A Copula-Based Framework for Studying Intergenerational Income Mobility

Anning Hu

Department of Sociology

Graduate School

Fudan University

Zhipeng Zhou

Department of Sociology

Cornell University

Abstract:

The sociological analysis of the mobility tables enhances the examination of the circulation mobility and helps one reveal the nuanced morphological patterns of mobility. In contrast, the economic analysis based on the measure of elasticity provides a handy way of covariate conditioning and statistically testing the similarities of mobility patterns across groups. In this article, we argue that the distinct methodological merits of these two approaches can be equipped by adopting a more comprehensive analytical framework using the copula functions: (1) The copula functions concern the dependence structure that is independent from the margins, which enable scholars to focus on the relative mobility; (2) The copula density, estimated either parametrically or non-parametrically, reveals the nuanced morphological mobility patterns; (3) By residualizing the marginal variables, the detected mobility pattern can be interpreted in a stronger causal sense; and (4) the Cramér–von Mises Test offers an easy-to-use statistic to conduct intergroup comparison of mobility patterns. The copula-based framework is illustrated by investigating the income mobility between 1978 and 2017 in the U.S., using the National Longitudinal Survey of Youth 1979 (NLSY79).

Keywords: Intergenerational Income Mobility; Mobility Tables; Intergenerational Income Elasticity; Copulas

- All correspondence to Anning Hu, 1118, Liberal Arts Building, Fudan University, 220, Handan Road, Shanghai, China. 200433. Email: huanning@fudan.edu.cn.

INTRODUCTION

Intergenerational mobility, as stated by Grusky and Weeden (2006: 85), is “one of the signature contributions of sociology over the last half-century.” Using sophisticated methods of mobility tables analysis, sociologists succeed in revealing various typological patterns of circulation mobility between parents and children. As such, the sociological studies of intergenerational mobility have always been serving as a major intellectual brainpower for understanding the extent of opportunity equality in a society (e.g., Grusky, Smeeding, and Snipp 2015; Hout 2015). This strand of research, following Ritzer’s definition, justifiably constitutes a unique “paradigm” (Ritzer 1975). However, the popularity of the sociological paradigm has been observed to dwindle in the 21st century (Grusky and Weeden 2006; Sakamoto and Wang 2020). With the increasing availability of administrative data on income, the economic framework based on the intergenerational income elasticity (IGE) has now been more frequently used by the sociologists (e.g., Beller and Hout 2006; Bloome 2014, 2017; Fox, Torche, and Waldfogel 2016). One consequence of this shift is the “balkanization” of the mobility research (Grusky and Cumberworth 2010) where the economic approach and sociological approach are increasingly seen to stand for distinctive enterprises.

This state of balkanization is understandable. The economic approach estimates the IGE to gauge the “global” status of intergenerational income mobility, and scholars adopting this approach are relatively more interested in the transition and social determinants of the IGE that can be handily examined in a conventional ordinary least squares (OLS) model (e.g., Chetty et al. 2014). In contrast, the occupation-based tabular studies lean more toward the distinction between the structural and circulation mobility through the log-linear model

(Christensen 1994), and the major theoretical contributions usually surround the various mobility typologies (e.g., quasi-independence, uniform association, and quasi-symmetry) established based on the “cell frequencies” of a mobility table (Hout 1983). Due to these different orientations, it goes as no surprise that occupational mobility and income mobility appear to belong to distinct schools of research, and consequently, some shortcomings inevitably emerge: the traditional academic questions pertaining to occupational mobility, such as the patterns of the circulation mobility, are rarely asked by income-mobility scholars, and meanwhile, the analytical convenience of the IGE analysis using the OLS model, such as the comparison of mobility status across groups and the conditioning of potentially confounding covariates, seem to be sociologically impertinent.

In this article, we argue that these shortcomings of the paradigmatic balkanization can be overcome by deploying a more comprehensive and inclusive framework that draws on the desirable mathematical properties of the copula functions. Specifically, this framework enables sociologists to (1) focus on the circulation income mobility that is net of the margins; (2) depict the nuanced morphological patterns of income mobility; (3) control for the covariates; and (4) statistically test the similarities of mobility patterns between groups of individuals. In the following section, we will summarize the major research questions that sociologists are interested in. Subsequently, we present the detailed discussions to address how the copula functions can help to answer these questions and methodologically bridging the sociological and economic approaches. The copula-based framework is then illustrated by examining the income mobility between 1978 and 2017 in the U.S., using the National Longitudinal Survey of Youth 1979 (NLSY79), which is followed by the concluding remarks.

RESEARCH QUESTIONS OF INTERGENERATIONAL MOBILITY

Several research questions that are of interest to sociologists surround intergenerational mobility. In this section, we will check these questions, and see how the mobility tables analysis and the IGE analysis could be used to partially address them. In the subsequent section, we will shift attention to the copula functions, and articulate in detail how the copula functions can better serve to investigate these research questions. Readers should note that the research questions discussed here are generally relevant for both occupation and income mobility.

Distinction between the Structural and Circulation Mobility

The distinction between the structural mobility and the circulation (or relative) mobility has been a long tradition in the sociological literature (Featherman and Hauser 1978; Torche 2015). It has been argued that the structural mobility is driven by the grand socio-political changes (i.e., industrialization) (e.g., Bearman and Deane 1992), so sociologists relatively pay more attention to the circulation mobility that better informs us the extent of social fluidity, i.e., children born in different classes of families have differential access to life chances (Breen and Jonsson 2007). The measure of circulation mobility that is widely adopted by the sociologists is the odds ratio, which is believed to be margin invariant insofar as to be distinguished from the structural mobility (Hout 1983).¹ However, readers should

¹ Take occupational mobility for an example, parents' and children's occupations are cross tabulated to constitute the mobility table. For parents' occupational category i and children's occupational category j , the expected frequency of observations F_{ij} is parameterized as $F_{ij} = \mu \tau_i \tau_j \tau_{ij}$. In this formula, μ refers to the sample size ($\log(\mu)$ is then taken to be the grand mean in the log-linear

note that the property of marginal invariance for the odds ratio does not go without critique (e.g., Sobel 1983; Harrison 1988), but more detailed clarifications on this theme go beyond the scope of this article.

