	3.64E-06	1.61E-07	1.56E-04	0.5808095	0.1000503	0.5553254	0.1765796	0.0978749
15	6.66E-03	1.01E-02	7.70E-03	3.75E-03	8.08E-03	0.1391703	3.80E-01	6.33E-01	5.17E-01	0.002552
16	1.01E-02	5.89E-03	8.08E-03	6.11E-03	1.40E-02	0.2138132	0.2022492	0.6583811	0.3053478	0.0028174
17	1.44E-02	7.12E-03	7.05E-03	5.64E-03	1.35E-02	0.1021736	0.3492355	0.6425682	0.4517612	0.0003575
18	1.02E-01	7.04E-02	6.59E-02	9.45E-02	1.16E-01	0.0089021	0.7045552	0.2108545	0.679877	1.54E-06
19	0.3567595	0.3155253	0.306185	0.3720351	0.3744862	0.0041331	0.5148415	0.1326386	0.4713843	5.17E-07
20	0.5792003	0.5676426	0.6066969	0.5377291	0.6184483	0.0799132	7.05E-01	4.40E-01	7.04E-01	7.05E-04
Summary										
Not significant	5	4	5	5	5	18	15	17	15	14
Significant	15	16	15	15	15	2	5	3	5	6
Total	20	20	20	20	20	20	20	20	20	20

Note: Each cell represents the p-value from a test of equality of median for the copula-based predicted consumption versus that of the actual panel for each year. This test is implemented for each vintile for each year (panel half). The significance level is shown at the 5% level. The second survey round is the base year. We use 400 simulations for each copula model.

Table 2.3. Estimated Parameters of Household Consumption Using Cross Sections, Vietnam 2006-2008

	2006	**2008**
Age	0.011***	0.009***
	(0.001)	(0.001)
Female	0.084***	0.113***
	(0.022)	(0.022)
Years of schooling	0.053***	0.056***
	(0.003)	(0.003)
Ethnic majority group	0.361***	0.383***
	(0.026)	(0.026)
Urban	0.433***	0.310***
	(0.024)	(0.023)
Constant	7.166***	7.492***
	(0.051)	(0.050)
σ_v	0.485	0.489
Adjusted R^2	0.407	0.370
N	3596	3701

Note: *p<0 .1, **p<0.05, ***p<0.01. Standard errors are in parentheses. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round.

Table 2.4. Unconditional Poverty Transitions Based on Synthetic Data for Two Periods, Vietnam 2006-2008 (Percentage)

First Period & Second Period	Gaussian	Clayton	Frank	FGM	Gumbel
Poor, Poor	**8.4**	10.3	8.1	6.4	8.0
	(0.5)	(0.3)	(0.3)	(0.3)	(0.3)
Poor, Nonpoor	**6.1***	4.4	**6.3***	7.8	**6.5**
	(0.4)	(0.3)	(0.4)	(0.4)	(0.4)
Nonpoor, Poor	6.3	**4.8***	7.0	8.7	7.2
	(0.5)	(0.3)	(0.3)	(0.3)	(0.3)
Nonpoor, Nonpoor	**79.1***	**80.5**	**78.6***	77.1	**78.4***
	(0.7)	(0.3)	(0.4)	(0.4)	(0.4)
N	3701	3701	3701	3701	3701

Note: Synthetic panels are constructed from cross sections for Vietnam using the Gaussian, Clayton, Frank, FGM, and Gumbel copulas. Predictions are obtained using the estimated parameters from the first and second survey rounds on data in the second survey round. Standard errors are obtained adjusting for complex survey design. All numbers are weighted using population weights. Poverty rates are in percent. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. Joint probabilities are shown. Estimates based on the synthetic panels that fall within the 95% CI and one standard error of those based on the actual panels are shown respectively in bold and in bold with a star "*". We use 400 simulations to obtain estimates.

Table 2.5. Consumption Dynamics for Two Periods, with Quintile Thresholds Fixed in 2nd Year, Vietnam 2006-2008 (Percentage)

