Centro de Estudios®
Espinosa Yglesias
PROMOVEMOS LA IGUALDAD
DE OPORTUNIDADES

Using Synthetic Panels to Estimate Intergenerational Mobility

Autores:

James E. Foster
George Washington University

Jonathan Rothbaum
George Washington University

Documento de trabajo no.

13/2015

Centro auspiciado por: ESRU Fundación Espinosa Rugarcía

Using Synthetic Panels to Estimate Intergenerational Mobility[♣]

James E. Foster[♠]
Jonathan Rothbaum[♣]

September 2015

Abstract

This paper studies the use of synthetic panels to estimate intergenerational income mobility in the absence of panel data. Using mobility curves, we show that while existing techniques for generating synthetic panels under strong parametric assumptions provide accurate estimates of transitions into and out of poverty, estimates of mobility a broader range of mobility measures are biased. We introduce a semiparametric technique using copulas to estimate synthetic panels. We validate our approach using data from the United States and show that our approach results in improved estimates of intra- and intergenerational mobility. We then apply the technique to estimate intergenerational income mobility in Mexico.

Keywords: mobility curves, synthetic panels, intergenerational mobility, Mexico, United States.

♣ Forthcoming in Vélez-Grajales, Huerta-Wong and Campos-Vázquez (eds.) (2015). México, ¿el motor inmóvil? Mexico: The Espinosa Yglesias Research Centre.
♠ George Washington University, Department of Economics and Institute for International Economic Policy, Email: fosterje@gwu.edu.
♣ George Washington University, Institute for International Economic Policy, Email: jrothbau@gwmail.gwu.edu.

Research support from The Espinosa Yglesias Research Centre (CEEY) is gratefully acknowledged.

1. Introduction

Inequality and inequality of opportunity are extremely important policy issues in many countries around the world. There has been a large amount of academic research documenting the high and widening levels of inequality in many developed and developing countries. Policies that target education and health outcomes are often introduced with the goal of decreasing inequality of opportunity and leveling the playing field between children from poor and wealthy households. However, due to the exceptionally high data requirements, we know very little about the level of equality of opportunity and mobility in most countries or how these may be changing over time. This is especially true for mobility of non-income measures of well-being such as consumption and wealth.

There have been attempts to estimate mobility without panel data on parent and child incomes. However, these papers generally estimate mobility of a proxy for income, such as education level (Hertz et al. 2007; Nimubona and Vencatachellum 2007; Emran and Shilpi 2012), occupational class (Emran and Shilpi 2011), indices of status (Torche 2005), or some combination of these variables (Behrman et al. 2001; Emran and Sun 2012) which only partially capture the relationship between parent and child incomes. Other works estimate the intergenerational elasticity of income in developing countries, for example by using two stage two sample instrumental variables (Guimaraes Ferreira and Veloso 2006; Dunn 2007; Nunez and Miranda 2010) or by looking at cases where parents and adult children reside together (Hertz 2001; Quheng, Gustafsson, and Shi 2012; Hnatkovska, Lahiri, and Paul 2013).

A similar problem exists in the literature on intragenerational mobility. In that context, the lack of panels following individuals or households over shorter time periods has limited researchers' ability to study transitions into and out of poverty and transitory and chronic poverty. To overcome the data limitations, work by Deaton (1985), Banks, Blundell, and Brugiavini (2001), and Antman and Mckenzie (Antman and McKenzie 2007) have used data from multiple cross sections and cohort level changes in income to study poverty and mobility.

Building on work by Elbers, Lanjouw, and Lanjouw (2003), Dang, Lanjouw, Luoto, and McKenzie (2011) (hereafter DLLM) proposed an approach to study poverty transitions by constructing synthetic panels using multiple cross sections. In their method, household income is estimated using time-invariant characteristics in two separate cross sections of data. For any household observed in the initial cross section but not in the final cross section, the probability of transitioning out of poverty depends on its estimated income in the final period as well as the correlation between the residual of the estimated income in the initial and final periods. By making conservative assumptions about the relationship between the initial and final income residuals, they establish upper and lower bounds on the proportion of the population that escapes or falls into poverty across any possible poverty line. In subsequent papers (Cruces et al. 2011; Dang and Lanjouw 2013), they make stronger assumptions about the relationship between initial and final income so that estimating mobility is reduced to estimating a single parameter. In a validation exercise using panel data, they show how this allows them to get more precise and generally accurate estimates of movements into and out of poverty.

The mobility curve (Foster and Rothbaum 2014) is a useful tool to validate the results of synthetic panel estimates of mobility as it allows the researcher to compare the estimate of true mobility (or poverty transitions) at all possible poverty lines simultaneously.

In this paper, we extend this research on synthetic panels in two ways. First, we show that a semiparametric copula specification using copulas generally provides a more accurate estimate of intragenerational and intergenerational mobility than the parametric specification in DLLM. We validate the semiparametric copula mobility estimates using a variety of data sets from the United States. We test parametric and semiparametric estimates for intragenerational mobility using the Panel Study of Income Dynamics Cross National Equivalence File (PSID-CNEF) and the public use file of the Current Population Survey Annual Social and Economic Supplement (CPS ASEC). Second, we apply the synthetic panel technique to the estimation of intergenerational mobility. Using the U.S. National Longitudinal Surveys of Youth (NLSY), we show that a more flexible copula model provides a more accurate estimate of the known intergenerational mobility than the parametric specification proposed by DLLM. After validating the use of synthetic panels to estimate intergenerational mobility, we apply the technique to measuring intergenerational mobility in Mexico, where panel data containing income of matched parents and their children does not exist.

To estimate Mexican intergenerational mobility, we use two sources of cross sectional data. The first, the ESRU Survey on Social Mobility in Mexico (EMOVI),[1] surveyed cross sections of Mexican households in 2006 and 2011 about their current demographic, household, and economic characteristics in addition to asking a series of retrospective questions about the circumstances of the household heads as children in their parent's households. We match a subgroup of the EMOVI households to households from their parents' cohort in the Mexican national statistical office's (INEGI) survey of household income and expenditures (ENIGH).[2] We construct synthetic panels from the EMOVI and ENIGH cross sections to estimate various measures of intergenerational mobility for Mexican children born between 1966 and 1981.

2. Synthetic Panels

The synthetic panel technique introduced by DLLM has a number of advantages over previous techniques. This method depends on fewer assumptions than other approaches and allows for comparisons within groups as well as between groups. This is important in the context of intergenerational mobility and equality of opportunity as we may interested in differences in mobility by quantile or race, among other decompositions.

We will begin with a brief discussion of the DLLM technique. Suppose there are two surveys of N_1 and N_2 individuals respectively that are random cross sections of the populations of interest. This could be the same population of households for intragenerational mobility or parent and child households for intergenerational mobility.

[1] Encuesta ESRU de Movilidad Social en México. ESRU is an acronym for the Fundación Espinosa Rugarcía, the foundation that funded the survey.

[2] INEGI: Instituto Nacional de Estadística y Geografía. ENIGH: Encuesta Nacional de Ingresos y Gastos de los Hogares.

Let x_{it} be a vector of household i's characteristics in period t. These characteristics can be time invariant (race, ethnicity, sex, place of birth, etc.), deterministic (age) or time-varying characteristics that can be recalled with accuracy (occupation type, employment status, household characteristics such as automobile ownership, size and characteristics of dwellings, etc.). Let y_{it} be a measure of economic status, such as income, wealth, or consumption, which we will call income for ease of exposition. Income can be expressed as a linear projection of the characteristics x_{it} as:

$$y_{it} = \beta_t' x_{it} + \epsilon_{it} \quad (2.1)$$

Given a poverty line or income cutoff c_t, the proportion of household that are upwardly mobile can be defined as:

$$P(y_{i1} \leq c_1 \text{ } and \text{ } y_{i2} > c_2). \quad (2.2)$$

This is the proportion of the population below the poverty line c_1 in period 1 but above the poverty line c_2 in period 2. For downward mobility, $P(y_{i1} > c_1 \text{ } and \text{ } y_{i2} \leq c_2)$ represents the transition from non-poor to poor status. Unfortunately without panel data, we do not observe y_{i1} and y_{i2} for the same households so we cannot calculate (2.2).

DLLM propose rewriting equation (2.2) by substituting (2.1) in for y_{it} so that:

$$P(\epsilon_{i1} \leq c_1 - \beta_1' x_{i1} \text{ } and \text{ } \epsilon_{i2} > c_2 - \beta_2' x_{i2}). \quad (2.3)$$

From this equation, upward mobility depends only on the joint distribution of the error terms, as all of the other terms $(c_t, \beta_t, x_{it}, \epsilon_{it})$ are known or can be estimated.

From (2.3), they estimate bounds on mobility by making assumptions about the relationship between ϵ_{i1} and ϵ_{i2}. They estimate an upper bound on mobility by assuming that the two error terms are completely independent of each other and using a bootstrap to estimate an average level of mobility. Their lower bound estimate of mobility is based on perfect correlation of the error terms which they implement by assuming $\epsilon_{i1} = \epsilon_{i2}$ for each household. They confirm that with panel data from Indonesia and Vietnam, the true estimate is nearly always contained by the bounds, even when the sample is decomposed into regional subgroups. However, the bounds themselves can be very wide. For example, at the poverty line in Indonesia between 1997 and 2000, the true level of upward mobility in the panel data was 0.08, but the lower bound was 0.03 and the upper bound was 0.12. Fields and Viollaz (2013) use data from Chile to test a variety concepts of mobility, including time dependence (correlation), positional movement, share movement, income movement, and mobility and inequality. They also find that the upper and lower bounds proposed are too wide and provide “limited information about poverty transition rates.”

To obtain more narrow estimates on the bounds of mobility or a point estimate requires further assumptions about the relationship between the ϵ_{it} residuals. One possible assumption, made by DLLM, is that ϵ_{i1} and ϵ_{i2} follow a bivariate normal distribution with correlation coefficient ρ and standard deviations σ_{ϵ_1} and σ_{ϵ_2}. One can obtain narrower bounds by assuming that there is a maximum and minimum possible correlation of the errors such that $0 < p_L < p_H < 1$. By narrowing the range of possible correlations to $\rho_L = 0.3$ and $\rho_H = 0.7$, the bounds on upward mobility across the poverty line for Indonesia are narrowed to $[0.08, 0.12]$ in their paper.

Cruces et al. (2011) estimate mobility under the bivariate normality assumption using panel data from Chile, Nicaragua, and Peru. They find that "the methodology performs well in predicting true mobility in and out of poverty by means of two rounds of cross-sectional data; true mobility lies within the two bounds most of the time." They also note that improving the model (including more characteristics in x_{it} to provide a better estimate of y_{it} and therefore smaller residuals) results in more accurate point estimates of mobility from synthetic panels and narrower bounds for a given set of parameters ρ_L and ρ_H.

If the true value for ρ were known or could be estimated, one could obtain a point estimate for mobility rather than bounds. Dang and Lanjouw (2013) propose a method using age cohorts to estimate the correlation between period 1 and period 2 income $\rho_{y_1 y_2}$. The income correlation is equal to:

$$\rho_{y_1 y_2} = \frac{\mathrm{cov}(y_{i1}, y_{i2})}{\sqrt{\mathrm{var}(y_{i1})\mathrm{var}(y_{i2})}} = \sqrt{\frac{\mathrm{var}(y_{i1})}{\mathrm{var}(y_{i2})}}\delta \quad (2.4)$$

Using the mean income for each cohort c in period t, $\bar{y}_{ct}$, δ can be estimate from the cohort regression:

$$\bar{y}_{c2} = \delta' \bar{y}_{c1} + \bar{v}_{c2} \quad (2.5)$$

For this estimate to be unbiased, the cohort variable must satisfy the conditions of an instrumental variable (including exogeneity and relevance).

With the estimated income correlations, the correlation between regression residuals can also be estimated by rearranging the terms in the correlation equation using (2.1) as:

$$\rho_{y_1 y_2} = \frac{\mathrm{cov}(y_{i1}, y_{i2})}{\sqrt{\mathrm{var}(y_{i1})\mathrm{var}(y_{i2})}} = \frac{\mathrm{cov}(\beta_1' x_{i1} + \epsilon_{i1}, \beta_2' x_{i2} + \epsilon_{i2})}{\sqrt{\mathrm{var}(y_{i1})\mathrm{var}(y_{i2})}} \quad (2.6)$$

If the two samples come from the same population (with the same distribution of x variables, an identifying assumption in DLLM), then (2.6) can be rewritten as:

$$\rho_{y_1 y_2} = \frac{\beta_1' var(x_i)\beta_2 + \rho\sqrt{\sigma_{\epsilon_1}^2 \sigma_{\epsilon_2}^2}}{\sqrt{\mathrm{var}(y_{i1})\mathrm{var}(y_{i2})}} \quad (2.7)$$

From (2.7), the correlation of regression residuals ρ can be estimated given an estimate of the correlation of incomes $\rho_{y_1 y_2}$ as all of the other terms can be estimated from the marginal distributions of the variables.

They test their mobility estimates with the estimated ρ's using data from a more expansive set of countries, including Bosnia-Herzegovina, Lao, the United States, Peru, and Vietnam. They find that their results are "quite accurate" and that they "are good not only for the general population but for smaller population groups as well."

3. Mobility Curves

In order to test and validate different synthetic panel techniques on intergenerational mobility, we first use mobility curves (Foster and Rothbaum 2014). Mobility curves plot transitions into and out of poverty at all possible poverty lines. This allows us both to compare our results to DLLM and to analyze how well synthetic panels predict income changes at other points in the distribution and not just at the poverty line or a small set of possible lines.

In this section we provide a brief summary of the mobility curve. The mobility curve is defined to compare how income gains and losses affect welfare. With $y_t = (y_{it}, \dots y_{nt})$ and $y = (y_1, y_2)$, upward and downward mobility at a given cutoff c are :

$$\begin{aligned} m_U(y,c) &= \frac{1}{n}\sum_{i=1}^{n} I(y_{1i} \leq c) I(y_{2i} > c) \\ m_D(y,c) &= \frac{1}{n}\sum_{i=1}^{n} I(y_{1i} > c) I(y_{2i} \leq c) \end{aligned} \quad (3.1)$$

Under the assumptions of utilitarian, time-separable social welfare, the mobility curve is defined so that if income gains resulted in a larger per capita increase in welfare in society B than in society A, B experienced more upward mobility. Let y^A and y^B be initial and final period incomes for societies A and B respectively. From equation (3.1), if $m_U(y^B, c) \geq m_U(y^A, c)$ for all possible cutoffs c and for some cutoffs $m_U(y^B, c) > m_U(y^A, c)$, then B experienced a larger per capita increase in welfare from upward mobility for any monotonically increasing utility function. By definition then, upward mobility is greater in B than A. The mobility curve is constructed by plotting m_U and m_D for all possible values of c. By looking at the mobility curves for B and A, we can easily see if B has more upward mobility than A. Figure 1 shows an example taken from Foster and Rothbaum (2014) with two societies A and B that share identical period 1 incomes, $y_1^A = y_1^B = (1{,}5)$, but where the period 2 incomes differ, $y_2^A = (2{,}4)$, and $y_2^B = (3{,}3)$. It is clear both from looking at the numbers and the mobility curves in Figure 1 that society B experienced greater upward mobility ($1 \rightarrow 3$ compared to $1 \rightarrow 2$) and greater downward mobility ($5 \rightarrow 3$ compared to $5 \rightarrow 4$) than society A.

As we noted above, mobility curves are also a way to look at poverty transitions as they plot transitions out of poverty (upward mobility) and into poverty (downward mobility) across all possible poverty lines. If $c_1 = c_2$, m_U in equation (3.1) is equal to equation (2.2).[3] By allowing us to simultaneously view mobility across all possible poverty lines, mobility curves make it easier to validate the accuracy of synthetic panels in estimating mobility at all possible poverty lines, rather than at one or a subset of arbitrarily chosen lines. As such, mobility curves provide a way to view a "distribution" of mobility that may not be possible with single index measures of mobility or transition matrices.

[3] The poverty lines c_1 and c_2 need not be the same for each household. For example, if household incomes are equivalence adjusted, then the poverty line for each household would be the same in equivalence-adjusted dollars but not in absolute income.

4. Synthetic Panels, Bivariate Normal Errors, and Mobility Curves

Dang and Lanjouw (2013) conducted a series of validations of the use of synthetic panels to estimate poverty transitions. In this section, we will discuss their validation results. As we are extending the use of the technique to intergenerational mobility, we also validate the synthetic panel technique for intergenerational mobility using the 1979 and 1997 National Longitudinal Surveys of Youth (NLSY) from the United States.

Dang and Lanjouw (2013) compare the synthetic panel estimates of movements into and out of poverty[4] in Bosnia-Herzegovina (2001-2004), Laos (2002/03-2007/08), Peru (2005-2006), the United States (2007-2009), and Vietnam (2006-2008). For 17 of the 20 transitions compared (poor to poor, poor to non-poor, non-poor to poor, non-poor to non-poor in each country), the true transition probability in the panel data and the synthetic panel estimates were not statistically significantly different from each other.

However, care must be taken with these results. As an illustration, in Figure 2 we plot the true (sample) mobility curve alongside the synthetic panel estimate of the mobility curve for the United States for 2004-2006 using data from the PSID-CNEF 2005 and 2007 waves. The synthetic panel was estimated for individuals aged 25-55 by first running an OLS regression on income with a small number of independent variables, including age, age squared, years of education, gender, and dummy variables for black and Hispanic.[5] The OLS regression results are in Table 1.

To estimate ρ, we used kendall's τ (ρ_τ) to reduce the effect of extreme outliers. Under the assumption of bivariate normality, the relationship between the ρ_τ and ρ is: $\rho = \sin\left(\frac{\pi}{2}\rho_\tau\right)$ (Demarta and McNeil 2007). For the PSID-CNEF, the actual correlation between the errors is 0.65 and the estimate from ρ_τ is 0.78. If we remove outliers whose predicted income is either 10 times higher or lower than their actual income, the correlation is 0.70.[6] By estimating ρ from ρ_τ, we also do not have to specify which outliers to remove as no single extreme value has a large effect on ρ_τ.

At the poverty line ($10,790 in 2007 dollars), the synthetic panel prediction for upward and downward mobility (corresponding to the poor to non-poor and non-poor to poor transitions in Dang and Lanjouw) is very similar to the true mobility observed in the panel. However, for middle and higher income cutoffs, DLLM mobility exceeds the bootstrapped 95% confidence band of true mobility (at nearly all cutoffs above $50,000 for upward mobility and $70,000 for downward mobility). In addition, at very low income levels (under $4,000 for upward and $10,000 for downward mobility), DLLM underpredicts mobility transitions.

This pattern is not unique to the PSID-CNEF data. Figure 3 shows the true mobility curve with confidence intervals and DLLM estimate for the 1-year panel from the public

[4] In the US, the Panel Study of Income Dynamics (PSID) poverty line was used and in Bosnia-Herzegovina, the 20th percentile of 2001 consumption was used. The authors used the official poverty line for each other country.

[5] These variables are very similar to the ones used in Dang and Lanjouw (2013) with the PSID. The goal is not to replicate their results exactly but to show how synthetic panels may provide a biased estimate of mobility at many points in the mobility curve and also yield their accurate poverty transition results.

[6] As an example for the case of intergenerational mobility, for the NLSY97 removing these outliers increases ρ fom 0.200 to 0.227, and the correlation estimated from ρ_τ is 0.246.

use CPS ASEC from 2005 to 2006.[7] We analyze mobility of total household income adjusted using the square root equivalence scale for all households with heads 25-54 years old in 2005. The OLS regression results for the CPS ASEC are shown in Table 2. The mobility curve results are very similar to the PSID, with the DLLM underestimating upward and downward mobility at very low income cutoffs and overestimating both at many middle and higher income cutoffs. In the CPS ASEC, the DLLM and true mobility curves converge at higher income cutoffs so that the two are not statistically significantly different in either upward or downward mobility at about $150,000 and above.

Using data from the 1979 and 1997 NLSYs, we can also test the synthetic panel technique with intergenerational data. Again we use the square root equivalence scale to adjust household incomes. We measure intergenerational income mobility for children in their early years in the workforce (ages 26-30) compared to their parents income when the children were in their teens (14-18 in the NLSY79 and 12-16 in the NLSY97). In each case, we average the household income over two years to reduce the attenuation bias from measurement error (Mazumder 2005) as much as possible. The NLSY79 OLS regression for parents and children includes age, age squared, years of education, dummies for industrial category of the household head's primary job, a dummy for urban status, region dummies, and dummies for non-Hispanic blacks, Hispanics and other non-whites. The OLS regression results are shown in Table 3.

Figure 4 shows the true mobility curve and the synthetic panel estimate for various correlation values. As the correlation between period 1 and 2 errors is about 0.24, which is used for the synthetic panel estimates shown in the figure.[8] Comparing the DLLM estimate to the true mobility curve, a number of things stand out. For both upward and downward mobility, as with PSID-CNEF and CPS ASEC intragenerational mobility, the DLLM estimate exceeded the 95% confidence band of the true mobility curve at many middle and high income cutoffs. Although not shown in the figure, by assuming bivariate normality the DLLM estimate actually exceeds the nonparametric DLLM upper bound estimate of no correlation between initial and final incomes in both upward and downward mobility at some cutoffs.

Figure 5 shows intergenerational mobility and the DLLM estimate for the NLSY97 sample. The OLS regression results for this group are in Table 4. In this case, fewer variables were used to show how the synthetic panel estimates performed even with a relatively low R^2 in the OLS regression.[9] The NLSY97 comparison between the DLLM estimates and the true mobility curve is very similar to the NLSY79 comparison. Again the DLLM mobility estimate exceeds the 95% confidence band for actual mobility at

[7] Confidence intervals estimated using CPS ASEC replicate weights. For a discussion of standard errors and replicates weights in the CPS ASEC, see "Estimating ASEC Variances with Replicate Weights" (US Census Bureau 2013). The CPS ASEC surveys household based on their address and does not follow individuals who move, but instead surveys the new residents at the address in the later wave. Therefore, a matching process must be done to construct a panel from the survey. For a thorough discussion of the construction of panels from the CPS ASEC, see Madrian and Lefgren (2000). As our goal is only to validate synthetic panel techniques, we use a simple method to create the panel. We include households whose head is of the same race and gender in both years and whose age is between one year less and two years greater in 2006 than in 2005.

[8] The true error correlations are 0.241 for the NLSY79 and 0.246 for the NLSY97.

[9] Cruces et al. (2011) show how increasing the number of explanatory variables improves the prediction and narrows the range between the upper and lower bounds as ϵ_{it} accounts for a smaller share of household income.

many middle and high income cutoffs. In both NLSYs, the DLLM mobility curves estimates for upward and downward mobility exceed the nonparametric upper bound over large ranges of cutoffs, for example including nearly all downward mobility cutoffs above $40,000 in both cases.