Compared with the mobility tables analysis, the economist framework mostly uses the analytical tool of the IGE. This construct pays much less attention to the structural-versus-circulation distinction.² Specifically, the IGE is routinely estimated by the coefficient β in the OLS model of $\log(\text{children's income}) = \alpha + \beta \log(\text{parents' income}) + \varepsilon$, where α is the intercept and ε is the random error. According to Chetty and colleagues (2014), this estimate is tantamount to $\rho \frac{\text{sd}[\log(\text{parents' income})]}{\text{sd}[\log(\text{children's income})]}$, where ρ is a linear correlation coefficient between $\log(\text{children's income})$ and $\log(\text{parents' income})$ and sd stands for standard deviation. Clearly, if the marginal income distributions of the two generations vary in terms of their extent of dispersion, the elasticity would change and deviate from ρ . To wrap up, the IGE does not make the structural-versus-circulation distinction in the sociological sense.³

Recognition of Mobility Patterns

Another question that has always been on the research agenda of sociologists is the pattern of mobility. By pattern, we mean the morphological characteristics that are embodied by the

model), τ_i is the row effect, τ_j is the column effect, and τ_{ij} is the circulation mobility effect, whose estimates take the form of the odds ratio.

² Although the IGE is sometimes taken by the economists to be a relative measure of mobility, this relativeness should not be treated to be equal to circulation mobility, since the relativeness means that the IGE shows the difference in log income between children of high versus low income parents (Chetty et al. 2014: 1560), and it does not mean that the IGE is independent from the marginal distributions of income of the two generations.

³ Of course, for empirical researchers, both parents' and children's income can be standardized, but this operation has been less frequently adopted. Moreover, even when standardized income measures are used, the linear correlation coefficient is still driven by the properties of the marginal distributions.

density or sparsity of counts on the two-dimensional “space” constituted by the parents’ and children’s status indicators (occupation, income, or education). Clearly, estimates of mobility patterns shed light on a wide range of theoretical themes. For instance, by seeing it, scholars are enabled to check the symmetry of mobility, i.e., whether or not the mobility from parents’ status group A to children’ status group B has the same extent of strength as that from status group B to status group A. Additionally, the pattern of mobility helps scholars identify the boundaries of mobility “cliques.” That is, the groups of statuses where within-clique mobility is easier than the one between cliques.

In the mobility tables analysis tradition, such a pattern of mobility is examined by virtue of configuring the design matrix for the cross table of occupations. By doing so, sociologists not only establish the nowadays well-known patterns of mobility, such as quasi-independence, uniform association, and quasi-symmetry (for more details see Hout 1983 and Wong 2010), but also show the extent of crossing or penetrability between occupational (or educational) barriers (e.g., the crossing effect model, Mare 1991). Relative to the sociological framework, the economic framework falls short of the nuanced description of the morphological mobility patterns. That is because the IGE, by construction, concerns the “the proportion of the income differences between two sets of parents that we on average will find in (and assessed at the geometric mean of) children’s earnings” (Gregg et al. 2017: 125). In this regard, the IGE informs us the “global” persistence of income across generations, so the more nuanced pattern of mobility or immobility at different “regions” of the multivariate income distribution is not the foci of the IGE.

Covariate Conditioning and Causality

As discussed by Grusky and Cumberworth (2010), sociologists are motivated to study intergenerational mobility by two questions: one is the “class formation” question that examines the extent to which a class structure has been institutionalized, and the other is the “equal opportunity” question that concerns the extent of dependence of one’s life chance access on family origin. These two questions differ from each other in terms of the necessity of conditioning on covariates. The class formation question is descriptive in the sense that it aims at showing the rigidity of immobility and then views it as evidence for the formation (or not) of clear-cut class boundaries. As such, there is no need to control for individual covariates, and the revealed immobility rigidity is limitedly related to causal reasoning. Compared with this question, the equal opportunity question differs in that it is explanatory, with well-articulated cause (characteristics of parents) and effect (characteristics of children). The ideal of equal opportunity cannot be affirmed or rejected, unless the link from family origin to children’s life chance access is real and stands for some substantive and causal relationship. Consequently, a causal interpretation of the empirical association is required, which, operationally, is tantamount to conditioning on the covariates that might confound such an intergenerational association.⁴

The class formation question, over the past several decades, has gradually faded out, and the equal opportunity question has now become the common foci for the social scientists of different disciplines (Evetts 2017). This intellectual trend, as implied by the discussions

⁴ For instance, a seemingly strong correlation of income between parents and children, that is, income immobility, might be attributed to the fact that they both live in urban areas with a high level of salary. Without fixing the variable of region, the intergenerational income association cannot be interpreted causally.

above, necessitates controlling for many covariates. But this is not an easy task. Although some advanced methods, such as the stereotype ordered regression (DiPrete 1990), have been developed to do so, the model configuration process can be complicated, rendering the interpretation of the analytical results not as straightforward as the other modeling strategies. This hardship, in contrast, can be easily overcome when using the IGE. As discussed earlier, the estimation of IGE only requires to fit an OLS model regressing the log transformed children's income on the log transformed parents' income, which is readily open to the addition of individual covariates, such as gender, age, and region, etc. Hence, compared with the sociological approach, the IGE-based economic approach has the relative merit of handily conditioning covariates and enhancing the causal interpretation of empirical results.