			2008					
			Poorest	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Richest**	**Total**
Panel A: True Panels	**2006**	**Poorest**	17.4	12.1	6.0	2.6	0.8	38.8
			(0.9)	(0.7)	(0.5)	(0.3)	(0.2)	(1.1)
		Quintile 2	1.9	5.4	7.2	4.7	1.5	20.6
			(0.3)	(0.5)	(0.6)	(0.5)	(0.2)	(0.9)
		Quintile 3	0.5	1.8	4.4	6.4	3.5	16.5
			(0.1)	(0.3)	(0.4)	(0.5)	(0.4)	(0.8)
		Quintile 4	0.2	0.7	2.0	5.0	5.7	13.6
			(0.1)	(0.2)	(0.3)	(0.5)	(0.5)	(0.8)
		Richest	0.0	0.2	0.4	1.4	8.6	10.5
			(0.0)	(0.1)	(0.1)	(0.3)	(0.6)	(0.7)
		Total	20.0	20.0	20.0	20.0	20.0	100
			(1.0)	(0.9)	(0.9)	(0.9)	(0.9)	
			2008					
			Poorest	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Richest**	**Total**
Panel B: Synthetic Panels	**2006**	**Poorest**	**16.7***	10.6	**6.3***	**2.6***	**0.5**	**36.7**
			(0.3)	(0.5)	(0.5)	(0.3)	(0.1)	(1.0)
		Quintile 2	**2.3**	**5.2***	5.6	**4.2**	**1.4***	18.7
			(0.3)	(0.5)	(0.4)	(0.4)	(0.2)	(1.0)
		Quintile 3	0.8	3.0	**4.8***	**5.6**	**3.2***	**17.3***
			(0.2)	(0.3)	(0.4)	(0.4)	(0.3)	(0.9)
		Quintile 4	**0.2***	1.1	2.7	**5.3***	**6.1***	15.5
			(0.1)	(0.2)	(0.3)	(0.4)	(0.5)	(0.9)
		Richest	0.0	**0.1***	0.6	2.2	**8.8***	11.8
			(0.0)	(0.1)	(0.2)	(0.3)	(0.5)	(0.6)
		Total	20.0	20.0	20.0	20.0	20.0	100
			(0.0)	(0.1)	(0.1)	(0.1)	(0.0)	

Note: Synthetic panels are constructed from cross sections for Vietnam. The second survey round is the base year. Standard errors are obtained adjusting for complex survey design. All numbers are weighted using population weights. Growth rates are in percent. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. Joint probabilities are shown. Estimates based on the synthetic panels that fall within the 95% CI and one standard error of those based on the actual panels are shown respectively in bold and in bold with a star "*". Quintile thresholds are fixed in the 2nd survey round. We use 400 simulations to obtain estimates.

Table 2.6. Consumption Dynamics for Two Periods, with Quintile Thresholds Fixed in 1st Year, Vietnam 2006-2008 (Percentage)

			2008					
			Poorest	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Richest**	**Total**
Panel A: True Panels	**2006**	**Poorest**	7.6	5.3	4.4	1.9	0.5	19.7
			(0.7)	(0.5)	(0.4)	(0.3)	(0.2)	(0.9)
		Quintile 2	1.1	3.9	7.3	5.4	1.9	19.6
			(0.2)	(0.4)	(0.6)	(0.5)	(0.3)	(0.9)
		Quintile 3	0.2	1.6	5.2	8.5	4.5	20.0
			(0.1)	(0.3)	(0.5)	(0.6)	(0.4)	(0.9)
		Quintile 4	0.0	0.8	1.8	7.2	10.4	20.2
			(0.0)	(0.2)	(0.3)	(0.6)	(0.6)	(0.9)
		Richest	0.0	0.1	0.6	2.5	17.2	20.5
			(0.0)	(0.1)	(0.2)	(0.3)	(0.8)	(0.8)
		Total	9.1	11.7	19.2	25.5	34.5	100
			(0.7)	(0.7)	(0.9)	(1.0)	(1.0)	

			2008					
			Poorest	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Richest**	**Total**
Panel B: Synthetic Panels	**2006**	**Poorest**	**7.8***	**5.8**	**4.4***	**1.7***	**0.3**	20.0
			(0.4)	(0.5)	(0.5)	(0.3)	(0.1)	(0.0)
		Quintile 2	1.9	**4.6**	**6.9***	**4.8**	**1.7***	20.0
			(0.3)	(0.5)	(0.6)	(0.4)	(0.3)	(0.0)
		Quintile 3	0.6	2.4	**5.9**	6.8	**4.4***	20.0
			(0.2)	(0.4)	(0.5)	(0.5)	(0.4)	(0.0)
		Quintile 4	0.1	**0.9***	3.4	**6.6**	9.0	20.0
			(0.1)	(0.2)	(0.4)	(0.5)	(0.5)	(0.0)
		Richest	**0.0***	**0.1***	**0.8**	**3.0**	**16.1**	20.0
			(0.0)	(0.1)	(0.2)	(0.3)	(0.3)	(0.0)
		Total	**10.4**	13.8	21.3	22.9	31.6	100
			(0.5)	(1.3)	(1.3)	(1.1)	(0.8)	

Note: Synthetic panels are constructed from cross sections for Vietnam. The second survey round is the base year. Standard errors are obtained adjusting for complex survey design. All numbers are weighted using population weights. Growth rates are in percent. Household heads' ages are restricted to between 25 and 55 for the first survey round and adjusted accordingly with the year difference for the second survey round. Joint probabilities are shown. Estimates based on the synthetic panels that fall within the 95% CI and one standard error of those based on the actual panels are shown respectively in bold and in bold with a star "*". Quintile thresholds are fixed in the 1st survey round. We use 400 simulations to obtain estimates.