5. Copulas and Dependence in Non-normal Error Distributions

We showed in the previous section that even with the known correlation between period 1 and period 2 incomes, assuming bivariate normally distributed errors in the log income regressions can yield biased estimates of mobility. These estimates can even exceed the nonparametric upper limit, which makes no assumptions about the marginal distribution of the error terms. Therefore, a major source of this bias could be from the normality assumption. DLLM test and reject the normality assumption in both the univariate and bivariate error distributions for both countries analyzed. Dang and Lanjouw (2013) plot the log of incomes and consumption against the normal distributions for a variety of countries, and in all cases the errors deviate from normality as well. DLLM acknowledge that "despite this rejection we will maintain the assumption…, and thereby illustrate the performance of our parametric bounding methods in a typical practical situation where the underlying distributional assumption may not hold precisely."[10]

In this paper we propose an alternative technique to utilize the observed distribution of errors with copulas to estimate the joint distribution of income instead of assuming bivariate normality. In this way, we can test how removing the assumption of bivariate normally distributed errors can improve the estimation of mobility measures using synthetic panels. Copulas are functions that relate multivariate distributions to their marginal distributions. They are especially useful when the variables are non-normal (Trivedi and Zimmer 2006). Copulas have been used in a wide variety of applications in economics, especially finance (Cherubini, Luciano, and Vecchiato 2004), but also modeling income and wealth (Jäntti, Sierminska, and Van Kerm 2012), tax incidence and inequity (Bø, Lambert, and Thoresen 2011), and the distribution of income (Zimmer and Kim 2012), inequality (Vinh, Griffiths, and Chotikapanich 2010), and mobility (Bonhomme and Robin 2009).

For any joint distribution F there exists a copula C that relates the joint to the marginal distributions F_1 and F_2 (Jaworski et al. 2010):

$$F(x_1, x_2) = C\big(F_1^{-1}(x_1), F_2^{-1}(x_2)\big). \qquad (5.1)$$

The choice of copula determines the dependence structure between ranks in the period 1 and period 2 marginal distributions.

Many classes of copulas have been used in the literature. In this paper, we use the Gaussian (Normal) copula, which simulates the dependence structure of the multivariate normal distribution. The dependence between ranks in the error distribution

[10] In Dang and Lanjouw (2013), the authors attempt to reduce the deviation from normality by implementing a Box-Cox transformation of period 1 and 2 incomes to minimize the skewness of the transformed income distributions. This reduces but does not eliminate the biases reported in the previous section. Although far less frequently, the synthetic panel generated with the Box-Cox transformation at ρ_{Actual} exceeds the non-parametric upper bound DLLM estimate at some cutoffs as well. In none of the data sets analyzed in this paper does this occur using the Gaussian copula.

will be the same as in the case of bivariate normality, but the marginal distribution F_1 and F_2 can be estimated from the empirical distributions in the data without imposing normality as in DLLM.We have chosen to use the Gaussian copula in this paper for a number of reasons. The first is simplicity. The Gaussian copula is determined by a single parameter, the correlation ρ between period 1 and 2 incomes just as the DLLM results are under bivariate normality. In this way, we can compare our results to theirs under the same correlation parameter to evaluate the different techniques. Another advantage of the Gaussian copula is that if the DLLM bivariate normality assumption holds, the copula and DLLM will yield the same result.

However, the Gaussian copula also has disadvantages relative to alternative copulas. One important shortcoming is that the Gaussian copula assumes no dependence between extreme values in the tails of the distribution (Demarta and McNeil 2007). Other more flexible copulas, such as the *t* copula provide additional parameter(s) which determine tail dependence. However, the goal in this paper is to show how relaxing the normality assumption using copulas can help in the estimation of mobility with synthetic panels, not to endorse the use of a specific copula.[11]

We generate the synthetic panel using the Gaussian copula to estimate mobility measures and then repeat the process as in a bootstrap to find the average mobility curve using the copula approach as follows:

1) Run the OLS regression of $\ln(y_{it})$ on independent variables x_{it} to get $\hat{\epsilon}_{i1}$ and $\hat{\epsilon}_{i2}$ for each individual or household from the two cross-sections.
2) Generate a synthetic panel dependency matrix r with n observations where, $r_t = (r_{1t}, \ldots, r_{nt}), r = (r_1, r_2)$ $r_{jt} \in [0,1]$ from the Gaussian copula with correlation parameter ρ. Each r_{jt} is the quantile in the error distribution for synthetic individual j in period t.
3) For each household, take a random draw from the copula distribution dependency vector r_j, where $r_j = (r_{j1}, r_{j2})$.
4) Use the kernel density to estimate the empirical cdf of the period 1 and period 2 error distributions, $\hat{F}_1$, $\hat{F}_2$. Then for each household i with synthetic panel error draw j, the synthetic panel period 1 and period 2 errors are $\tilde{\epsilon}_{i1} = \hat{F}_1^{-1}(r_{j1})$ and $\tilde{\epsilon}_{i2} = \hat{F}_2^{-1}(r_{j2})$.[12]
5) Repeat for all i individuals to get the complete synthetic panel income values with $\tilde{y}_{i1} = x_{i1}^{2\prime}\beta_1 + \tilde{\epsilon}_{i1}$ and $\tilde{y}_{i2} = x_{i2}'\,\beta_2 + \tilde{\epsilon}_{i2}$, where x_{i1}^2 are the period 1 covariates for each individual i observed in period 2 and construct the mobility curve from the synthetic incomes.
6) Repeat steps 1-5 with B replications and average the mobility curve over all replications to compute the average synthetic panel estimate of the mobility curve (as in DLLM for the upper bound estimate of mobility).

The next step is to validate the copula technique with existing data. Figure 2 and Figure 3 also show the copula estimate of mobility curves with the true value of ρ from

[11] We leave to potential future work the task of selecting the appropriate copula and estimating additional copula parameters beyond ρ.

[12] If the observations are equally weighted, the period 1 errors can be drawn directly from the observed error values. However, if the observations have weights, the empirical cdf must be estimated and the errors drawn from the estimated distribution.

the PSID-CNEF and CPS ASEC 1-year panel respectively. Both techniques provide a reasonably accurate of the experienced mobility, especially near the poverty line (approximately $11,000 in each).

In the online appendix, Appendix 2, we compare the absolute and squared deviations of the synthetic panel estimates from the true mobility observed in the panel, including for the quartile decomposition. These figures (A2.1-4 for the PSID-CNEF and A2.5-8 for the CPS ASEC) plot the deviation at each cutoff and the mean deviation up to each cutoff.[13] While the absolute and squared deviations and their means for the copula estimates are not always less than for the DLLM, the copula is less biased at the majority of cutoffs in both the PSID-CNEF and CPS ASEC. In the quartile decompositions, the copula outperforms the DLLM at nearly all cutoffs as well.

As we are studying intergenerational mobility, we again look at the 1979 and 1997 waves of the NLSY and compare the DLLM and Gaussian copula results. In both cases, we calculate the copula synthetic panel mobility estimate using the true period 1 and 2 error correlation ($\rho_{Actual} = 0.24$). The results for the NLSY79 are shown in Figure 4 and for the NLSY97 in Figure 5. In the deviation figures (A2.9-12 for the NLSY79 and A2.13-16 for the NLSY97), at nearly every cutoff, the copula outperforms the DLLM both at the aggregate level (the mobility curve m and aggregate mobility curve M) and for each quartile in the decomposition as well.

Figure 6 shows the complete mobility curve results including quartile decomposition for the NLSY79. Both the copula and DLLM overestimate upward mobility in the 1st quartile (DLLM at lower income levels in both). In the second quartile, both synthetic panel estimates have the upward and downward mobility curves shifted to the left, but in both cases, the copula is closer to the true value than the DLLM. The copula results are a very good match of the true results in the third quartile, and both the copula and the DLLM overestimate downward mobility in the 4th quartile. The results are nearly identical for the NLSY97 (not shown).

A major reason the DLLM results may be biased is that the predicted initial and final synthetic panel distributions do not match the data, due to the normality assumption. The DLLM mobility curves in the 2nd and 3rd quartiles are shifted to the left as a result. The copula, by utilizing the empirical error distributions, does not suffer as much from this shortcoming.

It is clear from these figures that neither the copula nor the DLLM estimates for mobility are perfect. However, the deviation comparisons in Appendix 2 show that the copula generally offers less biased estimate of true mobility than DLLM using these data sets.

We also calculate the true and synthetic panel estimates for a variety of other measures of mobility, including the intergenerational elasticity of income (IGE), the correlation between log incomes, and quintile transition matrices. These estimates are shown in Appendix 1. For each mobility measure, the results are reported both for ρ_{Actual} and a range of possible correlations $\rho_L < \rho_{Actual} < \rho_H$. In each of the four data

---

[13] The deviation at each cutoff is measures for $k = 1,2$ (absolute, squared) as $d_k(c) = \left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^k + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^k$. The mean deviation up to each cutoff (c) is the average of the deviations at all cutoffs up to and including c as
$\mu_k^d(c) = \frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i) - m_U^{Synthetic}(y,c_i)\right|^k + \left|m_D^{True}(y,c_i) - m_D^{Synthetic}(y,c_i)\right|^k\right)$.

sets we have discussed, the copula estimate of the IGE (Table A1.1) and log income correlation (Table A1.2) at ρ_{Actual} is closer to the true value than DLLM. In the two NLSYs, the copula IGE and correlation estimates are nearly the same as the true values.

The two synthetic panel techniques have nearly identical results in estimating the quintile transition matrices (Table A1.3). Both the DLLM and copula techniques were more accurate in estimating intergenerational transitions in the NLSYs than intragenerational ones in the PSID-CNEF and CPS ASEC. For majority of the cells in the quintile transition matrices of the NLSYs, the true transition probability is within the 95% confidence interval of the DLLM and copula estimates. The fact that the copula offers little to no advantage in estimating transitions supports the notion that the biases in the DLLM synthetic panels are due to distributional assumptions, as these assumptions are likely to be less relevant when analyzing rank changes as opposed to income changes.

The validation results from the Gaussian copula give us confidence in the applicability of this technique in estimating intergenerational mobility with synthetic panels. It also suggests that an avenue for future research is to test the use of copulas with greater dependence at the tails of the error distribution. This may reduce the biases caused by the overestimation of downward mobility at the top of the distribution and upward mobility at the bottom of the distribution.

6. EMOVI and Intergenerational Mobility in Mexico

Now we turn to applying this technique to estimating intergenerational mobility in Mexico, where the necessary panel data does not exist. To construct this estimate we use two distinct sources of data. One source of data must include information about the income (y_{i2}) and characteristics (x_{i2}) of child households. This data source must also include retrospective information about the parents of the child household cohort (x_{i1}^2). The second source of data must include income (y_{i1}) and characteristics (x_{i1}) for a cross section of individuals from the parent cohort. The x_{i1} variables in the parent data set must also match the x_{i1}^2 retrospective information from the child cohort.

For the child cohort and retrospective data, we use the EMOVI survey. This survey was conducted on cross sections of Mexican households in 2006 and 2011 with household heads between the ages of 25 and 64. The 2006 EMOVI includes 4,743 households that report positive income and supply information about their parent characteristics. The 2011 EMOVI includes 3,818 such individuals. The survey includes a number of questions about individual and household characteristics in the current generation, including assets, monthly income, occupation, and education. Crucially for this study, the EMOVI surveys also include a series of questions about the characteristics of their parents and the childhood households of the interviewed individuals. These questions are generally of the form, “When you were 14, …” and include information about parent occupations, education levels, locality, and household characteristics. The retrospective questions focus on characteristics that an individual is likely to recall with reasonable accuracy. For example, the children are not asked about their parent’s monthly income, but instead about whether their parents worked and what their occupation was. The children were also asked retrospective questions about household characteristics that are likely good predictors of income, such as ownership of cars, telephones, and televisions and access to electricity and indoor plumbing. The synthetic panel technique requires x_{it} variables that are only good *predictors* of income. The relationship need not be causal, so these questions about assets and access to services are especially valuable in generating accurate estimates of mobility.

We use the Mexican household income and expenditure survey, the ENIGH, as the source of data on the income and household characteristics of the parent cohort. The ENIGH is a nationally representative household survey in Mexico. The ENIGH includes information on household income, consumption, and various household characteristics. The ENIGH survey was conducted in 1984, 1989 and then every other year from 1992. The ENIGH and the EMOVI surveys are a good match for this analysis as there is a significant amount of overlap between the retrospective questions in the EMOVI surveys and the questions in the ENIGH.

We match a subset of the individuals in the EMOVI surveys with individuals from their parent cohorts in the ENIGH. Because the retrospective questions in the EMOVI survey ask about parent and household characteristics when the respondents were 14, we construct the synthetic panels for a subset of the individuals in the EMOVI samples. Individuals who were 36 in 2006 were 14 in 1984, the first year the ENIGH survey was conducted. We therefore restrict our focus to those individuals in the 2006 EMOVI who were between 30 and 39 (born 1966-1976) and match them to individuals in their parent cohort in the 1984 ENIGH. Individuals who were 36 in the 2011 EMOVI were 14 in 1989. As a result, we use the 1989 ENIGH as the cohort matched to the parents of the 2011 EMOVI individuals where between 30 and 39 (born 1971-1981). The household income used is individual equivalent household income using the square root equivalence scale. In both ENIGH surveys, we include only households with children in the sample to match the parent cohorts from the EMOVI surveys.

To be included in x_{i1}, x_{i2}, and x_{i1}^2, we restrict our attention to variables in all three survey sections, the parent cohort survey (ENIGH), the child household survey (EMOVI), and the survey of child households about their parent household characteristics (EMOVI retrospective questions). This includes education level, occupation, region, ownership of car, telephone, and television, and access to indoor plumbing and electricity. The regressions also include dummies for various city sizes using municipality data from the Mexican census from 1980 for the ENIGH84 and EMOVI06 parent cohort, from 1990 for the ENIGH89 and EMOVI11 parent cohort, and 2005 for the EMOVI06 and EMOVI11 child cohorts. Our sample of 30-39 year-old household heads is 1,388 for the EMOVI06 and 1,042 for the EMOVI11.

Table 5 shows the summary statistics for the regression variables in the ENIGH and EMOVI parent cohorts. If recall were accurate and the ENIGH surveys were a representative sample of parents from the EMOVI cohorts, the parent characteristics should be the same between each ENIGH and EMOVI parent cohort pair. For both pairs (ENIGH84-EMOVI06 and ENIGH89-EMOVI11) there are some variables with relatively large differences in means including, for example, the dummies for secondary education and indoor plumbing. However, in general, the means of the included variables for each parent cohort pair are similar. The regressions results are in Table 6 for the ENIGH84 and EMOVI06 and ENIGH89 and EMOVI11 parent-child cohorts. For both parent cohorts, the $R^2 = 0.50$. For the child cohorts, especially in the EMOVI11, the R^2 is lower.

This may be due in part due to the lower predictive power of some variables on income in the 2000s as compared to the 1980s such as owning a TV, having electricity, etc.[14]

In order to construct the synthetic panels, we must first estimate the OLS error correlation ρ for each period. To do this we follow Dang and Lanjouw (2013) in estimating $\rho_{y_1 y_2}$ using cohort groups. In our case, we construct cohorts of parent characteristics from the ENIGH84 and 89 samples and the parent characteristics of the EMOVI06 and 11 samples. We use age-city size cohorts for the parent generations. We create age cohorts in three year cohort groups from 35 to 52 as well as under 35 parents and over 52 parents in 1984 and 1989 respectively with each age cohort divided by the four city size dummies in Table 5. This yields 32 cohort groups. Using equations (2.4) we calculated $\rho_{y1 y_2}$ on the pooled ENIGH84/89 and EMOVI06/11 sample to increase our cohort group sizes. This yielded an estimated log income correlation of 0.39, which is slightly higher than but similar to the correlation in the NLSY samples ($\rho_{y_1 y_2}^{NLSY79} = 0.36$ and $\rho_{y_1 y_2}^{NLSY97} = 0.31$). We then estimated the OLS residual correlation using equation (2.7) using the x_i's from the EMOVI retrospective questions, which results in $\rho_{Estimated}^{84-06} = 0.12$ and $\rho_{Estimated}^{89-11} = 0.18$. The higher estimated ρ for the ENIGH89-EMOVI11 is due to the lower R^2 in the EMOVI11 child income regression.

Figure 7 shows the estimated intergenerational mobility curve for the 84-06 and 89-11 periods using the estimated correlations as well as upper and lower bound estimates with $\rho_U = 0.05$ and $\rho_L = 0.30$. The upper and lower bounds provide a very narrow range of mobility estimates, with a maximum difference between the two in both periods for any cutoff of 0.024. Due to the narrow bounds on the mobility estimates, we will proceed with our analysis using $\rho_{Estimated}$.[15] At $\rho_{Estimated}$, we find that upward and downward mobility were greater at lower income levels for the 84-06 cohort and greater at higher income levels for the 89-11 cohort. Figure 7 also shows the mobility the quartile decompositions for both the 84-06 and 89-11 periods. The gap between the mobility curves of the two cohorts are primarily due to differences in mobility of children in higher income quantiles.

However, these absolute mobility differences are primarily due to differences in the initial and final income distributions for both the parent and child cohorts in the 1984 and 1989 ENIGH and 2006 and 2011 EMOVI. Figure 8 shows the mobility curve and decomposition for mobility of rank. There are virtually no differences in rank mobility between the two ENIGH-EMOVI cohorts in any of the quartiles at any cutoff. This persistence of rank mobility over short periods of time is consistent with the NLSY data and other work on mobility over time in the US using tax data (Chetty, Hendren, Kline, Saez, et al. 2014).

Because of the stability in the rank mobility curves and in order to compare our results to other estimates of intergenerational mobility in Mexico and elsewhere, we also

[14] Because of the lower child regression R^2, we also tested a variation of the copula procedure detailed on page 11. In this variation, the child errors ϵ_{i2} were not replaced by a random draw from the empirical distribution based on the dependency matrix r. Instead, each individual's position in the error distribution was matched to the corresponding position in the dependency matrix where $r_{j2} = \hat{r}_{i2}$. The residual assigned to that individual's parents was then $\tilde{e}_{i1} = F^{-1}(r_{J1})$, from the corresponding period 1 rank from the dependency matrix. The mobility results were virtually identical for the two methods for all periods in both the Mexican and US NLSY data.

[15] Estimates of the different mobility measures using ρ_L and ρ_H are reported in Appendix 1.

look at quintile and decile transition matrices, shown in Table 7 and Table 8. In the EMOVI survey, individuals were asked to rate the relative socioeconomic status of their current household and their parent's household. Using the full 2011 EMOVI sample (and not just individuals 30-39 years old individuals as in this study), Velez, Campos and Huerta (2013) found that 48% of individuals who grew up in households in the lowest status quintile remained in the lowest status quintile as adults. In addition 52% of children who grew up in household in the highest status quintile remained there as well.

Using income and only EMOVI households headed by 30-39 year-olds, we estimate that in the two cohorts 35-38% of children who grew up in the poorest quintile of households remained there, and 37-39% of children born in households in the highest income quintile remained there. Children born in the poorest quintile of households are approximately five times more likely to end up in the poorest quintile than children born in the richest quintile.

This is generally similar to the results with the US data. 43% of children born to the poorest quintile remained there in the NLSY79, and 40% did so in the NLSY97. In both cases, the copula estimate was 4% below the true value and nearly within the same range as the estimates for Mexico. For children born to the top quintile, in both NLSYs 38% of children born in the top quintile remained there as adults. However, the copula estimates at ρ_{Actual} are 33% and 35%.[16] For children born in the top quintile, the NLSY sample values are nearly identical to our estimates for Mexico, but the NLSY copula estimates are lower.

The decile transition matrix (Table 8) shows just how great the gap in equality of opportunity is between children born in poor households and children from rich households. Averaging across the two Mexican cohorts, we estimate that children born in 1st decile households are about 11 times as likely to be in the first decile as adults than children born in the top decile. In addition, top decile children are nearly 10 times more likely to be in the top decile as adults than bottom decile children.[17] From the US NLSY data, the estimates are very similar. Averaging across the two NLYSs, bottom decile children are 10 times more likely to be bottom decile adults than top decile children (12 times in the copula estimate at ρ_{Actual}). Top decile NLSY children are 14 times more likely to be top decile NLSY adults (9 times in the copula estimate at ρ_{Actual}).

We can also compare rank mobility between Mexico and the United States by looking at the rank-rank slope as in Chetty, Hendren, Kline, and Saez (2014). This is the coefficient from a regression of child on parent ranks in their respective income distributions. Our estimates of the rank-rank slope for Mexico are 0.35 for the 84-06

[16] Using tax data for children roughly the age those in the NLSY97 and, Chetty, Hendren, Kline, and Saez (2014) estimate the probability of bottom quintile children remaining in the bottom quintile at 34% and the probability of top quintile children remaining in the top quintile at 37%. Their samples differ from ours in a couple of ways. They include households with zero income and do not equivalence adjust household income.

[17] In the ENIGH84-EMOVI06, bottom decile children are 15 times more likely to be in the bottom decile as adults than top decile children. From the ENIGH89-EMOVI11, we estimate that they are 8 times more likely. Averaging the probabilities of top and bottom decile children being in the bottom decile across the two cohorts, we estimate that bottom decile children are 11 times more likely to be in the bottom decile. For the probability of adult children being in the top decile, the estimate from averaging across the two cohorts is 10 times (11 times in the 84-06 cohort and 9 times in the 89-11 cohort).

cohort and 0.38 for the 89-11 cohort. This is very similar to the copula estimates of 0.39 for the NLSY79 and 0.35 for the NLSY97. Both NLSY copula estimates slightly underestimate the sample rank-rank slopes, which are 0.42 in the NLSY79 and 0.40 in the NLSY97.[18]

7. Conclusion

In this paper, we used a variety of mobility measures, such mobility curves, IGE, correlation, transition matrices, to show that synthetic panels can be used to provide reasonably accurate estimates of intergenerational income mobility in the absence of panel data. To do so, we introduced the use of copulas to improve the accuracy of synthetic panel mobility estimates over methods that impose strong distributional assumptions that are known to be generally invalid. We validated the use of copula-based synthetic panels on intra- and intergenerational mobility, and we used the copula-based synthetic panel technique to estimate income mobility in the context of Mexico, where no suitable panel data exists.

This allowed us to construct estimates of each of the aforementioned measures for income mobility in Mexico. Our estimates enable us to quantify the gaps in equality of opportunity available to children in Mexico, for example by comparing the prospects of children born in the richest households to those born in the poorest households. We estimate that children from the poorest decile of households are about 11 times more likely to be in the bottom decile as adults than children from the top decile, and top decile children are 10 times more likely to be in the top decile as adults than children in the bottom decile. These estimates suggest that rank mobility is very similar between the United States and Mexico.