It is necessary to mention that it is not our intent to argue for the equality between conditioning and causality. On the contrary, we are aware of the fact that, in many research scenarios, conditioning on the available covariates is insufficient to guarantee causality due to the omitted covariates. However, relative to simple description, covariate conditioning can highly improve causal interpretations (Shpitser, VanderWeele, and Robins 2012). In this regard, the causality in this article is discussed in the relative sense.

Comparison of Mobility Patterns

Sociologists of mobility research have long been interested in the similarities or differences in the pattern of mobility from one social group to another, so a comparative orientation has been indispensable. In the economic research of intergenerational mobility, the inter-group

comparison is also a frequently examined theme. Despite the convergence of research interests, the sociological and economic approaches adopt different analytical strategies.

Specifically, the comparison when using the mobility tables is accomplished by introducing a third dimension of variable in the log-linear model, which constitutes the so-called conditional association models (Wong 2010). Perhaps the most widely known one is the log-multiplicative layer effect model (Xie 1992) or the unidiff model (Erikson and Goldthorpe 1992). Suppose we want to compare the mobility pattern across three societies, this model decomposes the three-way interaction of parents' status, children's status, and society identity into two parts: the association scores that are deemed stable across societies and the level scores that captures how the associational pattern varies from one society to another (suppose one specific society is set as the reference). This model can be further modified to take into more complicated interactive patterns, as illustrated in Wong (2010).

Relative to the mobility tables approach, the IGE can easily accommodate comparison by adding the interaction terms to the OLS model. For instance, if we want to check the gender difference of intergenerational income mobility, we can fit the following model:

$\log(\text{children's income}) = \alpha + \beta \log(\text{parents' income}) + \theta \text{gender} + \lambda \text{gender} \times \log(\text{parents' income}) + \varepsilon$. Like the generic interaction effect analysis, the coefficient we are interested in is λ . It is hasty to give preference to one approach over the other, but from the perspective of practical researchers, the IGE approach is more appealing since it is easy to use, and the coefficient of interest can be routinely estimated and interpreted.

To summarize, the sociological approach based on the mobility tables is adept at revealing the mobility pattern of circulation mobility, but has to resort to complicated and

sometimes cumbersome modeling strategies to control for individual covariates and to make comparisons. In contrast, the IGE-based economic approach, by using the OLS modeling, can easily satisfy the requirement of conditioning on covariates and handle inter-group comparisons. However, the global measure of IGE is short of showing the nuanced mobility patterns, neither the distinction between structural and circulation mobility.

As discussed earlier, these paradigmatic differences balkanize the landscape of mobility research. As a result, when sociologists examine intergenerational income mobility, they are inclined to give up the traditional sociological approach and adopt the economic approach. As a result, some of the traditional sociological questions per mobility, such as the distinction between circulation and structural mobility and the nuanced pattern of mobility, cannot be addressed in the case of income mobility analysis. Meanwhile, the methodological convenience of the IGE approach seems to be independent from the sociological approach to mobility studies.

Moreover, with the increasing availability of administrative data on people's long-term income, the income aspect of intergenerational mobility gains much more academic attention. The bonding between intergenerational income mobility and the IGE approach partly contributes to the dwindling of the sociological approach over the past several decades. We view this bonding as unfortunate, and argue that it is possible to establish a more comprehensive analytical framework that can reconcile and draw on the strengths of the existing two approaches. The tool we draw on is the copula functions.

A COPULA-BASED FRAMEWORK

Copula Functions: A Brief Overview

The copula functions, by definition, refer to a battery of functions that are used to describe the associational patterns of a multivariate distribution. The word copula was derived from the Latin noun for "link," and the copula functions were introduced to the probability theory by Abe Sklar (1959), who showed that any multivariate distribution can be parameterized as a combination of the copula function C and the marginal cumulative distribution (CDF)s. For a given d -dimensional multivariate distribution $F(y_1, y_2, \dots, y_d)$, the Sklar Theorem proves the following equation:

$$F(y_1, y_2, \dots, y_d) = C[F(y_1), F(y_2), \dots, F(y_d)] \quad (1)$$

This formula suggests that a specific copula function C can serve to capture the dependence structure of the multiple marginal variables (converted to their CDFs), and the marginal CDFs and C jointly are sufficient to characterize the multivariate distribution $F(y_1, y_2, \dots, y_d)$.

The copula functions, as one mathematical construct to reflect the structure of statistical dependence, can be used for the research of intergenerational income mobility because the immobility, as the counter-construct of mobility, concerns precisely the strength of dependence between parents' status and children's status (i.e., income). In other words, by cross tabulating the incomes of the two generations (i.e., y_1 and y_2 in formula (1)), we can construct the joint income distribution $F(y_1, y_2)$, which, according to the Sklar Theorem, can be analyzed using the copula functions to show the dependence structure.

An easy way to show the connection between copula functions and mobility is to compare copula functions with the familiar occupational mobility tables. The key information for estimating the log-linear model and characterizing intergenerational mobility is the frequencies of cases in the different "cells" of the occupation cross table. For instance, in the scenario described by Figure 1 (a), most cases are located on the diagonal cells, so a considerable proportion of individuals do not have intergenerational mobility, either upwardly or downwardly. In order to purge out the undesirable influences from the marginal variables, sociologists computed the odds ratio.⁵ For Figure 1(a), the odds ratio of children to stay in the same occupation as their parents compared with otherwise is respectively 3.57 for the occupation pair A-B, 3.67 for the occupation pair B-C, and 13.54 for the occupation pair A-C.

Income mobility differs from occupational mobility in that income is a continuous variable, but we are still able to reveal the pattern of mobility by checking the joint probability density distribution. Like the case of mobility tables analysis, it is necessary to find a way to get rid of the influences from the margins. The copula functions can help to do so (more discussions are presented below). As shown in Figure 1 (b), the density of the copula (not the joint distribution) can well reveal the strong immobility, with the density masses concentrating mostly on the diagonal.