[18] The full results rank-rank slope results and copula estimates are in Table A1.4 in Appendix 1. Chetty, Hendren, Kline, and Saez (2014) estimate the rank-rank slope with tax data as 0.341, with the same caveats about their inclusion of households with zero income and no equivalence adjustment to household income. For context, they estimate the rank-rank slope in Denmark to be 0.180, implying a much lower degree of persistence in income rank across generations in Denmark than in either the United States or Mexico.

References

Antman, Francisca, and David J. McKenzie. 2007. “Earnings Mobility and Measurement Error: A Pseudo-Panel Approach.” *Economic Development and Cultural Change* 56 (1) (October): 125–161. doi:10.1086/520561.

Banks, James, Richard Blundell, and Agar Brugiavini. 2001. “Risk Pooling, Precautionary Saving and Consumption Growth.” *The Review of Economic Studies* 68 (4): 757–779.

Behrman, Jere R., Alejandro Gaviria, Miguel Székely, Nancy Birdsall, and Sebastian Galiani. 2001. “Intergenerational Mobility in Latin America.” *Economia* 2 (1): 1–44.

Bø, Erlend E., Peter J. Lambert, and Thor O. Thoresen. 2011. “Horizontal Inequity under a Dual Income Tax System: Principles and Measurement.” *International Tax and Public Finance* 19 (5): 625–640.

Bonhomme, Stephane, and Jean-Marc Robin. 2009. “Assessing the Equalizing Force of Mobility Using Short Panels: France, 1990-2000.” *The Review of Economic Studies* 76: 63–92.

Cherubini, Umberto, Elisa Luciano, and Walter Vecchiato. 2004. *Copula Methods in Finance*. Wiley.com.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States.” *NBER Working Paper 19843*.

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.” *NBER Working Paper 19844*.

Cruces, Guillermo, Peter Lanjouw, Leonardo Lucchetti, Elizaveta Perova, Renos Vakis, and Mariana Viollaz. 2011. “Intra-Generational Mobility and Repeated Cross-Sections: A Three-Country Validation Exercise.” *World Bank Policy Research Working Paper 5916*.

Dang, Hai-Anh, and Peter Lanjouw. 2013. “Measuring Poverty Dynamics with Synthetic Panels Based on Cross-Sections.” *World Bank Policy Research Working Paper 6504*.

Dang, Hai-Anh, Peter Lanjouw, Jill Luoto, and David McKenzie. 2011. “Using Repeated Cross-Sections to Explore Movements in and out of Poverty.” *World Bank Policy Research Working Paper 5550*.

Deaton, Angus. 1985. “Panel Data from Time Series of Cross-Sections.” *Journal of Econometrics* 30: 109–126.

Demarta, Stefano, and Alexander J. McNeil. 2007. “The T Copula and Related Copulas.” *International Statistical Review* 73 (1): 111–129.

Dunn, Christopher E. 2007. “The Intergenerational Transmission of Lifetime Earnings: Evidence from Brazil.” *The BE Journal of Economic Analysis & Policy* 7 (2).

Elbers, Chris, Jean O. Lanjouw, and Peter Lanjouw. 2003. “Micro–Level Estimation of Poverty and Inequality.” *Econometrica* 71 (1): 355–364.

Emran, M. Shahe, and Forhad Shilpi. 2011. “Intergenerational Occupational Mobility in Rural Economy.” *The Journal of Human Resources* 46 (2).

———. 2012. “Gender, Geography and Generations Intergenerational Educational Mobility in Post-Reform India.” *World Bank Policy Research Working Paper 6055*.

Emran, M. Shahe, and Yan Sun. 2012. “Magical Transition? Intergenerational Educational and Occupational Mobility in Rural China”: 1988–2002.

Fields, Gary S., and Mariana Viollaz. 2013. “Can the Limitations of Panel Datasets Be Overcome by Using Pseudo-Panels to Estimate Income Mobility?”

Foster, James E., and Jonathan Rothbaum. 2014. “The Mobility Curve: Measuring the Impact of Mobility on Welfare.” http://home.gwu.edu/~jrothbau/Rothbaum_Jonathan_MobilityCurves.pdf.

Guimaraes Ferreira, Sergio, and Fernando A. Veloso. 2006. “Intergenerational Mobility of Wages in Brazil.” *Brazilian Review of Econometrics* 26 (2): 181–211.

Hertz, Tom. 2001. “Education, Inequality and Economic Mobility in South Africa.” *PhD Dissertation*.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2007. “The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends.” *The BE Journal of Economic Analysis & Policy* 7 (2).

Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh B. Paul. 2013. “Breaking the Caste Barrier Intergenerational Mobility in India.” *Journal of Human Resources* 48 (2): 435–473.

Jäntti, Markus, Eva Sierminska, and Philippe Van Kerm. 2012. “Modelling the Joint Distribution of Income and Wealth.”

Jaworski, Piotr, Fabrizio Durante, Wolfgang Härdle, and Tomasz Rychlik, ed. 2010. *Copula Theory and Its Applications*.

Madrian, Brigitte C., and Lars John Lefgren. 2000. “An Approach to Longitudinally Matching Current Population Survey (CPS) Respondents.” *Journal of Economic and Social Measurement* 26: 31–62.

Mazumder, Bhashkar. 2005. “Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data.” *Review of Economics and Statistics* 87 (2): 235–255.

Nimubona, Alain-Désiré, and Désiré Vencatachellum. 2007. “Intergenerational Education Mobility of Black and White South Africans.” *Journal of Population Economics* 20 (1): 149–182.

Nunez, Javier I., and Leslie Miranda. 2010. “Intergenerational Income Mobility in a Less-Developed, High-Inequality Context: The Case of Chile.” *The BE Journal of Economic Analysis & Policy* 10 (1).

Quheng, Deng, Björn Gustafsson, and Li Shi. 2012. “Intergenerational Income Persistency in Urban China.” *IZA Discussion Paper No. 6907*.

Torche, Florencia. 2005. “Unequal But Fluid: Social Mobility in Chile in Comparative Perspective.” *American Sociological Review* 70 (3): 422–450.

Trivedi, Pravin K., and David M. Zimmer. 2006. *Copula Modeling: An Introduction for Practitioners*. *Foundations and Trends in Econometrics*. Vol. 1.

US Census Bureau. 2013. “Estimating ASEC Variances with Replicate Weights.” thedataweb.rm.census.gov/ftp/cps_ftp.html.

Velez Grajales, Roberto, Raymundo Campos Vazquez, and Juan Enrique Huerta Wong. 2013. “Informe Movilidad Social En Mexico 2013”. Mexico D.F., Mexico.

Vinh, Andrea, William E. Griffiths, and Duangkamon Chotikapanich. 2010. "Bivariate Income Distributions for Assessing Inequality and Poverty under Dependent Samples." *Economic Modelling* 27 (6): 1473–1483.

Zimmer, David M., and H. Youn Kim. 2012. "The Dependence Structure of Income Distribution." *Applied Economics* 44 (27): 3573–3583.

Tables

Table 1: OLS Regression Results for Log Income in the PSID-CNEF

	2005	2007
Years of Education	0.155***	0.157***
	(0.005)	(0.006)
Age	0.0425***	0.0590***
	(0.0136)	(0.0157)
Age squared	-0.000313*	-0.000538***
	(0.000171)	(0.000189)
Male	0.698***	0.593***
	(0.029)	(0.032)
Black	-0.362***	-0.429***
	(0.027)	(0.030)
Hispanic	-0.579***	-0.435***
	(0.084)	(0.093)
Constant	6.66***	6.45***
	(0.27)	(0.33)
R Squared	0.33	0.27
N	4,704	4,704

Notes: * p<0.1, ** p<0.05, *** p<0.01.
Robust standard errors in parenthesis.

Table 2: OLS Regression Results for Log Income in the CPS ASEC

	2005	2006
Education (High school excluded)		
Less than High School	-0.455***	-0.419***
	(0.020)	(0.018)
Some College	0.189***	0.186***
	(0.012)	(0.012)
College	0.534***	0.542***
	(0.013)	(0.013)
Masters	0.679***	0.692***
	(0.019)	(0.019)
Professional/PhD	0.941***	0.964***
	(0.030)	(0.029)
Age	0.005	0.003
	(0.006)	(0.006)
Age squared	-0.000097	-0.000125*
	(0.000073)	(0.000072)
Black	-0.436***	-0.442***
	(0.048)	(0.048)
Hispanic	-0.235***	-0.224***
	(0.014)	(0.014)
Constant	10.071***	10.095***
	(0.115)	(0.118)
R Squared	0.21	0.21
N	22,156	22,156

Notes: * p<0.1, ** p<0.05, *** p<0.01.
Robust standard errors in parenthesis.

Table 3: OLS Regression Results for Log Income in the NLSY79

	Parents	Children
Years of Education	0.063***	0.009***
	(0.002)	(0.002)
Urban	0.180***	0.137***
	(0.023)	(0.028)
Regions (North East excluded)		
North Central	0.059**	-0.186***
	(0.028)	(0.035)
South	-0.047*	-0.110***
	(0.027)	(0.032)
West	0.078***	-0.049
	(0.031)	(0.037)
Race/Ethnicity (White excluded)		
Black	-0.639***	-0.548***
	(0.026)	(0.030)
Hispanic	-0.403***	-0.372***
	(0.029)	(0.033)
Other Non-White	-0.448***	-0.158***
	(0.025)	(0.037)
Age	-0.006***	0.423*
	(0.002)	(0.217)
Age squared	0.00011***	-0.01260*
	(0.00004)	(0.00675)
Industry of Household Head (Retail excluded)		
Agriculture	-0.148**	-0.141
	(0.067)	(0.079)
Mining	-0.004	-0.090
	(0.370)	(0.150)
Manufacturing	-0.223*	0.113**
	(0.100)	(0.049)
Transportation	-0.018	0.233***
	(0.053)	(0.037)
Finance	-0.014	0.337***
	(0.130)	(0.053)
Business Services	0.029	0.444***
	(0.099)	(0.053)
Personal Services	-0.099	0.064
	(0.077)	(0.049)
Recreation	-0.201***	-0.223***
	(0.046)	(0.061)
Professional Services	-0.186*	0.217
	(0.099)	(0.117)
Other	-0.246***	0.261***
	(0.051)	(0.037)
Constant	9.54***	6.69***
	(0.05)	(1.73)
R Squared	0.33	0.21
N	5,005	5,005

Notes: * p<0.1, ** p<0.05, *** p<0.01.
Robust standard errors in parenthesis.

Table 4: OLS Regression Results for Log Income in the NLSY97

	Parents	Children
Years of Education	0.066***	0.059***
	(0.003)	(0.003)
Urban	0.105***	-0.010
	(0.030)	(0.032)
Regions (North East excluded)		
North Central	0.094**	-0.114***
	(0.042)	(0.043)
South	0.037	-0.039
	(0.040)	(0.040)
West	0.065	0.005
	(0.044)	(0.044)
Race/Ethnicity (White excluded)		
Black	-0.797***	-0.625***
	(0.034)	(0.033)
Hispanic	-0.599***	-0.221***
	(0.038)	(0.036)
Other Non-White	-0.123*	-0.109
	(0.069)	(0.069)
Constant	9.37***	9.75***
	(0.05)	(0.06)
R Squared	0.28	0.16
N	4,685	4,685

Notes: * p<0.1, ** p<0.05, *** p<0.01.
Robust standard errors in parenthesis.

Table 5: Summary Statistics for ENIGH and Parent Cohort of EMOVI

	ENIGH84	EMOVI06 Parents	ENIGH89	EMOVI11 Parents
Age (in parent survey year)	40.4	43.6	40.4	47.9
	(12.7)	(9.8)	(12.7)	(9.4)
Education (Primary education excluded)				
No or Some Primary	0.551	0.560	0.655	0.443
Secondary	0.131	0.079	0.131	0.211
University	0.055	0.027	0.004	0.058
Occupation (Retail excluded)				
Professionals	0.016	0.013	0.028	0.027
Technical	0.023	0.005	0.031	0.027
Education	0.027	0.011	0.029	0.008
Entertainment	0.010	0.008	0.008	0.002
Directors	0.018	0.001	0.022	0.001
Agriculture	0.319	0.325	0.263	0.209
Industrial Directors	0.022	0.007	0.025	0.009
Industrial Workers	0.226	0.250	0.224	0.261
Administrative	0.050	0.014	0.050	0.010
Retail (Ambulatory)	0.011	0.011	0.019	0.012
Service	0.044	0.027	0.046	0.034
Domestic Service	0.013	0.018	0.011	0.000
Transportation	0.051	0.130	0.058	0.092
Region (Northern excluded)				
Central	0.412	0.332	0.356	0.274
West	0.117	0.143	0.128	0.056
East	0.141	0.160	0.140	0.103
South	0.102	0.138	0.168	0.118
City Size (Large, > 100,000 excluded)				
Very Small (≤10,000)	0.099	0.118	0.057	0.065
Small (10,001-35,000)	0.238	0.220	0.168	0.156
Medium (35,001-100,000)	0.228	0.174	0.221	0.187
Owned Car	0.146	0.204	0.170	0.234
Indoor Plumbing	0.616	0.491	0.704	0.580
Electricity in Home	0.872	0.717	0.884	0.906
Owned Telephone	0.143	0.183	0.152	0.162
Owned TV	0.674	0.532	0.783	0.807
N	3,076	1,388	7,307	1,042

Notes: Standard deviation for household head age in parenthesis. All other variables are binary and the standard deviation is equal to $\left(p(1-p)\right)^{\frac{1}{2}}$.

Table 6: OLS Regression Results for Log Income in the ENIGH and EMOVI

	ENIGH84-EMOVI06		ENIGH89-EMOVI11	
	Parents (1984)	**Children (2006)**	**Parents (1989)**	**Children (2011)**
Age	0.0052	0.0318	0.0049	-0.0239
	(0.0043)	(0.1455)	(0.0030)	(0.1838)
Age squared	0.000021	0.000571	0.000020	0.000276
	(0.000046)	(0.002110)	(0.000031)	(0.002688)
Education (Primary education excluded)				
No/Some Primary	-0.198***	-0.151***	-0.170***	-0.025
	(0.027)	(0.056)	(0.020)	(0.079)
Secondary	0.106***	0.083**	0.177***	0.169***
	(0.035)	(0.037)	(0.027)	(0.052)
Tertiary	0.331***	0.647***	0.575***	0.544***
	(0.062)	(0.077)	(0.108)	(0.098)
Occupation (Retail excluded)				
Professionals	0.116	0.230*	0.252***	0.407***
	(0.092)	(0.114)	(0.051)	(0.121)
Technical	0.030	0.184*	0.045	0.331
	(0.071)	(0.103)	(0.043)	(0.165)
Education	0.008	0.287*	0.113***	0.165
	(0.076)	(0.137)	(0.045)	(0.168)
Entertainment	-0.108	0.173	0.185**	-0.016
	(0.102)	(0.179)	(0.074)	(0.317)
Directors	0.447***	0.229	0.460***	0.618***
	(0.081)	(0.179)	(0.053)	(0.226)
Agriculture	-0.132***	-0.329***	-0.112***	-0.079
	(0.034)	(0.070)	(0.024)	(0.093)
Industrial Directors	0.185***	0.208*	0.269***	0.176
	(0.071)	(0.132)	(0.050)	(0.165)
Industrial Workers	-0.082***	0.069*	-0.018	0.099*
	(0.032)	(0.041)	(0.022)	(0.058)
Administrative	0.053	0.083	0.151***	0.192***
	(0.051)	(0.077)	(0.036)	(0.086)
Retail (Ambulatory)	-0.095	-0.075	-0.079	-0.209*
	(0.083)	(0.135)	(0.056)	(0.135)
Service	-0.076*	-0.067	0.015	0.157**
	(0.049)	(0.072)	(0.037)	(0.087)
Domestic Service	-0.414***	-0.317***	-0.002	-0.157
	(0.095)	(0.110)	(0.062)	(0.119)
Transportation	0.010	-0.069	0.064**	0.087
	(0.047)	(0.067)	(0.033)	(0.213)
Region (Northern excluded)				
Central	-0.145***	-0.159***	-0.166***	-0.094*
	(0.026)	(0.042)	(0.019)	(0.061)
West	-0.138***	0.033	-0.133***	-0.233***
	(0.033)	(0.044)	(0.022)	(0.074)
East	-0.181***	-0.225***	-0.211***	-0.152**
	(0.032)	(0.067)	(0.023)	(0.068)
South	-0.045	-0.380***	-0.216***	-0.226***
	(0.032)	(0.069)	(0.024)	(0.067)

Table 6: OLS Regression Results for Log Income in the ENIGH and EMOVI, continued

	ENIGH84-EMOVI06		ENIGH89-EMOVI11	
	Parents (1984)	**Children (2006)**	**Parents (1989)**	**Children (2011)**
City Size (Large, > 100k excluded)				
Very Small (≤10k)	-0.127***	-0.114	-0.302***	-0.114
	(0.041)	(0.078)	(0.033)	(0.102)
Small (10k-35k)	-0.192***	-0.210***	-0.217***	-0.124**
	(0.028)	(0.049)	(0.022)	(0.059)
Medium (35k-100k)	-0.035	-0.053	-0.146***	-0.195***
	(0.027)	(0.053)	(0.021)	(0.051)
Owned Car	0.387***	0.232***	0.365***	0.284***
	(0.032)	(0.036)	(0.021)	(0.044)
Indoor Plumbing	0.200***	0.279***	0.180***	0.226***
	(0.024)	(0.051)	(0.018)	(0.053)
Electricity in Home	0.045	0.379**	0.126***	-0.141
	(0.035)	(0.132)	(0.025)	(0.141)
Owned Telephone	0.256***	0.157***	0.362***	0.022
	(0.033)	(0.036)	(0.021)	(0.047)
Owned TV	0.273***	0.239***	0.302***	0.140
	(0.027)	(0.071)	(0.021)	(0.132)
Constant	7.40***	6.5***	7.2***	7.9**
	(0.10)	(2.50)	(0.07)	(3.12)
R Squared	0.50	0.41	0.50	0.27
N	3,076	1,388	7,307	1,042

Notes: * p<0.1, ** p<0.05, *** p<0.01.
Robust standard errors in parenthesis.

Table 7: Copula Estimate of Quintile Transition Matrices with Standard Errors

Cohort 1: 1984-2006 ($\rho_{Estimate} = 0.12$)

Quintiles Parent	**Child 1st**	**2nd**	**3rd**	**4th**	**5th**
1st	34.6	26.2	17.5	13.6	7.9
	(2.5)	(2.2)	(2.0)	(1.6)	(1.5)
2nd	25.6	24.1	19.3	18.1	13.0
	(2.5)	(2.0)	(2.2)	(2.1)	(1.8)
3rd	19.0	20.9	21.0	20.9	18.0
	(2.4)	(2.2)	(2.1)	(2.2)	(1.9)
4th	13.1	17.2	21.6	23.9	24.4
	(1.8)	(2.2)	(2.4)	(2.3)	(2.0)
5th	7.4	11.8	20.4	23.7	36.9
	(1.6)	(1.7)	(2.0)	(1.8)	(2.4)

Cohort 2: 1989-2011 ($\rho_{Estimate} = 0.18$)

Quintiles Parent	**Child 1st**	**2nd**	**3rd**	**4th**	**5th**
1st	37.9	23.8	17.6	12.6	8.0
	(2.9)	(2.6)	(2.6)	(2.1)	(2.0)
2nd	25.3	24.5	19.8	17.9	12.2
	(2.4)	(2.7)	(2.6)	(2.3)	(2.1)
3rd	17.7	21.9	21.4	22.1	17.1
	(2.2)	(2.6)	(2.9)	(2.8)	(2.2)
4th	12.0	17.8	21.9	24.4	23.6
	(2.1)	(2.2)	(2.5)	(2.3)	(2.3)
5th	7.0	11.9	19.5	22.8	38.5
	(1.5)	(1.9)	(2.3)	(2.2)	(2.7)

Table 8: Copula Estimate of Decile Transition Matrices with Standard Errors

Cohort 1: 1984-2006 ($\rho_{Estimate} = 0.12$)

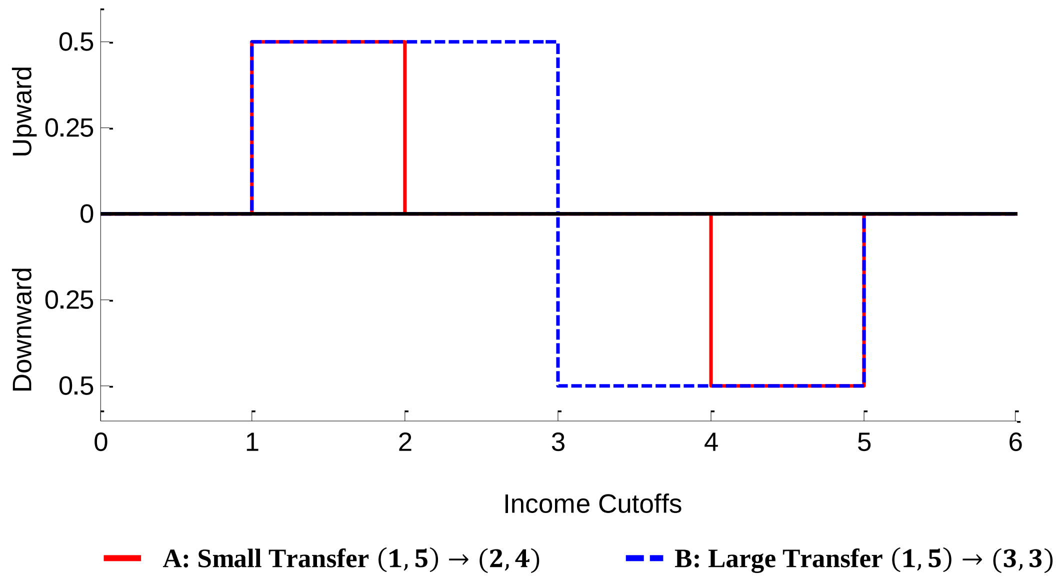
Deciles					**Child**					
Parent	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	21.7	15.5	14.8	12.2	7.8	9.0	7.0	5.0	3.9	2.5
	(3.3)	(2.7)	(2.7)	(2.7)	(2.2)	(2.4)	(2.2)	(1.6)	(1.4)	(1.2)
2	17.2	14.8	13.8	11.6	8.5	9.6	8.9	6.3	5.6	3.9
	(3.0)	(2.5)	(2.7)	(2.3)	(2.3)	(2.2)	(2.2)	(1.9)	(1.8)	(1.6)
3	13.9	13.2	13.5	11.8	8.8	9.6	9.8	7.7	6.9	4.9
	(2.8)	(2.9)	(2.5)	(2.5)	(2.3)	(2.4)	(2.6)	(2.2)	(2.2)	(1.6)
4	11.6	12.5	12.0	10.9	10.2	10.0	9.8	8.9	8.2	6.0
	(2.6)	(3.2)	(2.8)	(2.4)	(2.4)	(2.6)	(2.2)	(2.7)	(2.4)	(2.1)
5	9.6	11.1	10.7	11.0	10.6	10.1	10.1	9.5	9.8	7.0
	(2.5)	(2.3)	(2.4)	(2.6)	(2.6)	(2.2)	(2.1)	(2.3)	(2.6)	(2.1)
6	7.7	9.7	9.8	10.3	11.0	10.4	11.2	10.9	11.1	8.1
	(2.2)	(2.3)	(2.4)	(2.3)	(2.5)	(2.5)	(2.8)	(2.5)	(2.5)	(2.3)
7	6.5	8.4	8.4	9.5	11.2	10.6	11.6	11.7	11.5	10.7
	(1.9)	(2.3)	(2.4)	(2.6)	(2.8)	(2.4)	(2.5)	(2.9)	(2.3)	(2.3)
8	5.0	6.4	7.5	8.9	10.8	10.6	11.5	12.9	13.2	13.3
	(1.8)	(2.0)	(2.1)	(2.2)	(2.6)	(2.7)	(2.7)	(2.6)	(2.5)	(2.6)
9	3.5	5.1	5.4	7.6	10.8	10.8	11.0	14.2	14.9	16.8
	(1.6)	(1.8)	(1.7)	(2.2)	(2.3)	(2.4)	(2.5)	(2.4)	(2.9)	(2.6)
10	2.8	3.4	4.2	6.3	9.7	9.6	9.1	13.0	15.1	27.0
	(1.5)	(1.6)	(1.6)	(1.9)	(2.2)	(2.1)	(2.2)	(2.7)	(2.8)	(2.9)

Cohort 2: 1989-2011 ($\rho_{Estimate} = 0.18$)

Deciles					**Child**					
Parent	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	28.0	14.1	12.1	12.1	8.3	7.8	5.3	5.3	3.8	3.0
	(4.3)	(3.5)	(2.9)	(3.0)	(2.4)	(2.2)	(2.3)	(2.0)	(2.0)	(1.8)
2	19.3	14.3	11.8	11.7	9.6	9.4	7.5	7.1	5.6	3.6
	(3.5)	(2.9)	(2.5)	(3.2)	(2.4)	(2.7)	(2.1)	(2.4)	(2.2)	(1.8)
3	14.0	13.3	12.4	12.5	9.6	9.6	8.7	8.0	6.9	4.6
	(3.0)	(3.2)	(3.1)	(3.0)	(2.5)	(2.4)	(2.5)	(2.6)	(2.3)	(1.8)
4	10.9	12.5	12.5	11.5	10.2	10.2	10.1	9.0	7.5	5.3
	(3.0)	(2.9)	(3.0)	(2.9)	(2.9)	(3.1)	(2.5)	(2.6)	(2.2)	(2.0)
5	8.3	10.9	11.7	11.1	10.5	10.6	10.5	10.6	8.6	7.0
	(2.4)	(2.9)	(2.8)	(2.8)	(3.0)	(2.8)	(2.8)	(2.9)	(2.7)	(2.5)
6	6.4	9.8	10.7	10.3	10.7	11.1	11.8	11.3	10.0	8.6
	(2.0)	(2.5)	(2.5)	(3.1)	(3.2)	(2.9)	(2.9)	(3.0)	(3.1)	(2.3)
7	4.9	8.3	9.3	9.5	10.4	11.7	12.2	11.9	10.7	10.7
	(2.0)	(2.3)	(2.6)	(2.3)	(3.0)	(3.1)	(2.9)	(3.1)	(3.0)	(2.8)
8	3.4	7.3	8.6	8.3	10.5	11.3	11.9	12.7	12.9	12.9
	(1.7)	(2.3)	(2.5)	(2.7)	(2.7)	(3.2)	(2.8)	(2.8)	(3.3)	(3.3)
9	2.8	5.4	6.9	7.3	10.7	10.7	11.3	12.5	14.7	17.2
	(1.7)	(2.1)	(2.4)	(2.4)	(2.9)	(2.8)	(2.7)	(3.6)	(3.1)	(3.3)
10	1.8	4.0	3.9	5.5	9.1	8.6	10.5	11.3	18.8	26.3
	(1.2)	(1.5)	(1.8)	(1.9)	(2.4)	(2.5)	(2.5)	(2.9)	(3.1)	(3.1)

Figures

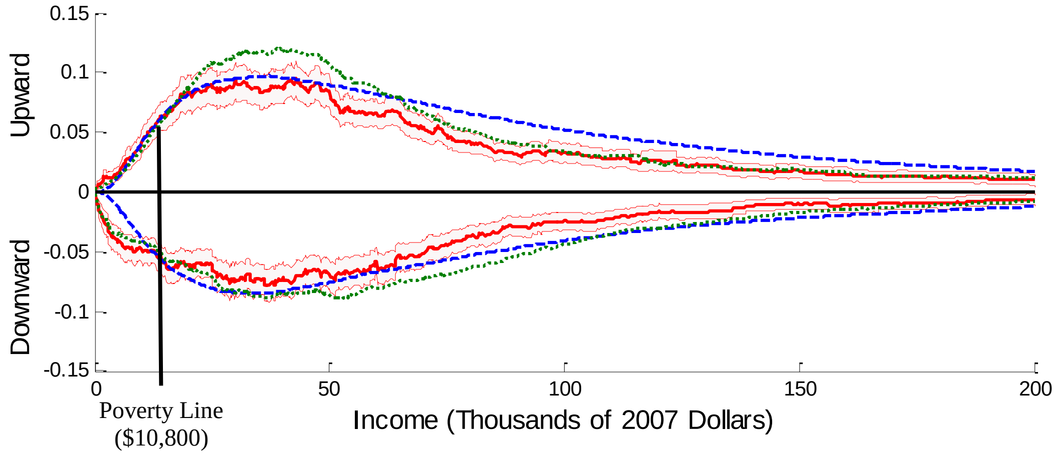
Figure 1: Mobility Curve Example



Source: Foster and Rothbaum 2014

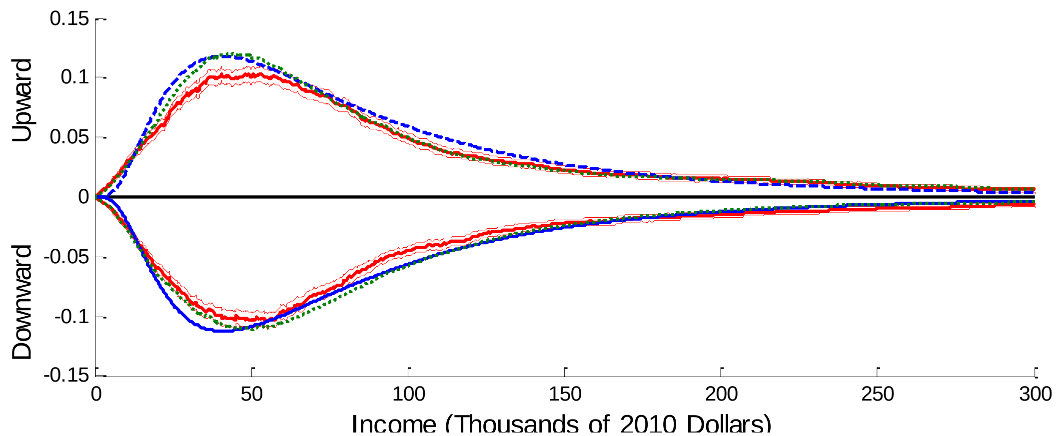
Notes: Example of mobility curve first-order dominance ordering in two societies, A and B. At each cutoff (c) on the x-axis, the mobility curve shows the share of the population that is upwardly mobile ($y_{i1} < c$ and $y_{i2} \geq c$) above the x-axis and the share that is downwardly mobile ($y_{i1} \geq c$ and $y_{i2} < c$) below the x-axis. Society B has more upward mobility than A ($1 \rightarrow 3$ vs. $1 \rightarrow 2$) which is also shown in the mobility curve by the fact that upward mobility for B is greater than or equal to upward mobility for A at all cutoffs. The same is true for downward mobility. Thus, B first-order mobility dominates A in both upward and downward mobility which also means that the welfare gain (loss) due to upward (downward) mobility is greater in B than A.

Figure 2: US PSID-CNEF 2005-2007 Synthetic vs. True Mobility



True Mobility (with 95% Confidence Bands) **DLLM Estimate (ρ_{Actual})** **Gaussian Copula (ρ_{Actual})**

Figure 3: US CPS ASEC Panel 2005-2006 Synthetic vs. True Mobility of Income



True Mobility (with 95% Confidence Bands) **DLLM Estimate (ρ_{Actual})** **Gaussian Copula (ρ_{Actual})**

Notes: True upward and downward mobility curve with 95% confidence intervals and DLLM and Guassian copula synthetic panel estimates using the known correlation of the OLS errors ρ_{Actual}. At each cutoff (c) on the x-axis, the mobility curve shows the share of the population that is upwardly mobile ($y_{i1} < c$ and $y_{i2} \geq c$) above the x-axis and the share that is downwardly mobile ($y_{i1} \geq c$ and $y_{i2} < c$) below the x-axis. The DLLM estimate underestimates upward and downward mobility at low income cutoffs and overestimates them at higher income cutoffs. However, the estimate of transitions into and out of poverty (mobility at the poverty line) are very accurate.

Figure 4: NLSY79 Intergenerational Mobility of Income
Comparison of True Mobility to DLLM and Gaussian Copula at $\rho_{Actual} = 0.24$

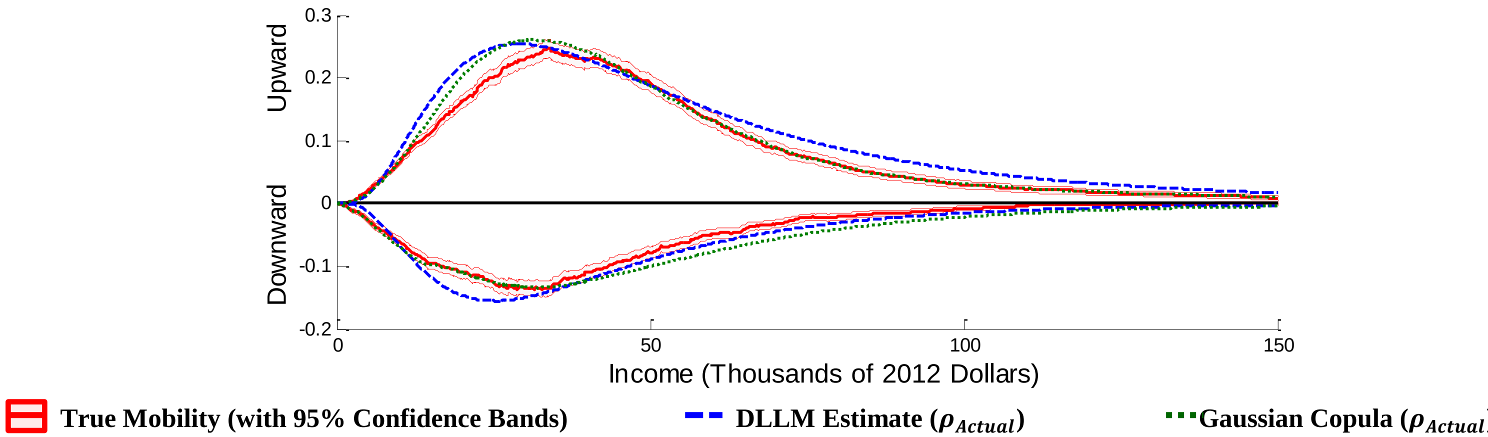
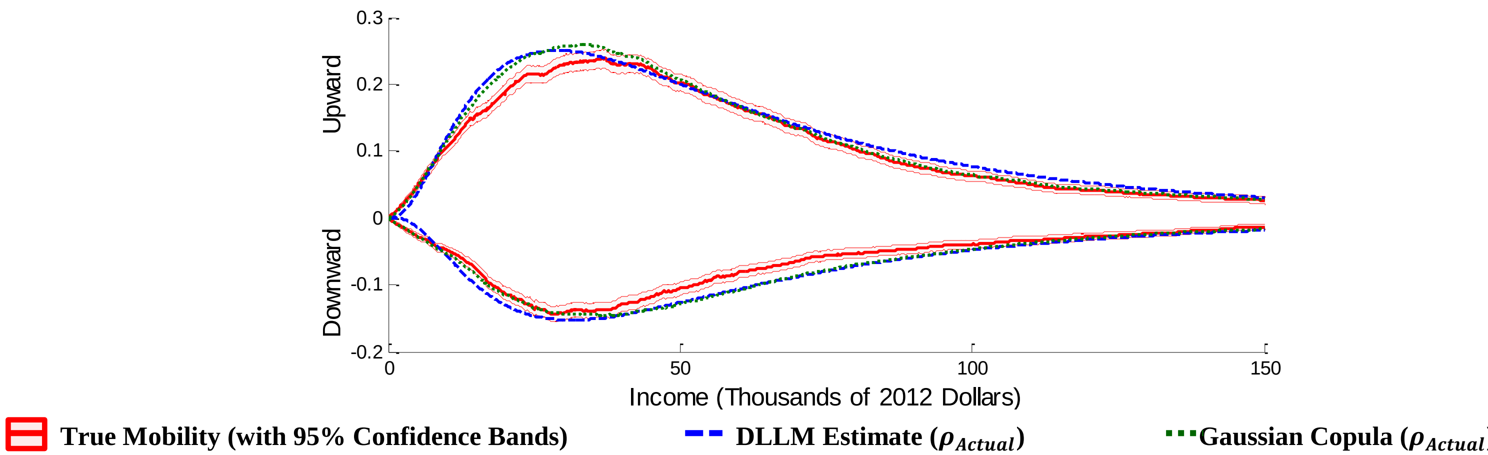
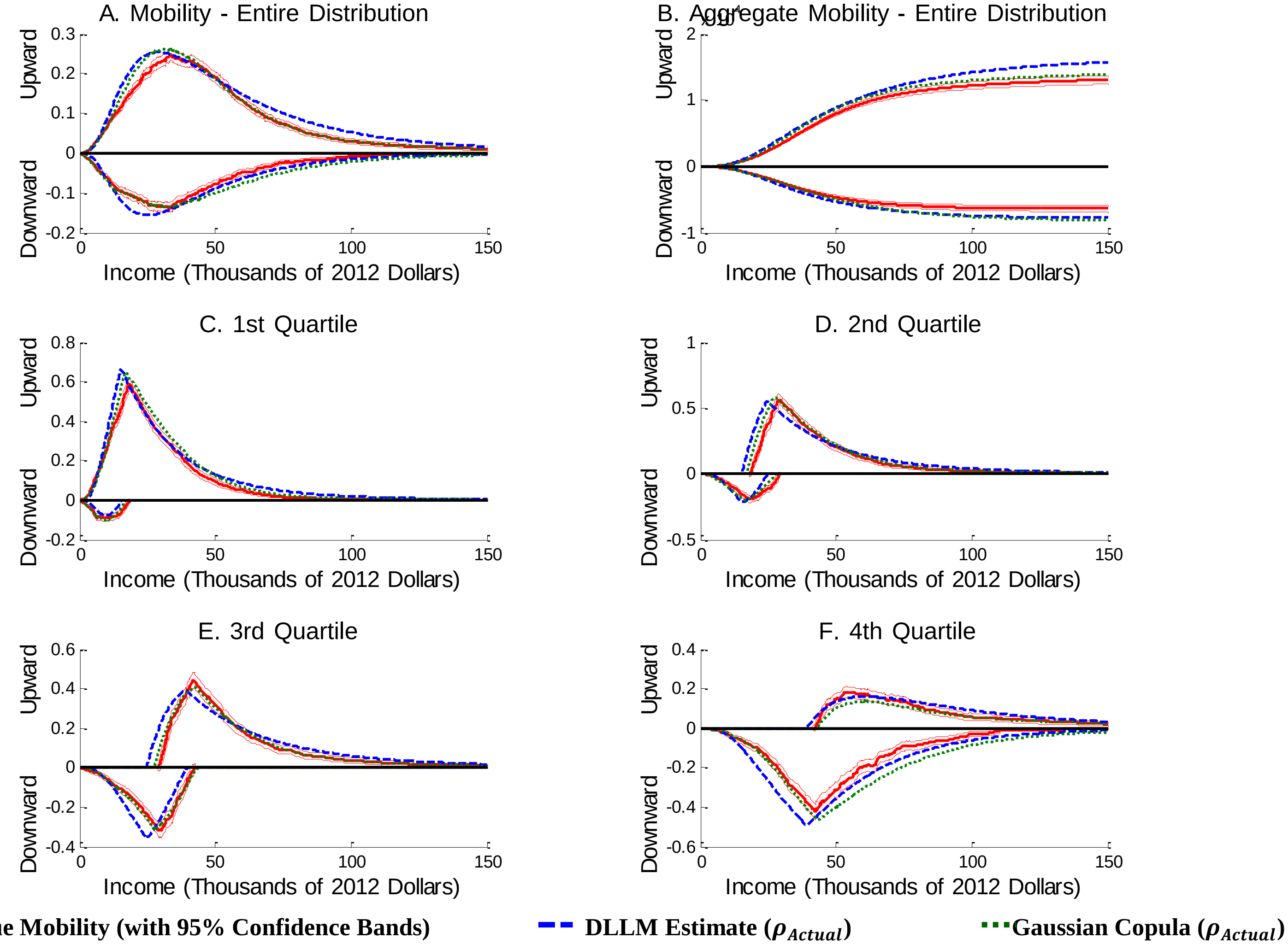


Figure 5: NLSY97 Intergenerational Mobility of Income
Comparison of True Mobility to DLLM and Gaussian Copula at $\rho_{Actual} = 0.24$



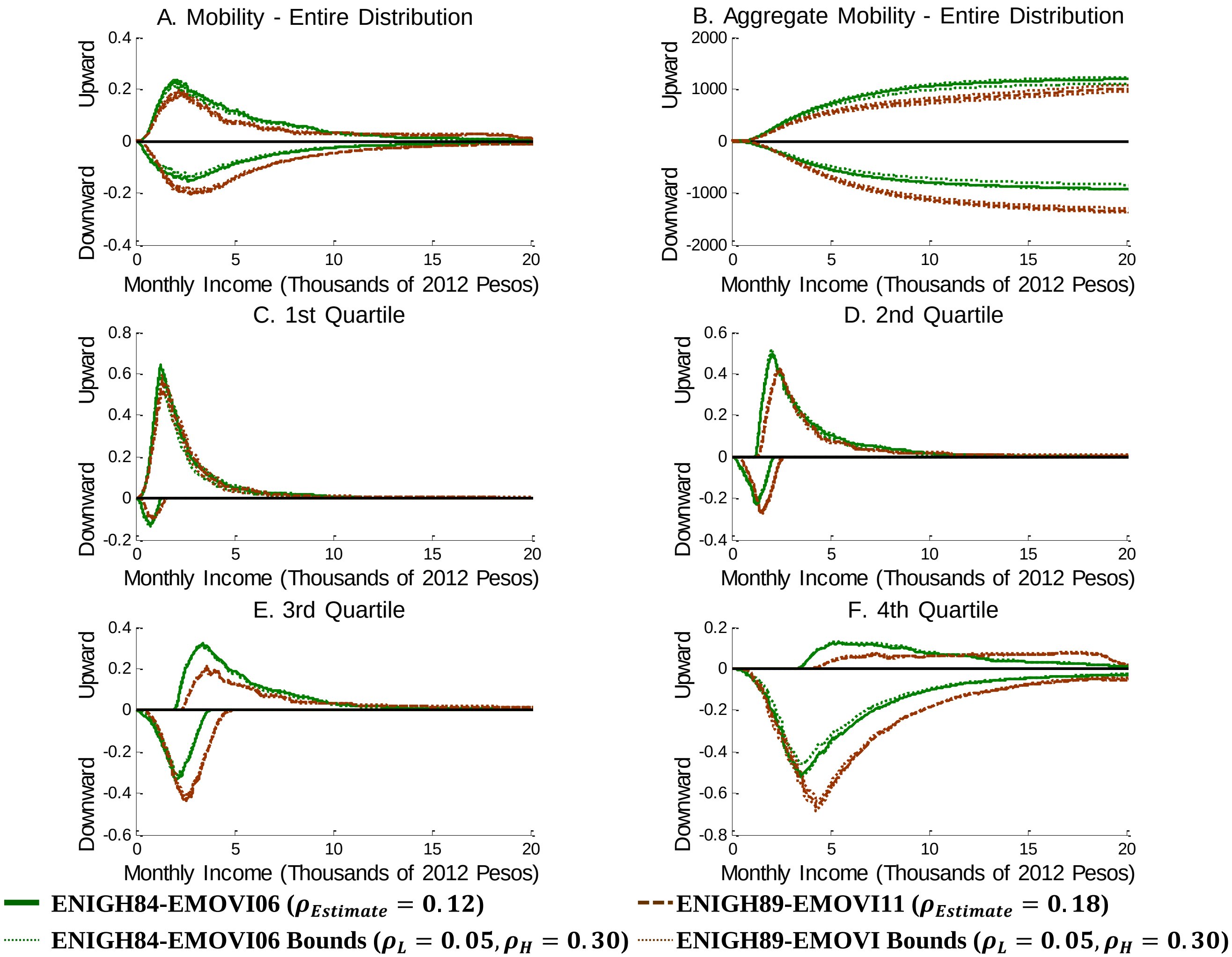
Notes: True upward and downward mobility curve with 95% confidence intervals and DLLM and Guassian copula synthetic panel estimates using the known correlation of the OLS errors ρ_{Actual}. At each cutoff (c) on the x-axis, the mobility curve shows the share of the population that is upwardly mobile ($y_{i1} < c$ and $y_{i2} \geq c$) above the x-axis and the share that is downwardly mobile ($y_{i1} \geq c$ and $y_{i2} < c$) below the x-axis. The DLLM estimate underestimates upward and downward mobility at low income cutoffs and overestimates them at higher income cutoffs. However, the estimate of transitions into and out of poverty (mobility at the poverty line) are very accurate.