[Figure 1 about here](#)

⁵ Again, readers should note that this purported independence from marginal distributions can be questionable.

To summarize, the density of copula functions can be adopted for the analysis of intergenerational income mobility. In the subsequent subsection, we will discuss how the copula functions better serve scholars to answer the four research questions presented earlier.

How Do the Copula Functions Enhance Intergenerational Income Mobility?

Circulation Mobility. One attractive property of the copula functions is that the morphology of the dependence between two marginal variables can be captured by the specific copula function net of the marginal distributions. This point of view can be readily seen from the Sklar Theorem, where the copula function C and the CDFs of the marginal variables are separate constructs. In this regard, the extent of mobility or immobility as revealed by the copula function can be viewed to be a kind of measure for the circulation mobility.

In previous studies, the association between copula functions and the circulation mobility can be sustained by two practices. One is the increasingly prevalent usage of the rank-rank correlation to study intergenerational mobility, that is, to define intergenerational income persistence to be the strong correlation between the ranking of parents' income and the ranking of children's income (e.g., Zhou and Xie 2019). One merit is that the ranking is less sensitive to the distributional characteristics of the margin. In light of the innate connection between ranking and the position on the CDF, the associational pattern between CDFs, as the gist of the copula functions, should gravitate toward the circulation mobility. The other practice is the decomposition used by economists. For instance, Chetty et al. (2017) defines the rate of absolute mobility to be the fraction of children that earn weakly more than their

parents. Suppose the sample size is N , this absolute mobility can be expressed as

$$\frac{\sum_i \mathbf{1}\{income_i^{children} > income_i^{parents}\}}{N}. \text{ This construct can be further decomposed. Let } r^{children} \text{ and } r^{parents}$$

respectively denote the ranking of parents' and children's income, and $Q(r^{children})$ and

$Q(r^{parents})$ respectively denote the r th quantile of the children and parent income

distributions, the rate of absolute mobility is then equal to

$$\int \mathbf{1}\{Q(r_i^{children}) > Q(r_i^{parents})\} C(r_i^{children}, r_i^{parents}) dr_i^{children} dr_i^{parents} \quad (2)$$

where $C(r_i^{children}, r_i^{parents})$ is the copula function. Clearly, $Q(r^{children})$ and $Q(r^{parents})$

concerns the marginal distributions of parents' and children's income, and

$C(r_i^{children}, r_i^{parents})$ measures the probability that each pair of ranks occurs net of the influences of the margins. Again, this decomposition lends support to the correspondence between the copula functions and the circulation mobility.

Mobility Patterns. Each specific copula function corresponds to a pattern of dependence, or to put it differently, a pattern of (im)mobility. Since the dependence structure may vary from one copula function to another, it goes as no surprise that there are many copula functions that have been developed for scholars to choose. In this article, we cannot exhaustively list all of them, but in order to place readers into perspective, we present four most frequently used copula functions: Normal, Clayton, Frank, and Gumbel, in Figure 2. As shown, the Normal copula function regulates a symmetric dependence pattern, the Clayton copula function is featured by heavy lower-tail dependence, the Gumbel copula function inversely shows a heavy upper-tail dependence structure, and heavy two-tail dependence can

be revealed by the Frank copula function.⁶ An accessible introduction to different copula functions can be found in Hofert et al. (2019) and Vuolo (2017).

Figure 2 about here

The identification of the copula function can be performed using the usual model selection procedure: fitting the data with different candidate copula functions and choose the one with the best fit based on the model evaluation measures, such as AIC (Akaike information criterion) or BIC (Bayesian information criterion). In addition to parametric estimation, we can also ascertain the copula function in the non-parametric way. This would require us to compute the empirical copula function. Specifically, given a series of simulated uniformly distributed variables u_1 and u_2 , the non-parametric empirical copula between parents' income and children's income $C(u_1, u_2)$ can be estimated as the following equation, where $\hat{F}_1$ and $\hat{F}_2$ are respectively the empirical CDFs:

$$C(u_1, u_2) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(\hat{F}_1(\text{parents' income}) \leq u_1, \hat{F}_2(\text{children's income}) \leq u_2) \quad (3)$$

Relative to the parametric estimation of the copula function, the non-parametric approach can better accommodate the specificities of the copula density, but the estimation process may also call for the adoption of some kernel methods for the sake of smoothing.

Covariate Conditioning. Causal interpretation of empirical results, as widely practiced, calls for conditioning on many confounding covariates. When using the copula functions to

⁶ The morphology can be tuned by changing the parameter θ of the respective copula function, and θ has a one-to-one correspondence to the Kendall's tau. However, this parameter, like the IGE or the rank-rank correlation, is a global measure of dependence and it is insufficient to show the (im)mobility pattern.

study intergenerational income mobility, this conditioning can be performed by residualization, that is, we first control for the influences of the covariates on the marginal variables, and then examine the mobility pattern based on the residuals. Intuitively, the residuals are independent from the covariates, so the estimated pattern of mobility can be interpreted more causally than otherwise.

Residualization is a widely adopted strategy to purge out the influences from undesirable confounding covariates (Zhou 2019), and using the copula functions to examine the residuals is not new either to the social scientists. For instance, the copula functions have been proposed to be integrated in the endogenous switching model (e.g., Smith 2003). Given a binary selection variable S , the outcome model for individual i can be written as:

$$\begin{cases} y_{1i} = \mathbf{x}_{1i}'\beta_1 + \varepsilon_{1i} & \text{if } S_i = 1 \\ y_{0i} = \mathbf{x}_{0i}'\beta_0 + \varepsilon_{0i} & \text{if } S_i = 0 \end{cases} \quad (4)$$

In model (4), $\mathbf{x}_{1i}$ and $\mathbf{x}_{0i}$ are the vectors of covariates and the random error terms are respectively ε_{1i} and ε_{0i} . The two error terms are assumed to be independent from each other in the OLS modeling, but in the scenario of selection bias, they can be inter-correlated. In this case, in order to consistently estimate β_1 and β_0 , the dependence structure of ε_{1i} and ε_{0i} should be modelled out, and that is the place where the copula functions can play a role. Clearly, the copula function for the association between ε_{1i} and ε_{0i} is estimated after conditioning on $\mathbf{x}_{1i}$ and $\mathbf{x}_{0i}$, so the dependence structure *per se* can be interpreted more causally.

On the link between the copula functions and causality, several caveats should be mentioned. First, the copula functions can be used to enhance causality not only for the endogenous switching model, but also for general simultaneous equation models where the error terms are believed to be mutually dependent, i.e., the educational transition model by Mare (1980). Second, β_1 , β_0 , and the copula functions can be estimated together by maximizing their joint likelihood function. This operation would accommodate the extra variations of the residuals from the estimation of the outcome models, but fall short of interpretability (Achen 2005). In this regard, we follow the conventional two-step practice by first computing the residuals and then estimating the copula functions (Vuolo 2017).

Comparison. The comparison of a global measure of intergenerational income mobility is easier than the comparison of mobility patterns, since the latter concerns whether or not the mobility morphology differs across some pre-defined groups. When using the copula functions to describe the (im)mobility structure, we can use the Cramér–von Mises (CVM) statistic to perform the mobility pattern comparison.

Specifically, suppose we have two groups with sample size n_1 and n_2 , we can first estimate their empirical copula functions, denoted respectively by $C_{n_1}(u_1, u_2)$ and $D_{n_2}(u_1, u_2)$, and then compute their “difference”, that is, the CVM, as follows:

$$CVM = \int_{[0,1]^d} \frac{C_{n_1}(u_1, u_2) - D_{n_2}(u_1, u_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} du \quad (5)$$

According to Rémillard and Scaillet (2009), formula (5) takes the form of

$$\frac{1}{\frac{1}{n_1} + \frac{1}{n_2}} \times (A - 2B + C) \quad (6)$$

where $A = \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \prod_{s=1}^d [1 - \hat{F}(x_{is,n_1}) \vee \hat{F}(x_{js,n_1})]$, $B = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \prod_{s=1}^d [1 - \hat{F}(x_{is,n_1}) \vee \hat{F}(y_{js,n_2})]$,
 $C = \frac{1}{n_2^2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} \prod_{s=1}^d [1 - \hat{F}(y_{is,n_2}) \vee \hat{F}(y_{js,n_2})]$, and $a \vee b = \max(a, b)$.

We use the CVM statistic to test the hypothesis that the two groups have the same (im)mobility pattern ($H_0: C_{n_1} = D_{n_2}$ and $H_1: C_{n_1} \neq D_{n_2}$).

The randomization method can be used to compute the p value of the CVM. For two groups A and B, their empirical CDFs are expressed respectively as

$$\hat{F}_{n_1}(\mathbf{A}) = [\hat{F}_{1A}(\text{parents' income}), \hat{F}_{2A}(\text{children's income})]$$

$$\hat{F}_{n_2}(\mathbf{B}) = [\hat{F}_{1B}(\text{parents' income}), \hat{F}_{2B}(\text{children's income})]$$

These two matrices are concatenated in the form of $\begin{bmatrix} \hat{F}_{n_1 \times d}(\mathbf{A}) \\ \hat{F}_{n_2 \times d}(\mathbf{B}) \end{bmatrix}$. Subsequently, we randomize the rows. For each round of randomization, two new groups are constructed, one group including the first n_1 rows and the other including the rest n_2 rows. The CVM based on the randomized data is then computed. Suppose we have M rounds of randomization, the p value can be computed as $1 - \frac{\sum_{i=1}^M \mathbf{I}(\text{CVM}_i \leq \text{CVM}_{\text{original data}})}{M}$, where $\text{CVM}_{\text{original data}}$ refers to the CVM using the observed data.

Table 1 about here

The discussions thus far can be summarized in Table 1. By taking advantage of the desirable properties of the copula functions, sociologists are enabled to analyze intergenerational income mobility in the familiar “sociological” fashion while at the same time glean the methodological merits in the IGE-based approach, i.e., to set focus on the nuanced morphological pattern of the circulation mobility, to interpret the analytical results in

a stronger causal sense, and to conduct intergroup comparison of mobility patterns. All of these advantages suggest that the copula functions can be a promising tool for the sociologists of mobility research.

ILLUSTRATION WITH THE NLSY79 DATA

Data and Measures

To illustrate the copula-based analytical framework, we draw on the NLSY79 to study the intergenerational income mobility in the U.S., from the 1970's to the early twenty-first century. A random sample of people aged 14 to 22 was collected in 1979 and these respondents were interviewed annually through 1994 and biennially thereafter. This research design facilitates the analysis of intergenerational social mobility by comparing family income in childhood with individual income in adulthood. In this article, we study income between 1978 and 2017 (survey years 1979 to 2018). To obtain reported parents' income without over-representing late home-leavers, we follow previous studies and exclude those respondents who were older than 19 in 1979 (Bloome 2014). After leaving out a fraction of cases with missing variables, the final sample size is 5539.

In light of the vital role of family income for the well-being of children (Bloome 2014), we examine the total family income instead of father's income only. By definition, the total family income includes labor market earnings, asserts, and transfers from government and nongovernment programs. We first adjust for inflation over years according to the reference year 2018, using the core Consumer Price Index Retroactive Series (R-CPI-U-RS).