Figure 6: Complete Mobility Curve Results for NLSY79 with Quartile Decomposition
Comparison of True Mobility to DLLM and Gaussian Copula at $\rho_{Actual} = 0.24$



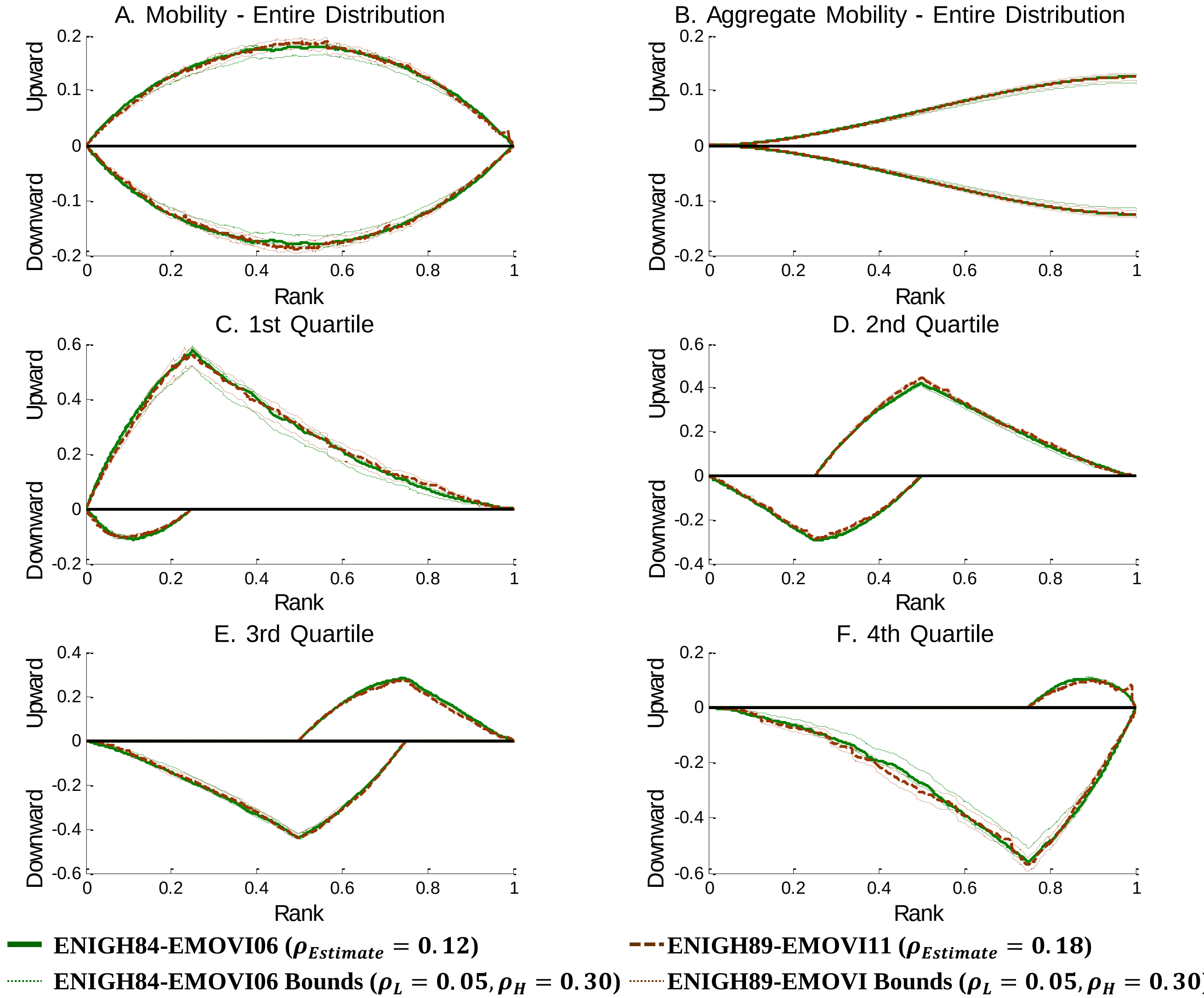
Notes: True upward and downward mobility curve and quartile decomposition for NLSY79 with 95% confidence intervals and DLLM and Guassian copula synthetic panel estimates using the known correlation of the OLS errors ρ_{Actual}. At each cutoff (c) on the x-axis, the mobility curve shows the share of the population that is upwardly mobile ($y_{i1} < c$ and $y_{i2} \geq c$) above the x-axis and the share that is downwardly mobile ($y_{i1} \geq c$ and $y_{i2} < c$) below the x-axis.

Figure 7: Mobility Curve and Quartile Decomposition Estimate of Intergenerational Mobility in Mexico Using Copula with Bounds



Notes: Copula estimate of the upward and downward mobility curve and quartile decomposition for the ENIGH84-EMOVI06 (born 1966-1976) and ENIGH89-EMOVI11 (born 1971-1981) using the estimated correlation of the OLS errors $\rho_{Estimated}$ and lower and upper bounds. At each cutoff (c) on the x-axis, the mobility curve shows the share of the population that is upwardly mobile ($y_{i1} < c$ and $y_{i2} \geq c$) above the x-axis and the share that is downwardly mobile ($y_{i1} \geq c$ and $y_{i2} < c$) below the x-axis.

Figure 8: Quartile Decomposition of Rank Mobility Curve Estimate of Intergenerational Mobility in Mexico Using Copula with Bounds



Notes: Copula estimate of the rank upward and downward mobility curve and quartile decomposition for the ENIGH84-EMOVI06 (born 1966-1976) and ENIGH89-EMOVI11 (born 1971-1981) using the estimated correlation of the OLS errors $\rho_{Estimated}$ and lower and upper bounds. At each cutoff (c) on the x-axis, the mobility curve shows the share of the population that is upwardly mobile ($F_1^{-1}(y_{i1}) < c$ and $F_2^{-1}(y_{i2}) \geq c$) above the x-axis and the share that is downwardly mobile ($F_1^{-1}(y_{i1}) \geq c$ and $F_2^{-1}(y_{i2}) < c$) below the x-axis.

Appendix 1. Synthetic Panels Estimates of Other Measures of Mobility

Table A1.1: Comparison of True Intergenerational Elasticity to Synthetic Panel Estimates

	True	DLLM			Copula		
PSID-CNEF (2005-2007)	0.76	$\rho_L = 0.63$ 0.77	$\rho_{Actual} = 0.78$ 0.87	$\rho_H = 0.88$ 0.94	$\rho_L = 0.63$ 0.71	$\rho_{Actual} = 0.78$ 0.86	$\rho_H = 0.88$ 0.96
CPS ASEC (2005-2006)	0.67	$\rho_L = 0.58$ 0.67	$\rho_{Actual} = 0.68$ 0.75	$\rho_H = 0.78$ 0.82	$\rho_L = 0.58$ 0.63	$\rho_{Actual} = 0.68$ 0.71	$\rho_H = 0.78$ 0.79
NLSY79	0.44	$\rho_L = 0.09$ 0.24	$\rho_{Actual} = 0.24$ 0.35	$\rho_H = 0.39$ 0.47	$\rho_L = 0.09$ 0.32	$\rho_{Actual} = 0.24$ 0.43	$\rho_H = 0.39$ 0.54
NLSY97	0.33	$\rho_L = 0.09$ 0.21	$\rho_{Actual} = 0.24$ 0.31	$\rho_H = 0.39$ 0.41	$\rho_L = 0.09$ 0.23	$\rho_{Actual} = 0.24$ 0.33	$\rho_H = 0.39$ 0.44
ENIGH84-EMOVI06		$\rho_L = 0.05$ 0.26	$\rho_{Estimate} = 0.12$ 0.30	$\rho_H = 0.30$ 0.42	$\rho_L = 0.05$ 0.32	$\rho_{Estimate} = 0.12$ 0.35	$\rho_H = 0.30$ 0.45
ENIGH89-EMOVI11		$\rho_L = 0.05$ 0.23	$\rho_{Estimate} = 0.18$ 0.27	$\rho_H = 0.30$ 0.37	$\rho_L = 0.05$ 0.28	$\rho_{Estimate} = 0.18$ 0.33	$\rho_H = 0.30$ 0.40

Table A1.2: Comparison of True Log Income Correlation to Synthetic Panel Estimates

	True	DLLM			Copula		
PSID-CNEF (2005-2007)	0.70	$\rho_L = 0.63$ 0.74	$\rho_{Actual} = 0.78$ 0.84	$\rho_H = 0.88$ 0.91	$\rho_L = 0.63$ 0.65	$\rho_{Actual} = 0.78$ 0.78	$\rho_H = 0.88$ 0.88
CPS ASEC (2005-2006)	0.67	$\rho_L = 0.58$ 0.67	$\rho_{Actual} = 0.68$ 0.75	$\rho_H = 0.78$ 0.83	$\rho_L = 0.58$ 0.65	$\rho_{Actual} = 0.68$ 0.73	$\rho_H = 0.78$ 0.81
NLSY79	0.36	$\rho_L = 0.09$ 0.20	$\rho_{Actual} = 0.24$ 0.31	$\rho_H = 0.39$ 0.42	$\rho_L = 0.09$ 0.25	$\rho_{Actual} = 0.24$ 0.35	$\rho_H = 0.39$ 0.46
NLSY97	0.31	$\rho_L = 0.09$ 0.20	$\rho_{Actual} = 0.24$ 0.32	$\rho_H = 0.39$ 0.44	$\rho_L = 0.09$ 0.20	$\rho_{Actual} = 0.24$ 0.32	$\rho_H = 0.39$ 0.43
ENIGH84-EMOVI06		$\rho_L = 0.05$ 0.31	$\rho_{Estimate} = 0.12$ 0.34	$\rho_H = 0.30$ 0.44	$\rho_L = 0.05$ 0.27	$\rho_{Estimate} = 0.12$ 0.39	$\rho_H = 0.30$ 0.49
ENIGH89-EMOVI11		$\rho_L = 0.05$ 0.27	$\rho_{Estimate} = 0.18$ 0.30	$\rho_H = 0.30$ 0.41	$\rho_L = 0.05$ 0.36	$\rho_{Estimate} = 0.18$ 0.42	$\rho_H = 0.30$ 0.48

Notes: These tables compare the true log income correlation with the estimate from DLLM and Gaussian copula-based synthetic panels using the known OLS error correlation (ρ_{Actual}) and a plausible lower and upper bound of possible correlations (ρ_L and ρ_H). For the ENIGH84-EMOVI06 (born 1966-1976) and ENIGH89-EMOVI11 (born 1971-1981) estimates of intergenerational mobility in Mexico, the true ρ is not known, and we specify a best estimate of the error correlation $\rho_{Estimate}$.

Table A1.3: Comparison of Quintile Transition Matrices

PSID-CNEF (2005-2007)

True

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	65.7	22.5	7.4	3.7	0.4
2nd	23.0	47.9	22.1	6.0	1.3
3rd	5.8	21.6	45.5	20.0	7.2
4th	3.1	5.7	21.2	52.0	17.9
5th	2.4	2.3	3.7	18.1	73.5

DLLM $[\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88]$

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	60.1, 68.9, 76.9	25.1, 23.7, 20.6	10.7, 6.3, 2.3	3.5, 0.9, 0.1	0.5, 0.0, 0.0
2nd	25.3, 24.0, 20.4	32.9, 39.6, 49.8	24.5, 26.1, 25.5	13.3, 9.3, 4.2	3.9, 1.0, 0.1
3rd	10.6, 6.0, 2.4	24.9, 26.3, 25.1	28.5, 35.1, 44.6	24.8, 26.0, 25.5	11.2, 6.6, 2.5
4th	3.3, 0.9, 0.1	13.5, 9.4, 4.5	25.1, 26.0, 25.1	32.4, 39.4, 49.0	25.8, 24.2, 21.3
5th	0.5, 0.1, 0.0	3.6, 1.0, 0.1	11.2, 6.4, 2.5	26.0, 24.4, 21.2	58.8, 68.3, 76.3

Copula $[\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88]$

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	55.5, 66.1, 74.5	25.4, 24.0, 21.5	12.4, 7.6, 3.4	5.4, 2.0, 0.4	1.2, 0.2, 0.0
2nd	24.9, 22.8, 20.4	31.2, 37.8, 46.4	24.5, 26.0, 26.4	14.6, 11.5, 6.4	4.7, 1.9, 0.4
3rd	12.3, 8.2, 4.2	23.8, 24.9, 24.8	27.8, 33.3, 40.7	24.3, 25.4, 26.4	11.8, 8.1, 3.9
4th	5.7, 2.5, 0.7	14.5, 11.0, 6.8	23.9, 25.4, 25.7	30.6, 37.1, 45.0	25.3, 24.0, 21.9
5th	1.5, 0.3, 0.0	5.0, 2.2, 0.5	11.5, 7.8, 4.0	25.0, 24.0, 21.7	57.1, 65.8, 73.9

CPS ASEC (2005-2006)

True

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	64.6	21.3	8.0	3.7	2.4
2nd	21.6	43.0	20.7	9.6	5.1
3rd	7.8	21.0	40.8	20.6	9.9
4th	3.7	9.0	21.2	44.7	21.5
5th	2.3	5.8	9.3	21.3	61.2

DLLM $[\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88]$

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	54.5, 60.3, 67.1	26.0, 25.5, 24.4	13.0, 10.5, 7.2	5.4, 3.3, 1.3	1.2, 0.4, 0.1
2nd	26.0, 25.5, 24.4	29.6, 33.0, 38.2	23.7, 24.8, 25.9	15.3, 13.4, 10.2	5.4, 3.3, 1.3
3rd	12.9, 10.5, 7.2	23.7, 24.8, 25.9	26.7, 29.4, 34.0	23.7, 24.8, 25.8	13.0, 10.5, 7.2
4th	5.4, 3.3, 1.3	15.3, 13.5, 10.2	23.7, 24.8, 25.8	29.7, 33.0, 38.2	26.0, 25.4, 24.5
5th	1.2, 0.4, 0.1	5.5, 3.3, 1.4	13.0, 10.5, 7.2	26.0, 25.5, 24.5	54.4, 60.3, 66.9

Copula $[\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88]$

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	54.5, 61.3, 68.3	25.3, 24.8, 23.4	12.3, 9.9, 6.7	5.5, 3.4, 1.5	1.4, 0.6, 0.1
2nd	26.0, 25.4, 23.8	30.2, 33.5, 39.1	23.5, 24.5, 25.3	14.7, 13.1, 10.3	5.6, 3.5, 1.6
3rd	12.4, 9.8, 6.6	24.1, 25.0, 25.7	27.0, 29.9, 34.4	23.6, 24.6, 25.6	13.0, 10.7, 7.7
4th	5.0, 3.1, 1.3	15.0, 13.2, 10.3	24.1, 25.1, 26.0	30.1, 33.1, 38.0	25.9, 25.5, 24.5
5th	1.2, 0.5, 0.1	5.5, 3.4, 1.5	13.1, 10.7, 7.7	26.1, 25.7, 24.6	54.1, 59.7, 66.1

Notes: This table compares the true quintile transition matrix with the estimates from DLLM and Gaussian copula-based synthetic panels using the known OLS error correlation (ρ_{Actual}) and a plausible lower and upper bound of possible correlations (ρ_L and ρ_H). Synthetic panel estimates for each quintile to quintile transition are for $\rho_L, \rho_{Actual}, \rho_H$. The ρ_{Actual} standard errors were generated by bootstrap with 100 replications.

True value lies within $[\rho_L, \rho_H]$ range (hatched cells)

True value lies within 95% confidence interval of ρ_{Actual} estimate (light grey cells)

True Value lies within 95% confidence of ρ_{Actual} and $[\rho_L, \rho_H]$ range (dark grey cells)

Table A1.3: Comparison of Quintile Transition Matrices, continued

NLSY79

True

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	43.2	24.8	16.0	10.6	5.3
2nd	22.8	24.2	22.9	16.8	13.3
3rd	13.8	21.6	24.2	21.6	18.9
4th	13.0	16.7	18.7	26.4	25.2
5th	7.1	12.9	18.2	24.4	37.5

DLLM [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	29.8, 34.6, 39.8	22.6, 23.6, 24.7	19.2, 18.4, 17.4	16.0, 14.3, 12.0	12.3, 9.1, 6.0
2nd	22.3, 23.6, 24.7	21.2, 22.3, 23.8	20.4, 20.5, 21.3	19.2, 18.9, 18.0	16.9, 14.8, 12.1
3rd	18.9, 18.3, 17.4	19.9, 20.6, 21.4	20.4, 21.3, 21.8	20.4, 20.6, 21.4	20.4, 19.2, 18.0
4th	16.1, 14.2, 11.8	19.0, 18.5, 18.0	20.1, 20.7, 21.3	21.6, 22.5, 23.9	23.1, 24.1, 25.1
5th	12.8, 9.4, 6.3	17.3, 15.0, 12.1	19.8, 19.1, 18.2	22.9, 23.7, 24.7	27.3, 32.8, 38.8

Copula [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	33.2, 38.5, 44.0	22.3, 23.1, 23.6	19.0, 17.6, 16.6	14.9, 12.9, 10.5	10.6, 7.8, 5.2
2nd	23.6, 24.3, 25.3	21.7, 22.9, 24.4	20.3, 21.0, 21.2	18.5, 18.3, 17.7	15.9, 13.5, 11.5
3rd	18.6, 17.8, 16.3	20.5, 20.8, 21.5	20.4, 20.9, 22.1	20.9, 21.4, 22.2	19.6, 19.1, 17.9
4th	14.4, 12.4, 9.9	19.0, 18.6, 18.4	20.4, 21.1, 21.7	22.4, 23.1, 24.6	23.8, 24.8, 25.4
5th	10.1, 6.9, 4.4	16.6, 14.6, 12.1	20.0, 19.4, 18.4	23.3, 24.3, 25.0	30.2, 34.9, 40.1

NLSY97

True

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	39.9	22.8	17.3	11.4	8.5
2nd	21.7	25.7	22.1	19.3	11.3
3rd	17.3	22.5	21.5	19.7	19.1
4th	11.5	16.5	24.3	23.9	23.6
5th	9.5	12.6	14.8	25.7	37.6

DLLM [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	30.0, 35.0, 40.9	22.5, 24.0, 24.9	18.9, 18.4, 17.0	16.1, 13.6, 11.5	12.4, 8.9, 5.7
2nd	22.6, 23.9, 24.9	21.4, 22.4, 24.3	20.3, 20.6, 21.4	18.9, 18.6, 17.7	16.8, 14.6, 11.7
3rd	18.6, 18.1, 17.2	20.1, 20.6, 21.4	20.4, 21.0, 22.3	20.6, 21.1, 21.3	20.3, 19.2, 17.8
4th	15.9, 13.9, 11.2	18.7, 18.2, 17.6	20.6, 20.9, 21.6	21.7, 22.6, 24.4	23.1, 24.4, 25.2
5th	12.8, 9.1, 5.8	17.2, 14.8, 11.9	19.9, 19.1, 17.6	22.7, 24.1, 25.2	27.3, 32.8, 39.5

Copula [$\rho_L = 0.63, \rho_{Actual} = 0.78, \rho_H = 0.88$]

Parent Quintile	Child Quintile 1st	2nd	3rd	4th	5th
1st	30.1, 35.5, 41.4	22.1, 23.3, 24.1	18.9, 18.2, 16.8	16.1, 13.8, 11.7	12.7, 9.1, 6.0
2nd	23.4, 24.2, 25.2	21.7, 22.7, 24.2	20.2, 20.5, 21.3	18.6, 18.4, 17.6	16.0, 14.2, 11.8
3rd	18.8, 18.0, 16.8	20.1, 20.7, 21.6	20.6, 21.4, 22.3	20.4, 20.9, 21.6	20.1, 19.0, 17.7
4th	15.6, 13.4, 11.0	18.9, 18.4, 18.1	20.6, 21.1, 21.8	21.7, 22.9, 24.2	23.2, 24.2, 25.0
5th	12.1, 8.8, 5.5	17.1, 14.8, 12.1	19.6, 18.8, 17.9	23.2, 24.0, 24.9	27.9, 33.4, 39.5

Notes: This table compares the true quintile transition matrix with the estimates from DLLM and Gaussian copula-based synthetic panels using the known OLS error correlation (ρ_{Actual}) and a plausible lower and upper bound of possible correlations (ρ_L and ρ_H). Synthetic panel estimates for each quintile to quintile transition are for $\rho_L, \rho_{Actual}, \rho_H$. The ρ_{Actual} standard errors were generated by bootstrap with 100 replications.

True value lies within [ρ_L, ρ_H] range

True value lies within 95% confidence interval of ρ_{Actual} estimate

True Value lies within 95% confidence of ρ_{Actual} and [ρ_L, ρ_H] range

Table A1.4: Rank-Rank Slope

PSID-CNEF (ρ_{Actual} = 0.78)				CPS ASEC (ρ_{Actual} = 0.68)			
ρ	True	DLLM	Copula	ρ	True	DLLM	Copula
0.63	0.820	0.737	0.678	0.58	0.720	0.655	0.644
0.68		0.770	0.712	0.63		0.694	0.682
0.73		0.805	0.745	**0.68**		**0.734**	**0.721**
0.78		**0.841**	**0.783**	0.73		0.774	0.761
0.83		0.875	0.823	0.78		0.815	0.802
0.88		0.912	0.865				

NLSY79 (ρ_{Actual} = 0.24)				NLSY97 (ρ_{Actual} = 0.24)			
ρ	True	DLLM	Copula	ρ	True	DLLM	Copula
0.09	0.418	0.218	0.288	0.09	0.403	0.222	0.243
0.14		0.248	0.321	0.14		0.258	0.279
0.19		0.282	0.352	0.19		0.294	0.314
0.24		**0.316**	**0.386**	**0.24**		**0.327**	**0.350**
0.29		0.349	0.416	0.29		0.363	0.382
0.34		0.382	0.449	0.34		0.399	0.418
0.39		0.416	0.484	0.39		0.436	0.454

ENIGH84-EMOVI06				ENIGH89-EMOVI11			
ρ	True	DLLM	Copula	ρ	True	DLLM	Copula
0.05	N/A	0.237	0.321	0.05	N/A	0.233	0.316
0.10		0.265	0.343	0.10		0.260	0.344
0.12		**0.272**	**0.354**	0.12		0.272	0.349
0.15		0.295	0.372	0.15		0.287	0.363
0.18		0.313	0.386	**0.18**		**0.304**	**0.376**
0.20		0.326	0.394	0.20		0.318	0.394
0.25		0.348	0.418	0.25		0.349	0.416
0.30		0.383	0.446	0.30		0.383	0.440

True value lies within 95% confidence interval

No statistically significant difference between true value and estimate

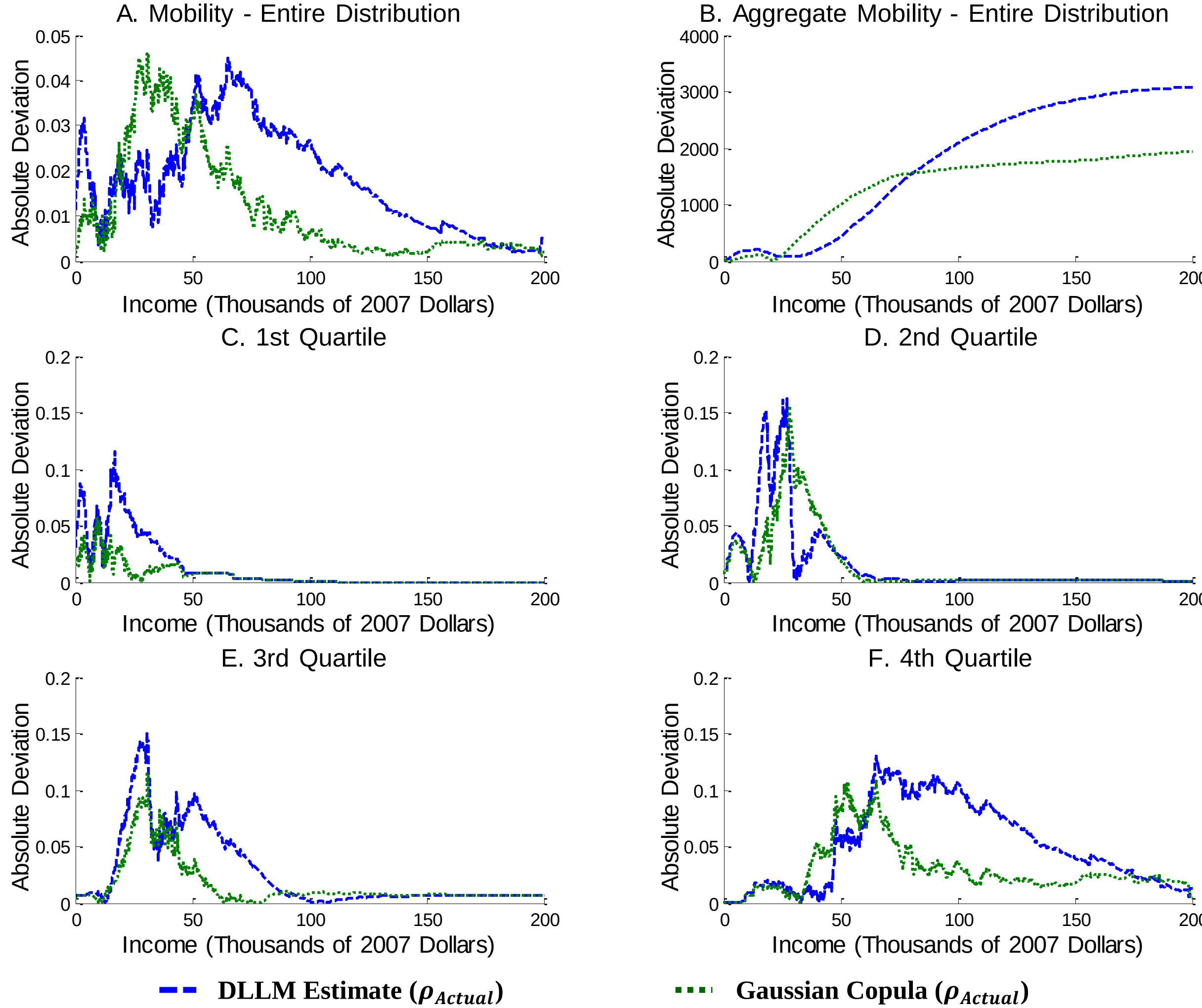
True Value lies within 95% confidence of ρ_{Actual} and not statistically significantly different from estimate

Notes: The rank-rank slope is the coefficient from the regression of child income rank on parent income rank. This table compares the true and synthetic panel estimate of the rank-rank slope for each data set for a variety of ρ values including the actual error correlation when known and the estimated correlation for the ENIGH84-EMOVI06 (born 1966-1976) and ENIGH89-EMOVI11 (born 1971-1981).

Appendix 2 (Online Only). Comparison of True and Synthetic Mobility Curve Estimates

Figure A2.1: Absolute Deviation of Synthetic Estimate (ρ_{Actual}) from True Mobility in the PSID-CNEF (2005-2007)

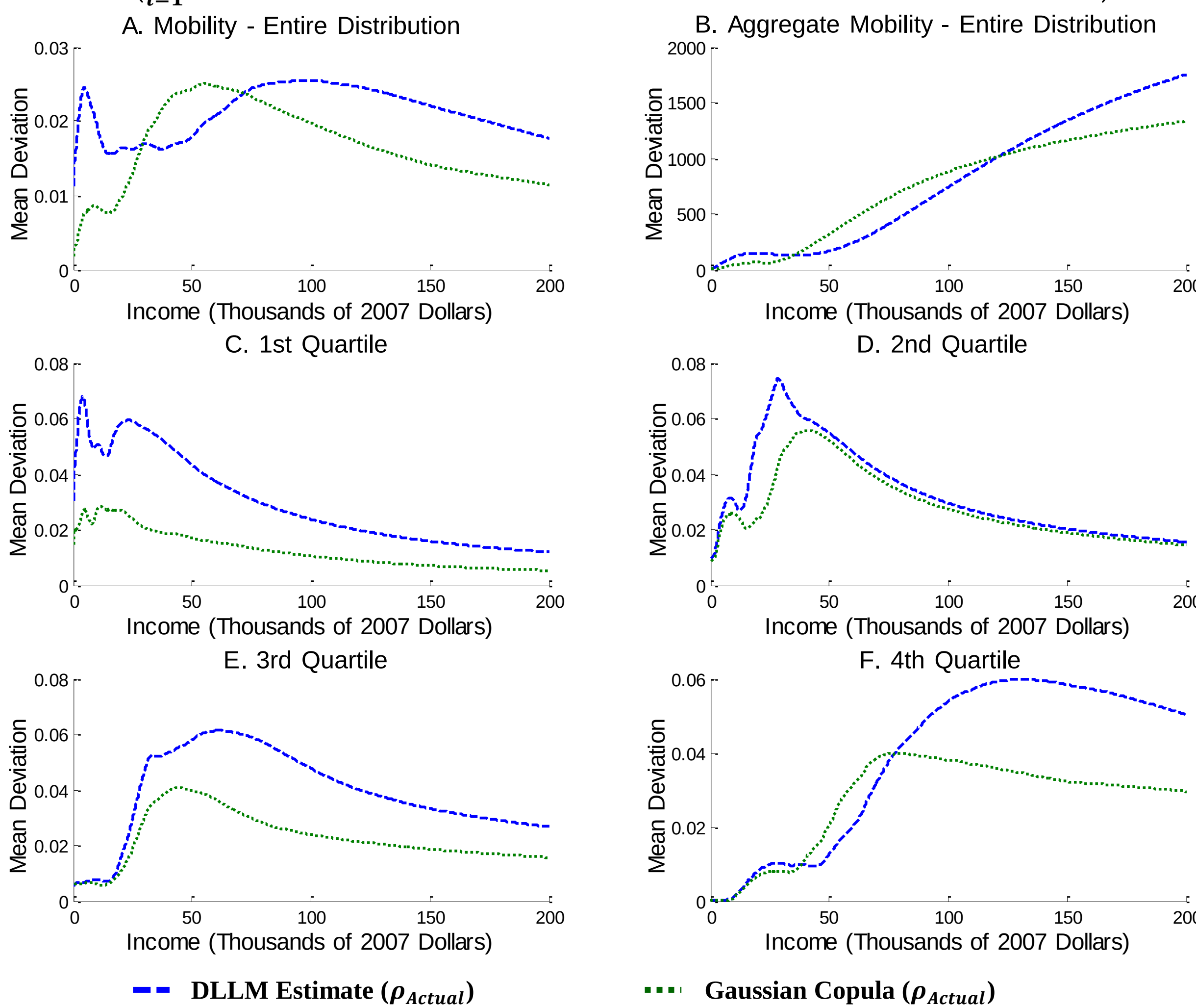
$$\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right| + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|$$



Notes: This table compares the absolute deviation from the true sample mobility curve. At each cutoff, the absolute deviation is the absolute value of the difference between each synthetic panel estimate and the true curve summed for both upward and downward mobility, so that the absolute deviation = $\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right| + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.2: Mean Absolute Deviation up to Each Cutoff of Synthetic Estimate (ρ_{Actual}) from True Mobility in the PSID-CNEF (2005-2007)

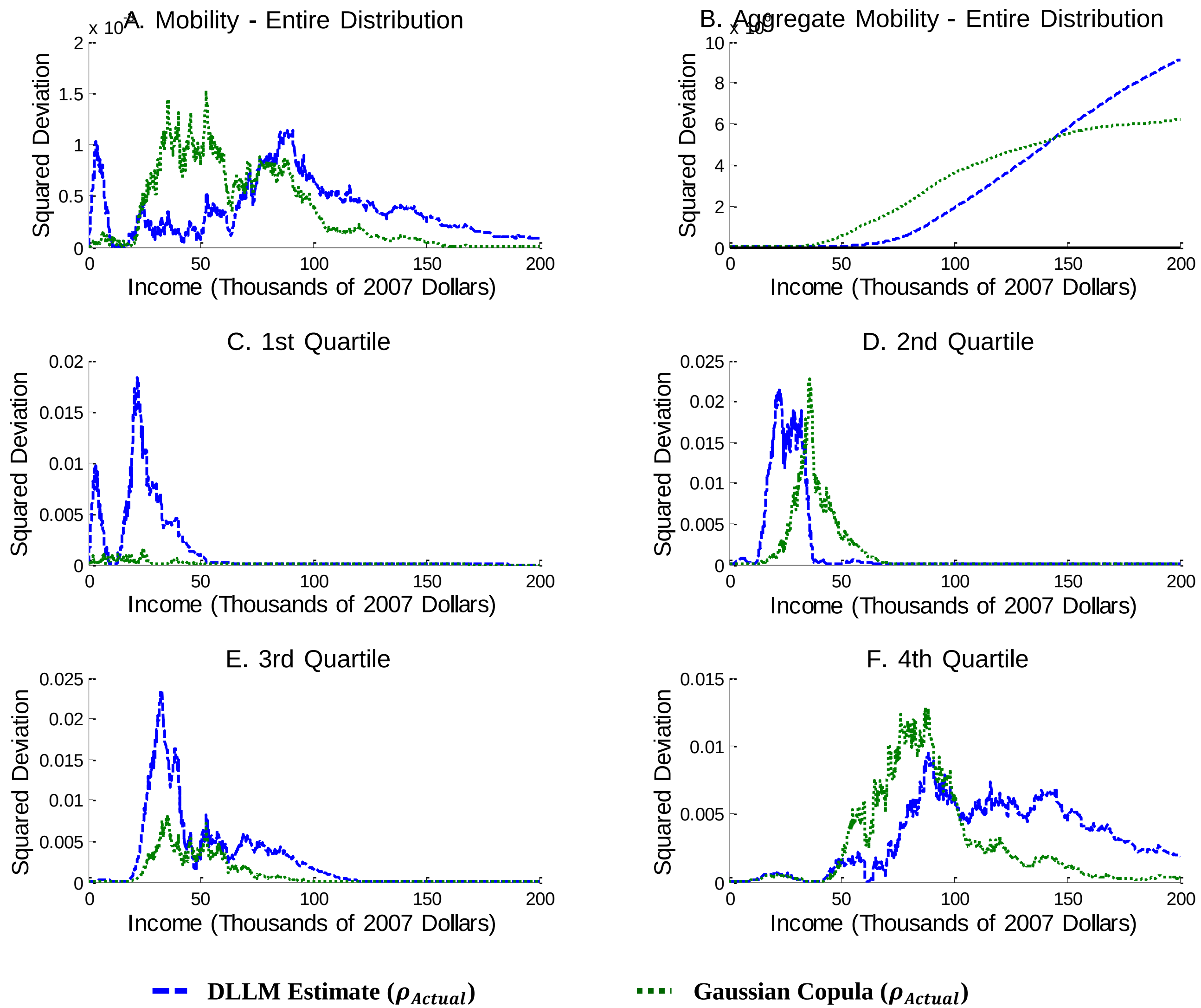
$$\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|\right)$$



Notes: This table compares the mean deviation from the true sample mobility curve. At each cutoff, the mean deviation is the average absolute deviation up to the cutoff, so that for $c_i \in [0, c]$ approximated by a discrete number of cutoffs n_c, the mean deviation $= \frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|\right)$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.3: Squared Deviation of Synthetic Estimate (ρ_{Actual}) from True Mobility in the PSID-CNEF (2005-2007)

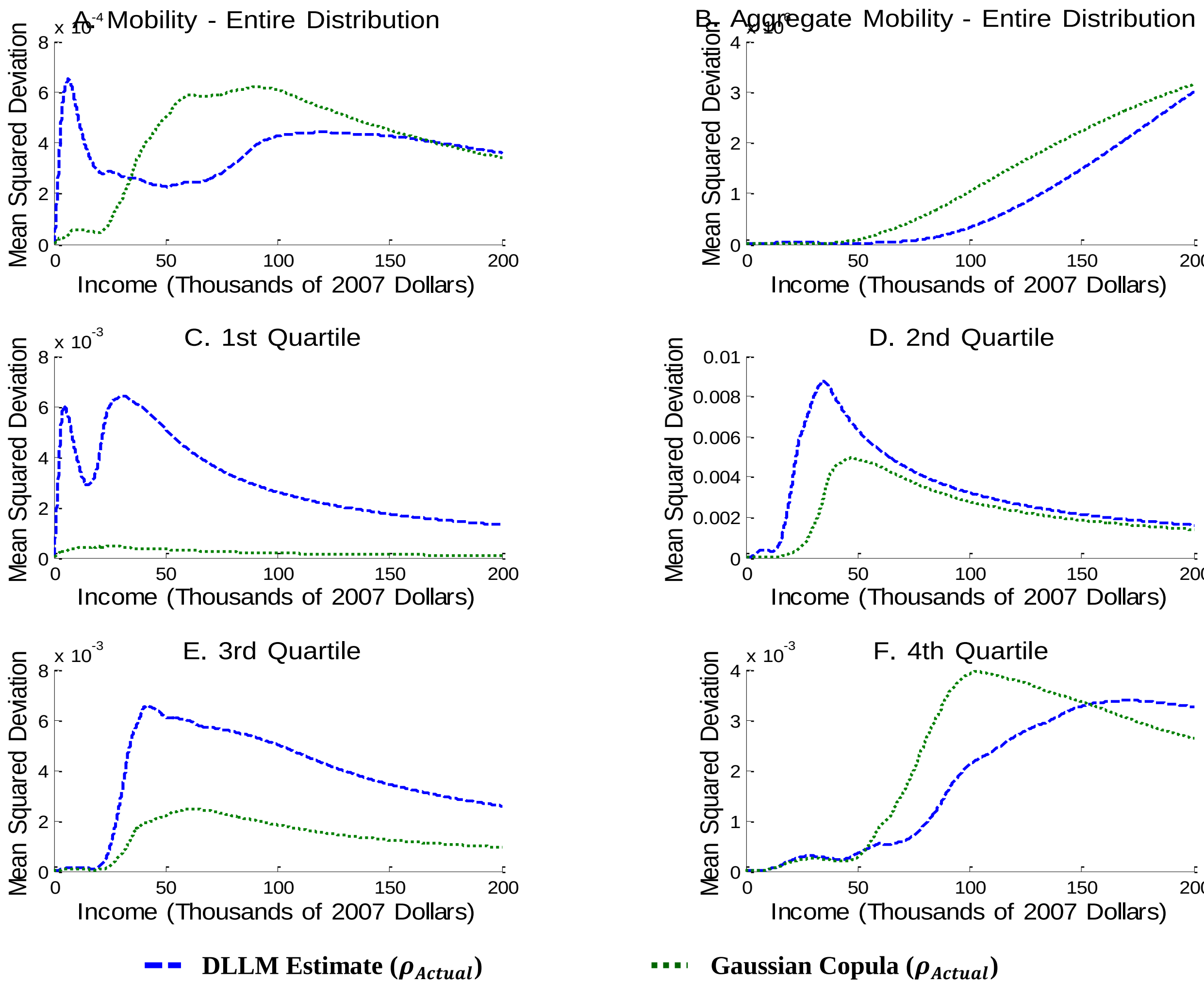
$$\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^2 + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^2$$



Notes: This table compares the squared deviation from the true sample mobility curve. At each cutoff, the squared deviation is the absolute value of the squared difference between each synthetic panel estimate and the true curve summed for both upward and downward mobility, so that the squared deviation = $\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^2 + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^2$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.4: Mean Squared Deviation up to Each Cutoff of Synthetic Estimate (ρ_{Actual}) from True Mobility in the PSID-CNEF (2005-2007)

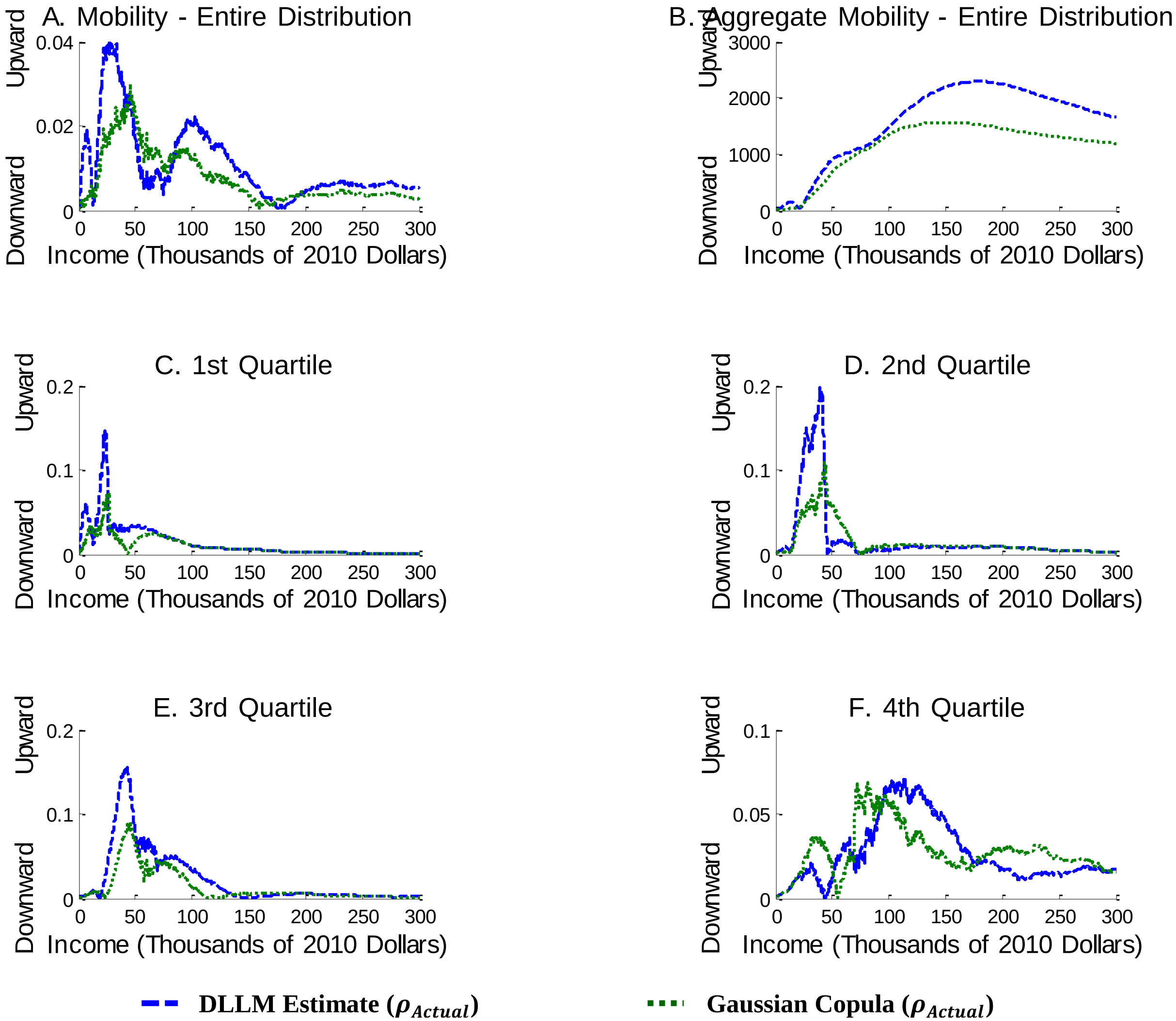
$$\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|^2+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|^2\right)$$



Notes: This table compares the mean squared deviation from the true sample mobility curve. At each cutoff, the mean squared deviation is the average squared deviation up to the cutoff, so that for $c_i \in [0, c]$ approximated by a discrete number of cutoffs n_c, the mean squared deviation $= \frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|^2+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|^2\right)$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.5: Absolute Deviation of Synthetic Estimate (ρ_{Actual}) from True Mobility in the CPS ASEC (2005-2006)

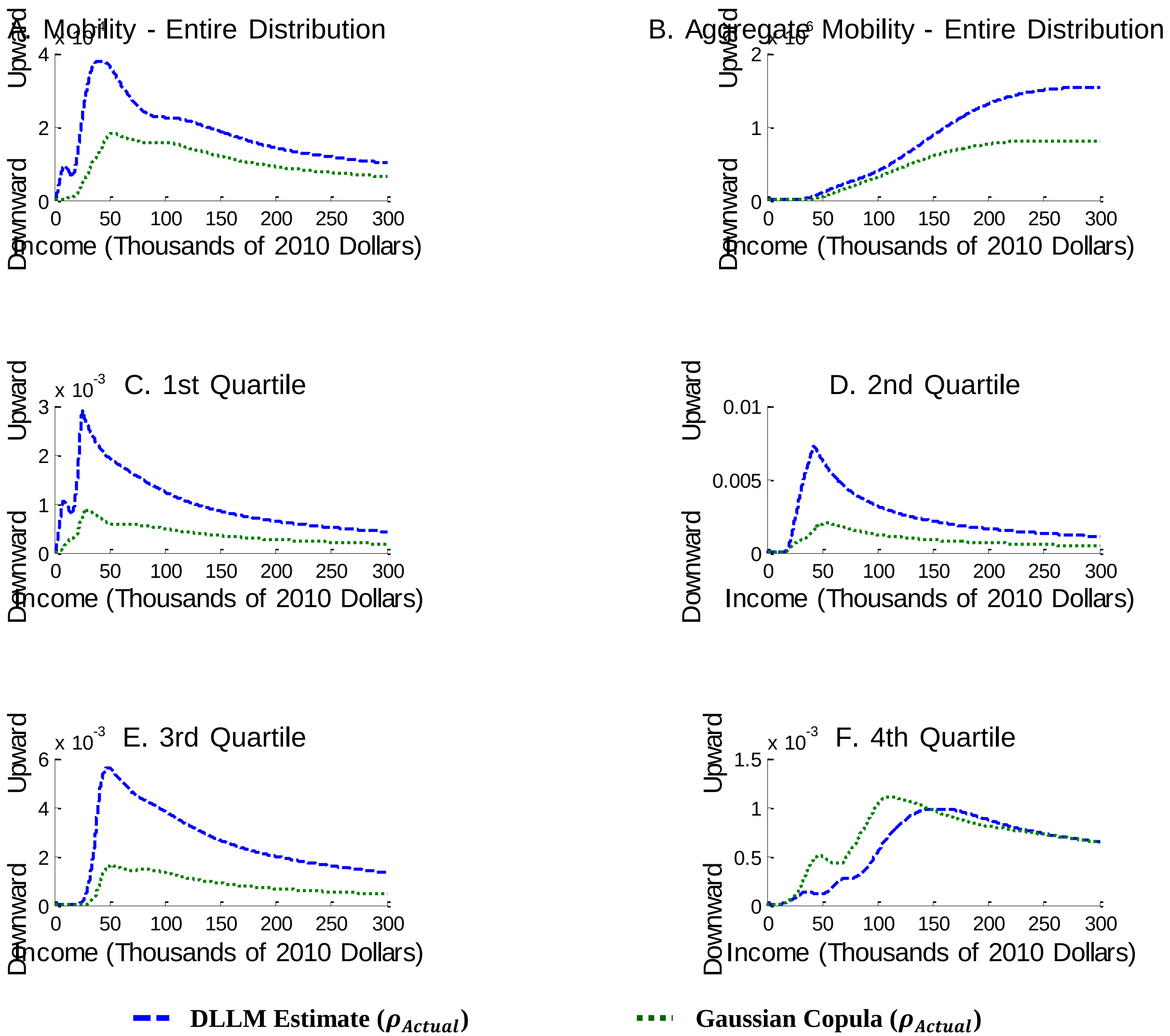
$$\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right| + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|$$