Subsequently, we average the total family income across 1978 to 1982 (survey years 1979 to

1983) as one effort to curtail the influences of measurement error (Mazumder 2005), and then adjust it for family needs by dividing it by the square root of family size (Bloome, Dyer, and Zhou 2018). Previous studies have suggested that individuals' permanent income could be best measured when people are above 30 years old (Haider and Solon 2006; Chetty et al. 2014), so we average the CPI-adjusted annual income of adult respondents who are over 30 years old through to the year 2018.

Some covariates are used to show the effect of residualization on the pattern of mobility. For family income, we take into account parents' educational attainment (years of formal education), parents' birthplace (1= in U.S., 0=non-U.S.), whether or not the family resided in the south (1=yes, 0= no) when the respondent was 14 years old, and whether or not the family resided in urban areas (1=yes, 0=no) when the respondent was 14 years old. As for adult's permanent income, we consider respondents' age in 1979, race (1=Hispanic, 2=black, 3=otherwise), and gender (1=male, 0=female). Descriptive analysis of these variables can be found in the Appendix Table A1. The results of the OLS models can be found in the Appendix Tables A2 and A3.

From these models, we can generate the residuals of both family income and adult income, whose distributions are free from the influences of these covariates. We can then compare the copula function of the original non-residualized data with that of the residuals to check the influences of the covariates on the mobility patterns.

One caveat should be mentioned. As discussed earlier, these covariates cannot constitute a full list of potential confounders for the association between family income and adult

income. In other words, we cannot rule out the possibility of the omitted-variable bias in the analytical results presented below. This limitation should be kept in mind.⁷

Analytical Results

The estimate of the copula functions requires the conversion of the observed marginal data to their CDF. This step, known as the probability integral transformation, applies a given theoretical distribution to the marginal variable to compute the probability of observing a value that is equal to or lower than a given observed value. If the theoretical distribution fits the data, the probability would be uniform on the unit line. This is called the uniformity assumption, and it should be tested before moving on to the investigation of the copula functions (Vuolo 2017).

In this study, the theoretical distribution used to fit the data is the normal distribution, as suggested by the histograms shown in Figure 3 where the empirical distributions of log family income, log adult income, and their residuals appear to follow a normal curve. Using the normal distribution to convert the observed data to the CDFs, we have the analytical results presented in Figure 4. Unsurprisingly, the uniformity assumption can be confirmed (as evidenced by the linear pattern on the unit line).

[Figure 3 about here](#)

[Figure 4 about here](#)

⁷ However, controlling for too many covariates might not be always desirable. If some variable is the mediator between parents' income and children's income (i.e., educational attainment of children), fixing the mediator would underestimate the extent of intergenerational transmission of status. Also, the collider bias might be committed if the covariate refers to some common consequence of both parents' and children's income (Elwert and Winship 2014).

After confirming the uniformity assumption, we estimate the copula functions. We first estimate the parametric copula function. Specifically, we fitted 37 different copula functions to the data, and choose the one with the best fit by checking the values of log likelihood, AIC and BIC (For a full list of the 37 copula functions, please see Czado 2019). The resultant copula function is the Survival BB8 (parameter 1 = 2.73, parameter 2 = 0.6) for the dependence structure between original family income and original adult income (Log likelihood = 218.77; AIC= -433.54; BIC=-420.53), and its density is presented in Figure 5 (a). A pattern of strong immobility (that is, strong dependence) is detected at the regions where log family income and log adult income are valued both high or both low. That finding lends support to the intergenerational income persistence for the financially most advantaged and most disadvantaged groups. Relatively, the middle regions have less density mass, meaning that for the middle-income citizens (tentatively defined to be those whose incomes are ranked in the middle region of their respective distribution), a stronger extent of income mobility can be affirmed.

Figure 5 about here

The mobility pattern presented might be subject to the potential confounding of covariates, so we estimated the copula functions for the residualized log family income and residualized log adult income. The copula function with the best fit is still the Survival BB8 (parameter 1= 1.31, parameter 2 = 0.9), but the parameters differ (Log Likelihood= 60.51; AIC=-117.03; BIC=-104.02). The density is described in Figure 5 (b). Although the copula function before and after residualizing belongs to the same family, their differing parameters determine different copula density surfaces. The strong immobility for the most

disadvantaged group still holds, but the extent of immobility at the high-income end is not as strong as that using the original data. In other words, the seemingly strong intergenerational transmission of income advantages for the richest individuals is partly driven by the covariate confoundedness. So, does this difference stand for some statistically significant finding? To answer this question, we deploy the Cramér–von Mises Test to perform the copula comparison. The estimated Cramér–von Mises statistic is 0.553, which is significant at the 0.001 level. This result suggests that residualization and conditioning on the covariates can significantly alter the copula density. Again, readers should note that the covariates we consider here are by no means exhaustive. But conditioning on a very limited number of basic socio-demographic variables can already significantly change the pattern of income mobility. This fact indicates that conditioning on covariates deserves more attention for scholars of mobility research.

Figure 5 (c) and 5(d) present the non-parametric estimate of the copula function density, respectively for the original data and the residualized data.⁸ Since the basic patterns do not change materially from the parametrically estimated ones, we do not discuss them in details.

To better understand the mobility pattern revealed by the copula functions, we can "slice" the estimated copula density at different spots of the log family income. By doing so, we can see how children's income would be ranked in their own generation when parents' income varies, net of the influences of the marginal distributions of both parents' and children's

⁸ Both densities are estimated non-parametrically using the transformation local likelihood method (Geenens, Charpentier, and Paindaveine 2017). The bandwidth matrix for these two subfigures is $\begin{pmatrix} 0.71 & -0.71 \\ 0.71 & 0.71 \end{pmatrix}$. The model goodness of fit indices are Log Likelihood=231.72, AIC=-434.29, and BIC= -339.49 for the original data, and Log Likelihood=74.84, AIC=-120.18, and BIC= -24.23 for the residualized data.

income. Specifically, we select the 20th, 40th, 60th, and 80th quantiles for both parents' and children's income. By definition, a smaller quantile corresponds to a relatively lower ranking of income. We then see the probabilities of children's income to reach the four quantiles when their parents' income changes across the four quantiles. The results can be found in Figure 6.