Notes: This table compares the absolute deviation from the true sample mobility curve. At each cutoff, the absolute deviation is the absolute value of the difference between each synthetic panel estimate and the true curve summed for both upward and downward mobility, so that the absolute deviation = $\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right| + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.6: Mean Absolute Deviation up to Each Cutoff of Synthetic Estimate (ρ_{Actual}) from True Mobility in the CPS ASEC (2005-2006)

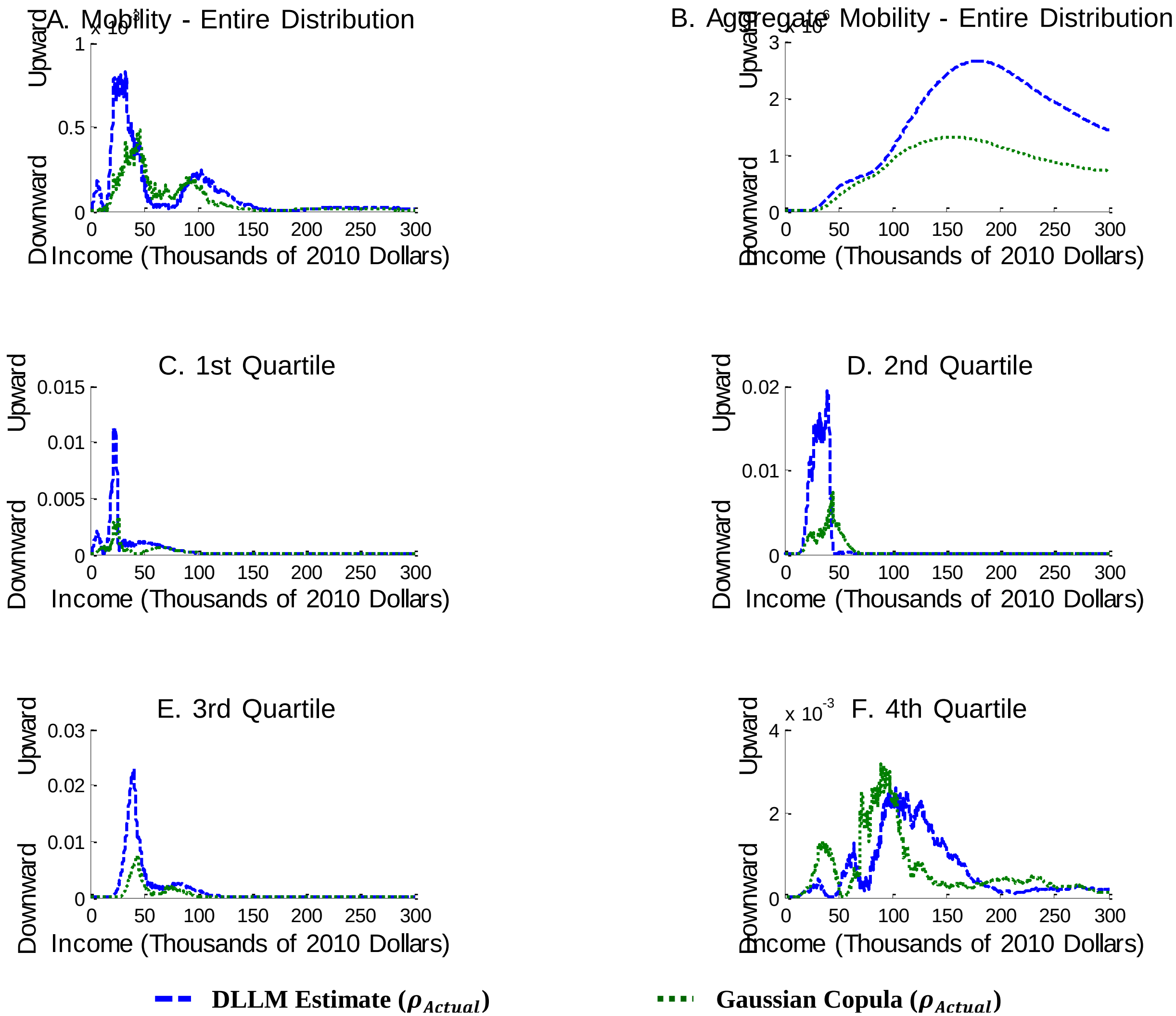
$$\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c)-m_U^{Synthetic}(y,c)\right|+\left|m_D^{True}(y,c)-m_D^{Synthetic}(y,c)\right|\right)$$



Notes: This table compares the mean deviation from the true sample mobility curve. At each cutoff, the mean deviation is the average absolute deviation up to the cutoff, so that for $c_i \in [0, c]$ approximated by a discrete number of cutoffs n_c, the mean deviation $= \frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|\right)$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.7: Squared Deviation of Synthetic Estimate (ρ_{Actual}) from True Mobility in the CPS ASEC (2005-2006)

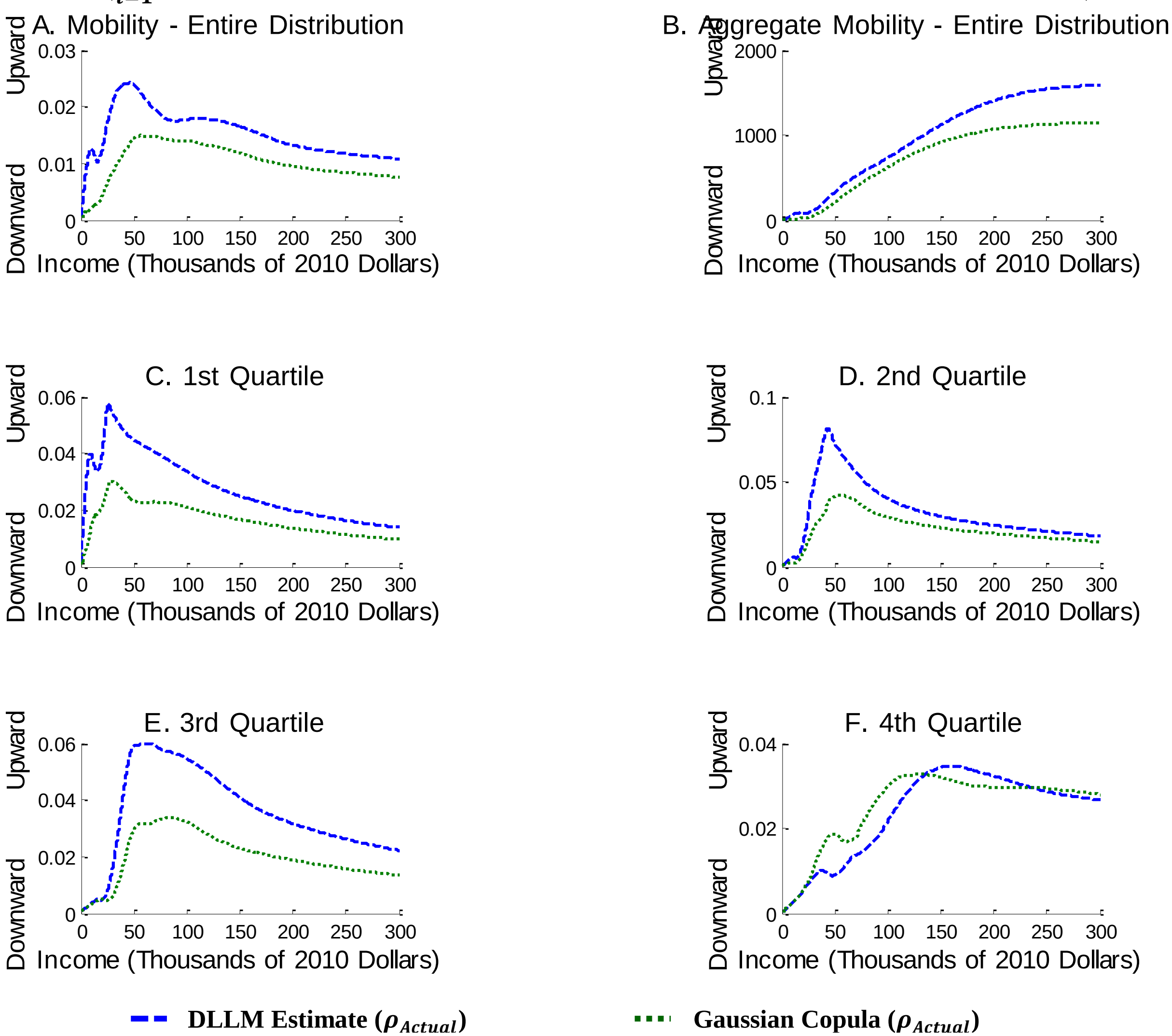
$$\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^2 + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^2$$



Notes: This table compares the squared deviation from the true sample mobility curve. At each cutoff, the squared deviation is the absolute value of the squared difference between each synthetic panel estimate and the true curve summed for both upward and downward mobility, so that the squared deviation = $\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^2 + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^2$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.8: Mean Squared Deviation up to Each Cutoff of Synthetic Estimate (ρ_{Actual}) from True Mobility in the CPS ASEC (2005-2006)

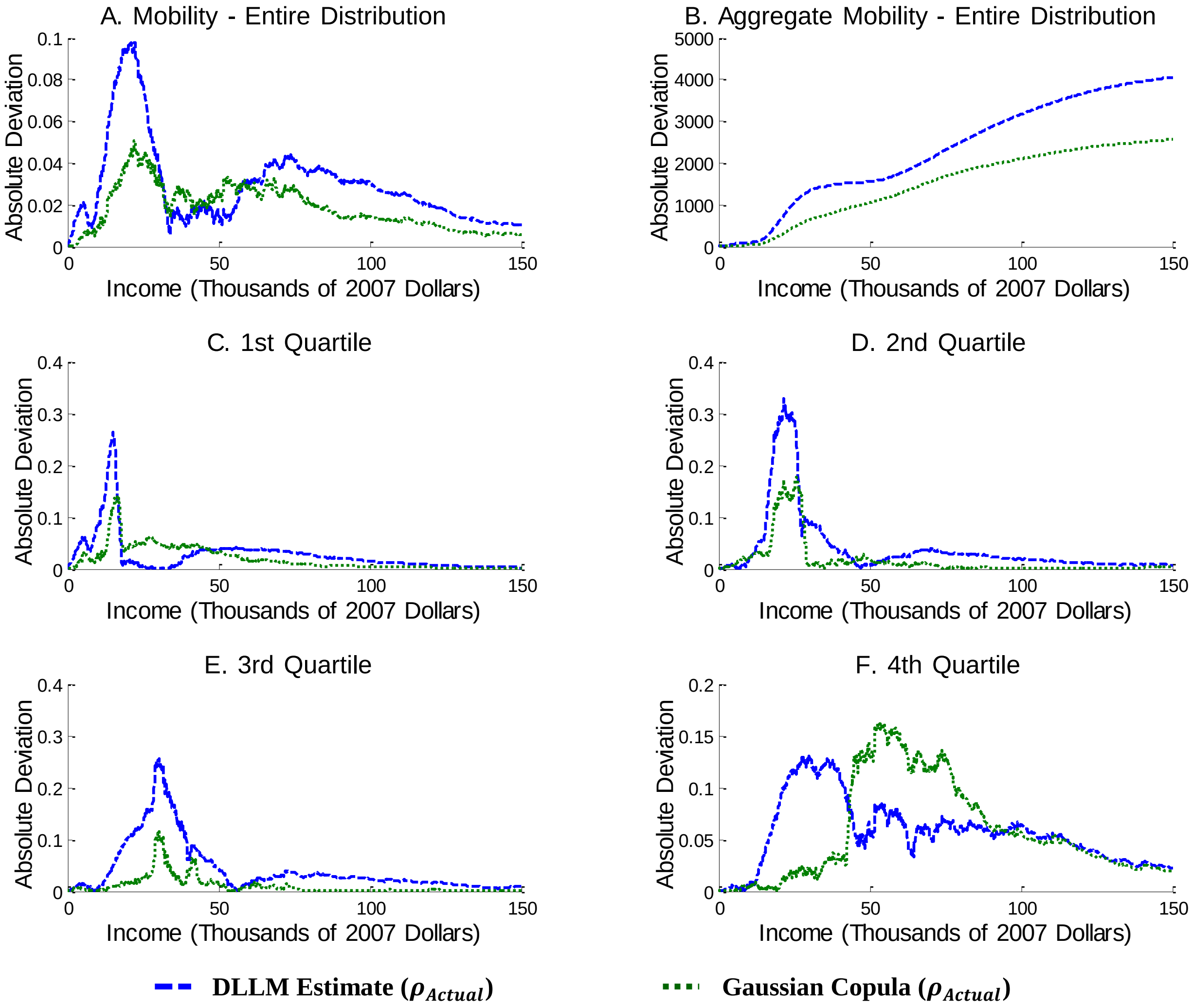
$$\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|^2+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|^2\right)$$



Notes: This table compares the mean squared deviation from the true sample mobility curve. At each cutoff, the mean squared deviation is the average squared deviation up to the cutoff, so that for $c_i \in [0, c]$ approximated by a discrete number of cutoffs n_c, the mean squared deviation $= \frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|^2+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|^2\right)$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.9: Absolute Deviation of Synthetic Estimate (ρ_{Actual}) from True Mobility in the NLSY79

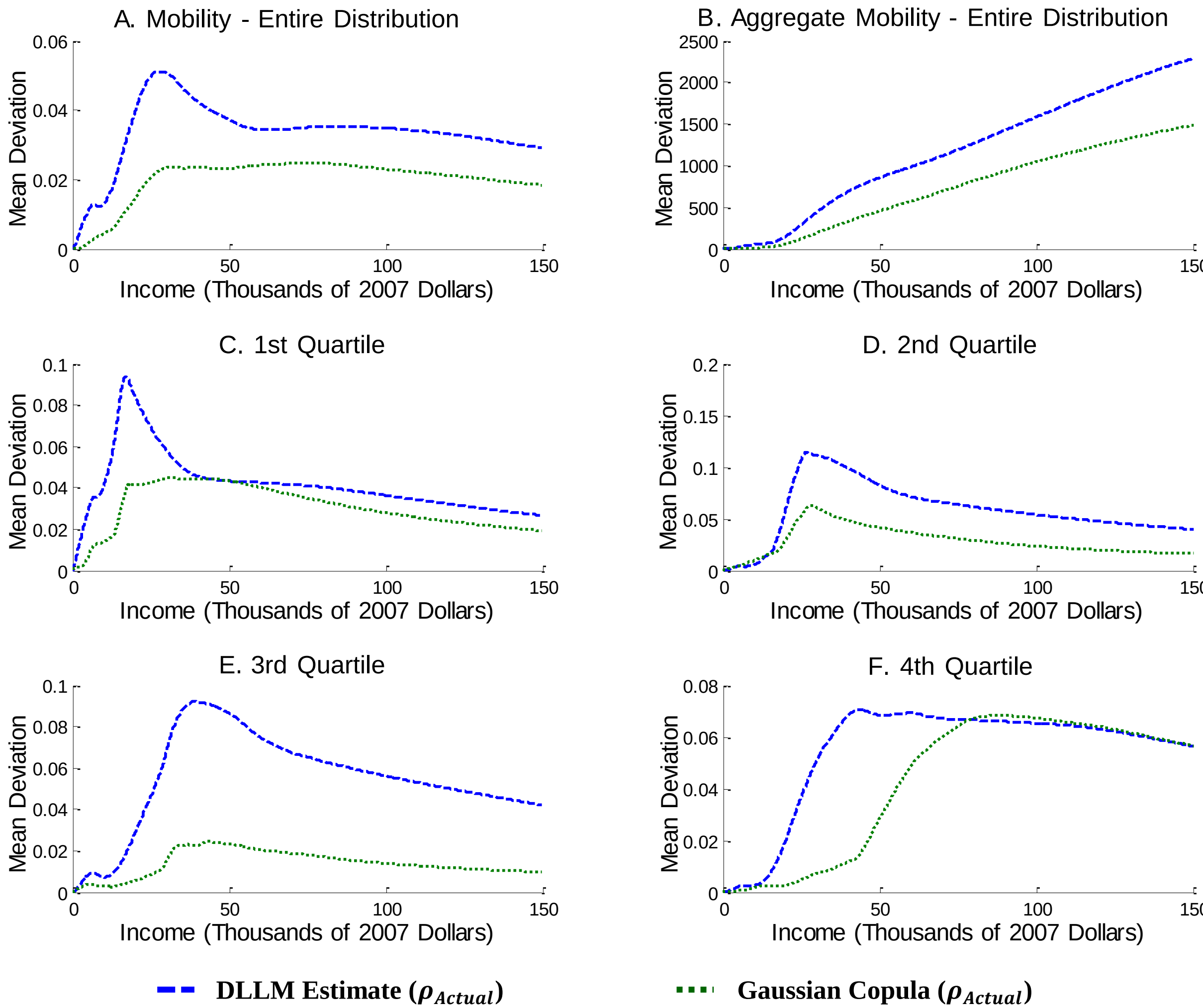
$$\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right| + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|$$



Notes: This table compares the absolute deviation from the true sample mobility curve. At each cutoff, the absolute deviation is the absolute value of the difference between each synthetic panel estimate and the true curve summed for both upward and downward mobility, so that the absolute deviation = $\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right| + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.10: Mean Absolute Deviation up to Each Cutoff of Synthetic Estimate (ρ_{Actual}) from True Mobility in the NLSY79

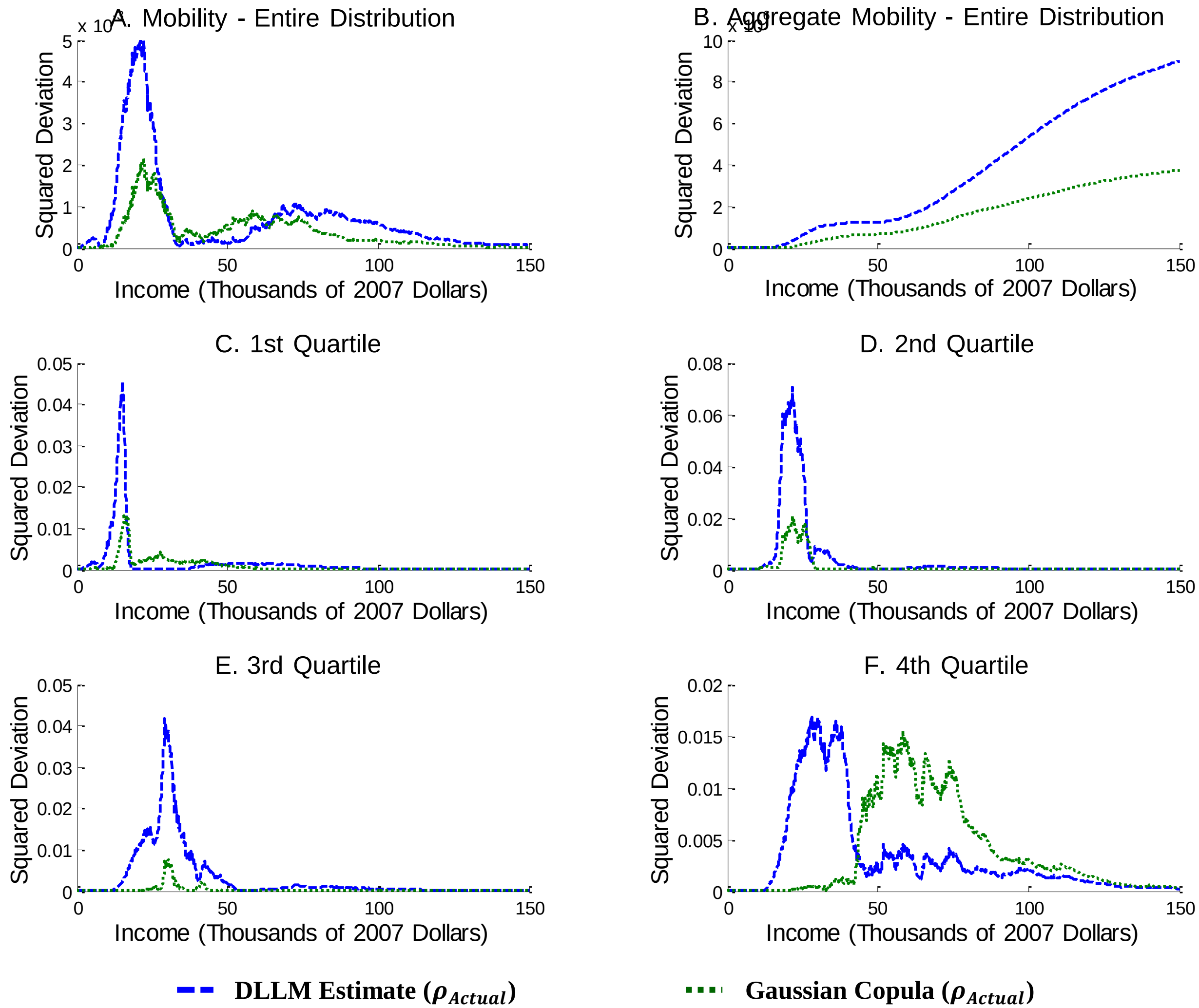
$$\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c)-m_U^{Synthetic}(y,c)\right|+\left|m_D^{True}(y,c)-m_D^{Synthetic}(y,c)\right|\right)$$



Notes: This table compares the mean deviation from the true sample mobility curve. At each cutoff, the mean deviation is the average absolute deviation up to the cutoff, so that for $c_i \in [0, c]$ approximated by a discrete number of cutoffs n_c, the mean deviation $= \frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|\right)$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.11: Squared Deviation of Synthetic Estimate (ρ_{Actual}) from True Mobility in the NLSY79

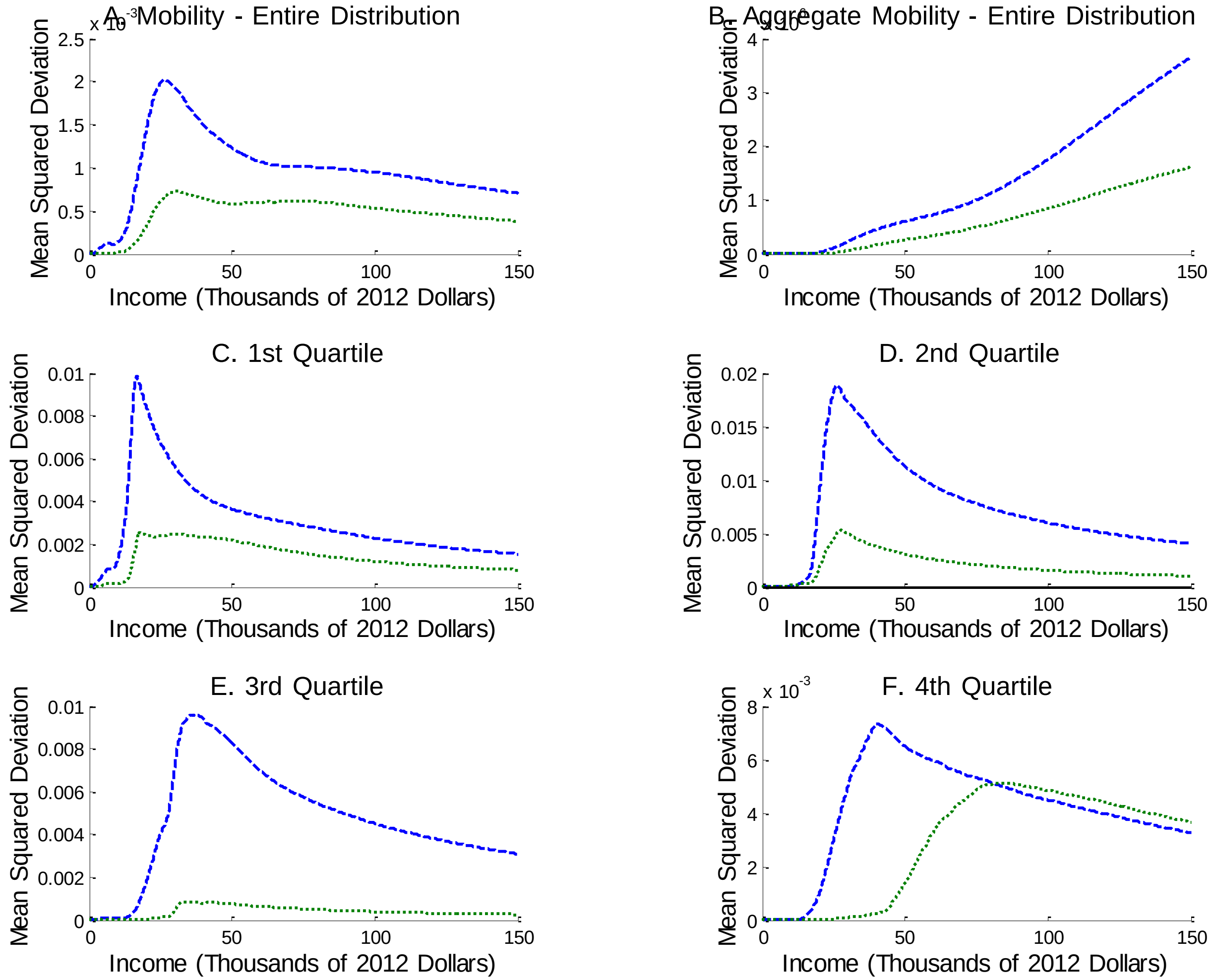
$$\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^2 + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^2$$



Notes: This table compares the squared deviation from the true sample mobility curve. At each cutoff, the squared deviation is the absolute value of the squared difference between each synthetic panel estimate and the true curve summed for both upward and downward mobility, so that the squared deviation = $\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^2 + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^2$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.12: Mean Squared Deviation up to Each Cutoff of Synthetic Estimate (ρ_{Actual}) from True Mobility in the NLSY79

$$\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|^2+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|^2\right)$$

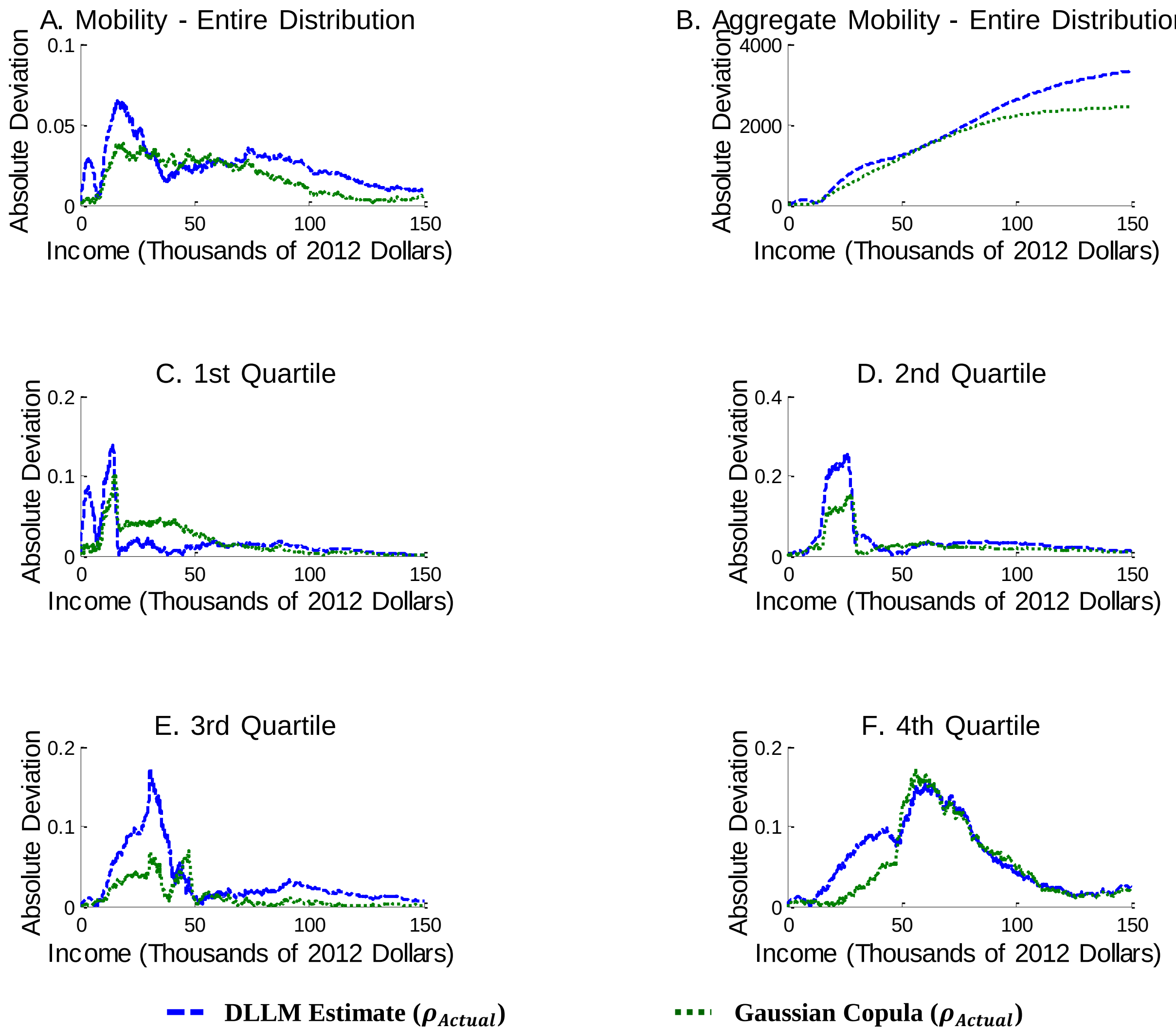


DLLM Estimate (ρ_{Actual}) **Gaussian Copula (ρ_{Actual})**

Notes: This table compares the mean squared deviation from the true sample mobility curve. At each cutoff, the mean squared deviation is the average squared deviation up to the cutoff, so that for $c_i \in [0, c]$ approximated by a discrete number of cutoffs n_c, the mean squared deviation $= \frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|^2+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|^2\right)$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.13: Absolute Deviation of Synthetic Estimate (ρ_{Actual}) from True Mobility in the NLSY97

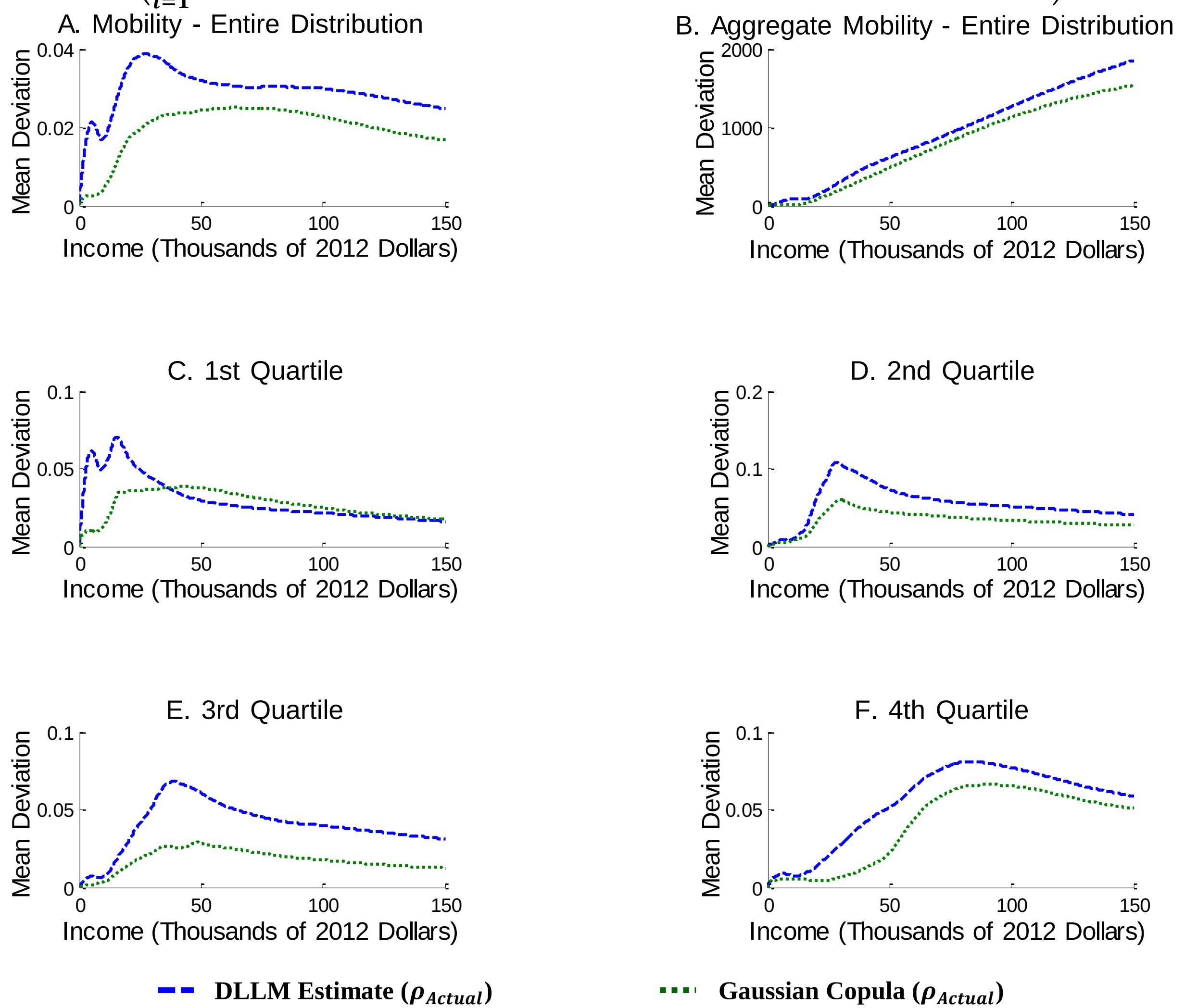
$$\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right| + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|$$



Notes: This table compares the absolute deviation from the true sample mobility curve. At each cutoff, the absolute deviation is the absolute value of the difference between each synthetic panel estimate and the true curve summed for both upward and downward mobility, so that the absolute deviation $= \left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right| + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.14: Mean Absolute Deviation up to Each Cutoff of Synthetic Estimate (ρ_{Actual}) from True Mobility in the NLSY97

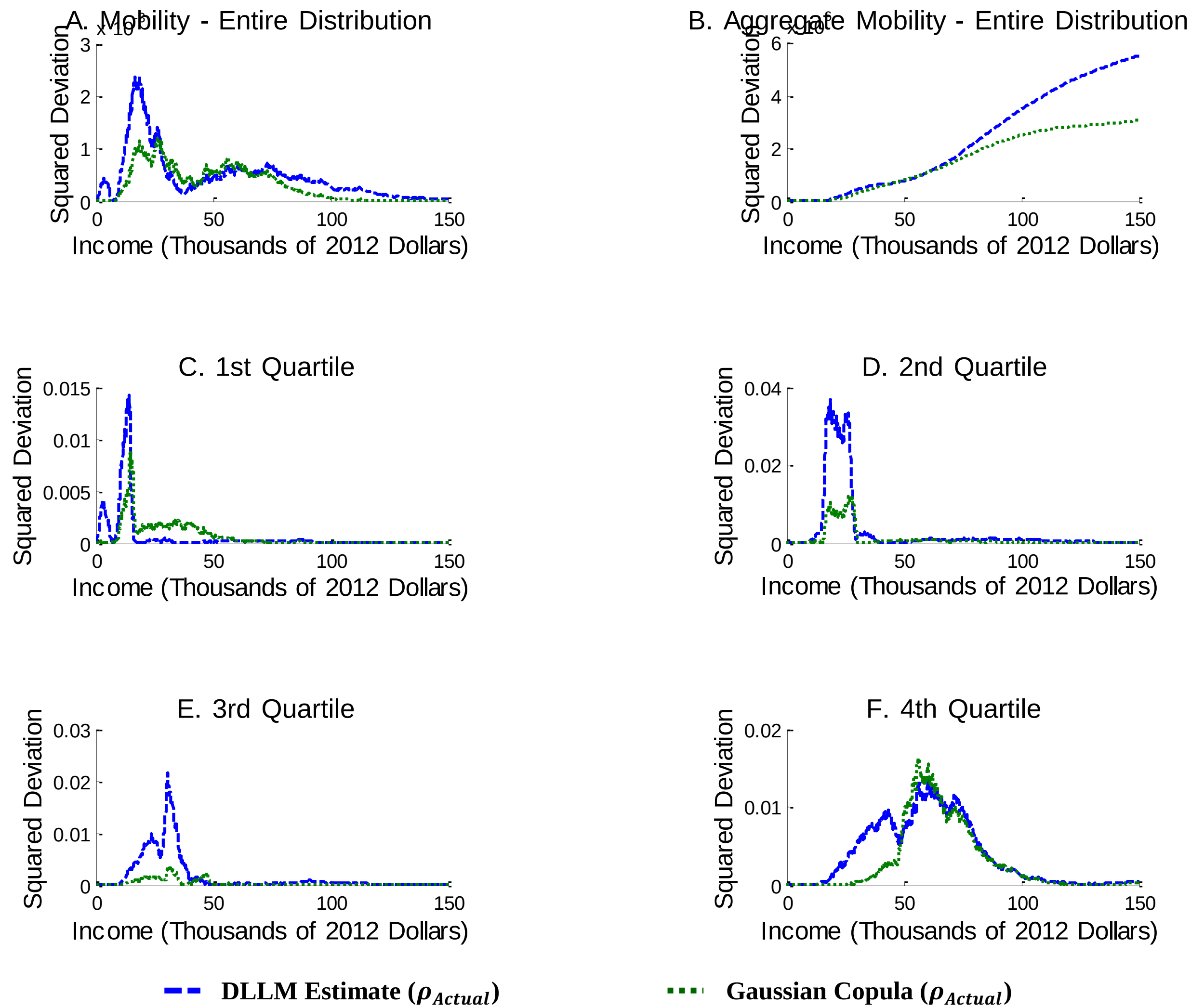
$$\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c)-m_U^{Synthetic}(y,c)\right|+\left|m_D^{True}(y,c)-m_D^{Synthetic}(y,c)\right|\right)$$



Notes: This table compares the mean deviation from the true sample mobility curve. At each cutoff, the mean deviation is the average absolute deviation up to the cutoff, so that for $c_i \in [0, c]$ approximated by a discrete number of cutoffs n_c, the mean deviation $= \frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|\right)$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.15: Squared Deviation of Synthetic Estimate (ρ_{Actual}) from True Mobility in the NLSY97

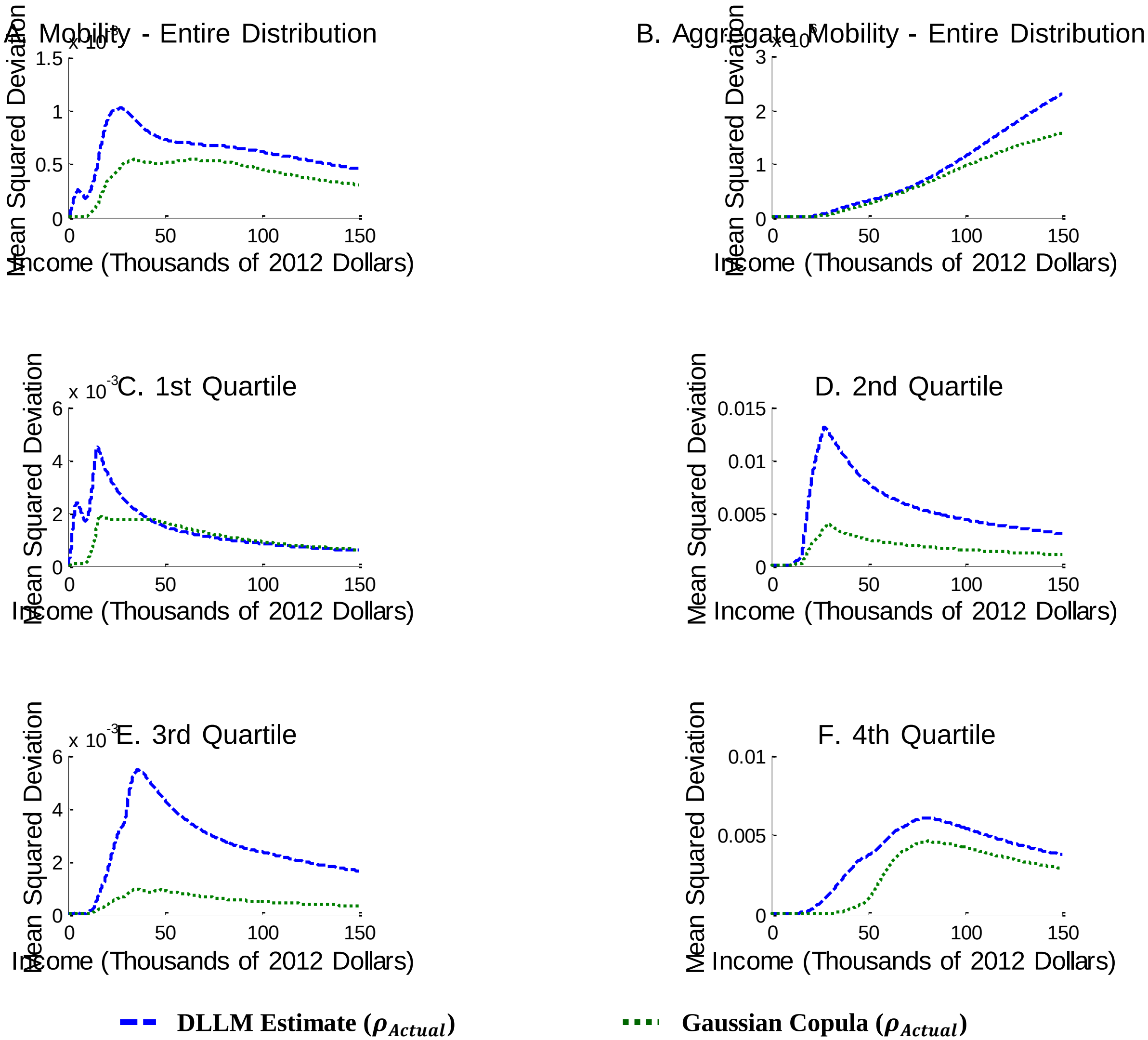
$$\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^2 + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^2$$



Notes: This table compares the squared deviation from the true sample mobility curve. At each cutoff, the squared deviation is the absolute value of the squared difference between each synthetic panel estimate and the true curve summed for both upward and downward mobility, so that the squared deviation = $\left|m_U^{True}(y,c) - m_U^{Synthetic}(y,c)\right|^2 + \left|m_D^{True}(y,c) - m_D^{Synthetic}(y,c)\right|^2$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).

Figure A2.16: Mean Squared Deviation up to Each Cutoff of Synthetic Estimate (ρ_{Actual}) from True Mobility in the NLSY97

$$\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|^2+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|^2\right)$$



Notes: This table compares the mean squared deviation from the true sample mobility curve. At each cutoff, the mean squared deviation is the average squared deviation up to the cutoff, so that for $c_i \in [0,c]$ approximated by a discrete number of cutoffs n_c, the mean squared deviation $=\frac{1}{n_c}\left(\sum_{i=1}^{nc}\left|m_U^{True}(y,c_i)-m_U^{Synthetic}(y,c_i)\right|^2+\left|m_D^{True}(y,c_i)-m_D^{Synthetic}(y,c_i)\right|^2\right)$. The DLLM and Gaussian copula-based synthetic panels in the comparison use the known OLS error correlation (ρ_{Actual}).