Figure 6 about here

Figure 6 (a) uses the original data. It is shown that those whose parents' income is located in the middle range (the 40th and 60th quantiles) have roughly similar probabilities of having their own income fall into the four quantiles. However, respondents from the disadvantaged families (parents' income at the 20th quantile) mostly concentrate in the 20th quantile income in their own generation, and their likelihood of moving upward to the higher quantiles rapidly and linearly declines (marked out by the left dotted square). The situation for the advantaged group (parents' income at the 80th quantile) suggests their parents' income advantage can be very likely to be maintained by their children, and the probability of falling down into a lower income quantile steeply declines (marked out by the right dotted square).

Figure 6 (b) report the analytical results using the residualized data. In general, a similar pattern is shown. However, the extent of “maintenance” of income (dis)advantages for the low-income and high-income groups is less stark than that of the original data. For those whose parents' income is at the 20th quantile, the “slope” of the declining probability of moving to an upper income quantile is not much less steep than that of the original data. This is also the case for the 80th quantile of parents' income, where the extent of drop in the probabilities of children's falling into a lower income quantile is more gentle. Since the

extent of “obliquity” in the transitioning probabilities of moving into the other income quantiles can be understood to reflect the income immobility, we may conclude that conditioning on covariates reduces the extent of income immobility at the two extreme groups, which again necessitates covariate conditioning when examining income mobility, especially at the top or the bottom.

[Figure 7 about here](#)

Our last analysis illustrates the comparison of mobility patterns between groups of individuals. Specifically, we look into the gender difference. We first estimate the copula densities non-parametrically respectively for the male and female respondents. Note that the residualized log adult income has already controlled for the variable of gender, so we only residualize the log family income.

The results can be found in Figure 7. The Cramér–von Mises statistic for the original data is 0.143, which is statistically significant at the 0.001 level. However, this statistic fails to be statistically significant after we residualize the log family income (The Cramér–von Mises statistic =0.05, and the p value =0.06). In this regard, it seems that the gender difference in the pattern of intergenerational income (im)mobility is partly driven by the confounding covariates. After fixing the basic socio-demographic variables for (only) the log family income, the gender difference in the income mobility patterns ceases to be significant.

CONCLUDING REMARKS

Despite the common interest in intergenerational mobility, the empirical approaches diverge between sociologists and economists. This “balkanization” of the mobility research is

unfortunate, since many of the academic questions sociologists traditionally explore, such as the patterns of the circulation mobility, are rarely asked by the research on income mobility. At the same time, the analytical convenience of the economic approach in terms of the comparison of mobility status across groups and the conditioning of potentially confounding covariates seem to be kept outside from the mobility tables school.

In fact, the divergence of analytical approaches precisely implies the room for mutual learning. In this article, we argue that the distinct methodological merits of these two approaches can be equipped by a more comprehensive analytical framework using the copula functions: (1) The copula functions concern the dependence structure that is independent from the margins, which enable scholars to focus on the relative mobility; (2) The copula density, estimated either parametrically or non-parametrically, reveals the nuanced morphological mobility patterns; (3) By residualizing the marginal variables, the detected mobility pattern can be interpreted in a stronger causal sense; and (4) the Cramér–von Mises Test offers an easy-to-use statistic to conduct intergroup comparison of mobility patterns. The copula-based framework is illustrated by investigating the income mobility between 1978 and 2017 in the U.S., using the National Longitudinal Survey of Youth 1979 (NLSY79).

The potential of the copula functions for mobility research is not fully exposited. For instance, in addition to the familiar two-generation mobility, recent studies shift attention to the mobility involving three or more generations (Mare 2014). This multi-generation mobility question calls for the estimate of the conditional mobility pattern between two generations (e.g., parents and children) given a particular characteristic of the third generation (e.g., grandparents). In this case, the vine copula methods might be used. A similar research

scenario is the interaction between inter- and intra-generational mobility, since they may condition each other (Cheng and Song 2019).

The copula functions are constructed to capture the dependence structure of multiple continuous variables, so they confront with limitations when dealing with categorical variables. In this scenario, the copula function can be not identifiable. Because of this, it is rare for mobility researchers to use the copula functions to analyze occupational mobility. To smoothen the categorical variable, random jitters may be used (Denuit and Lambert 2005).

References

- Achen, Christopher H. 2005. "Two-Step Hierarchical Estimation: Beyond Regression Analysis." *Political Analysis* 13(4):447-56.
- Bearman, P., and G. Deane. 1992. "The Structure of Opportunity: Middle-Class Mobility in England, 1548-1689." *American Journal of Sociology* 98:30 - 66.
- Beller, Emily, and Michael Hout. 2006. "Intergenerational Social Mobility: The United States in Comparative Perspective." *The Future of Children* 16(2):19-36.
- Bloome, Deirdre. 2014. "Income Inequality and Intergenerational Income Mobility in the United States." *Social Forces* 93(3):1047-80.
- . 2017. "Childhood Family Structure and Intergenerational Income Mobility in the United States." *Demography* 54(2):541-69.
- Bloome, Deirdre, Shauna Dyer, and Xiang Zhou. 2018. "Educational inequality, educational expansion, and intergenerational income persistence in the United States." *American Sociological Review* 83(6):1215-53.

- Breen, Richard, Jan Jonsson, xa, and O. 2007. "Explaining Change in Social Fluidity: Educational Equalization and Educational Expansion in Twentieth‐Century Sweden." *American Journal of Sociology* 112(6):1775-810.
- Cheng, Siwei, and Xi Song. 2019. "Linked Lives, Linked Trajectories: Intergenerational Association of Intragenerational Income Mobility." *American Sociological Review* 84(6):1037-68.
- Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The fading American dream: Trends in absolute income mobility since 1940." *Science* 356(6336):398-406.
- 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 *." *The Quarterly Journal of Economics* 129(4):1553-623.
- Christensen, Ronald. 1994. *Log-Linear Models*: Springer-Verlag New York.
- Czado, C. 2019. *Analyzing Dependent Data with Vine Copulas: A Practical Guide With R*: Springer International Publishing.
- Denuit, Michel, and Philippe Lambert. 2005. "Constraints on concordance measures in bivariate discrete data." *Journal of Multivariate Analysis* 93(1):40-57.
- DiPrete, Thomas A. 1990. "Adding Covariates to Loglinear Models for the Study of Social Mobility." *American Sociological Review* 55(5):757-73.
- Elwert, Felix, and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40(1):31-53.
- Erikson, Robert, and John Harry Goldthorpe. 1992. *The constant flux : a study of class mobility*

- in industrial societies*: Oxford : Clarendon press.
- Evetts, J. 2017. *The Sociology of Educational Ideas*: Taylor & Francis.
- Featherman, David L., and Robert M. Hauser. 1978. "Opportunity and change." edited by Robert Mason Hauser. New York :: Academic Press.
- Fox, L., F. Torche, and J. Waldfogel. 2016. "Intergenerational mobility." in *The Oxford Handbook of the Social Science of Poverty*, edited by David Brady and Linda M. Burton. Oxford: Oxford University Press.
- Geenens, Gery, Arthur Charpentier, and Davy Paindaveine. 2017. "Probit transformation for nonparametric kernel estimation of the copula density." *Bernoulli* 23(3):1848-73, 26.
- Gregg, Paul, Jan O. Jonsson, Lindsey Macmillan, and Carina Mood. 2017. "The Role of Education for Intergenerational Income Mobility: A comparison of the United States, Great Britain, and Sweden." *Social Forces* 96(1):121-52.
- Grusky, David B., and Kim A. Weeden. 2006. "Does the Sociological Approach to Studying Social Mobility Have a Future?" Pp. 85-108 in *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, edited by Stephen Morgan, Gary Fields, and David B. Grusky. 2006. Stanford, Calif: Stanford University Press.
- Grusky, David B., and Erin Cumberworth. 2010. "A National Protocol for Measuring Intergenerational Mobility?" Paper prepared for the *National Academy of Science Workshop on Advancing Social Science Theory: The Importance of Common Metrics*.
- Grusky, David B., Timothy M. Smeeding, and C. Matthew Snipp. 2015. "A New Infrastructure for Monitoring Social Mobility in the United States." *The ANNALS of the American Academy of Political and Social Science* 657(1):63-82.

- Haider, Steven, and Gary Solon. 2006. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96:1308–20.
- Harrison, Roderick J. 1988. "Do odds ratios really control for the availability of occupational positions in status contingency tables?" *European Sociological Review* 4(1):65-79.
- Hofert, M., I. Kojadinovic, M. Mächler, and J. Yan. 2019. *Elements of Copula Modeling with R*: Springer International Publishing.
- Hout, Michael. 1983. *Mobility Tables*. Thousand Oaks, California: Sage Publications.
- . 2015. "A Summary of What We Know about Social Mobility." *The ANNALS of the American Academy of Political and Social Science* 657(1):27-36.
- Mare, Robert D. 1980. "Social Background and School Continuation Decisions." *Journal of the American Statistical Association* 75(370):295-305.
- . 1991. "Five Decades of Educational Assortative Mating." *American Sociological Review* 56(1):15-32.
- . 2014. "Multigenerational aspects of social stratification: Issues for further research." *Research in Social Stratification and Mobility* 35:121-28.
- 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):235–55.
- Rémillard, Bruno, and Olivier Scaillet. 2009. "Testing for equality between two copulas." *Journal of Multivariate Analysis* 100(3):377-86.
- Ritzer, George. 1975. "Sociology: A Multiple Paradigm Science." *The American Sociologist* 10(3):156-67.

- Sakamoto, Arthur, and Sharron Xuanren Wang. 2020. "The declining significance of occupation in research on intergenerational mobility." *Research in Social Stratification and Mobility* 70:100521.
- Shpitser, I., T. VanderWeele, and J. M. Robins. 2012. "On the validity of covariate adjustment for estimating causal effects." *arXiv e-prints*:arXiv:1203.3515.
- Sklar, A. (1959) Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8, 229-231.
- Smith, Murray D. 2003. "Modelling sample selection using Archimedean copulas." *The Econometrics Journal* 6(1):99-123.
- Sobel, Michael E. 1983. "Structural Mobility, Circulation Mobility and the Analysis of Occupational Mobility: A Conceptual Mismatch." *American Sociological Review* 48(5):721-27.
- Torche, Florencia. 2015. "Analyses of Intergenerational Mobility: An Interdisciplinary Review." *The ANNALS of the American Academy of Political and Social Science* 657(1):37-62.
- Vuolo, Mike. 2017. "Copula Models for Sociology: Measures of Dependence and Probabilities for Joint Distributions." *Sociological methods & research* 46(3):604-48.
- Xie, Yu. 1992. "The Log-Multiplicative Layer Effect Model for Comparing Mobility Tables." *American Sociological Review* 57(3):380-95.
- Zhou, Xiang. 2019. "Equalization or Selection? Reassessing the “Meritocratic Power” of a College Degree in Intergenerational Income Mobility." *American Sociological Review* 84(3):459-85.

Zhou, Xiang, and Yu Xie. 2019. "Market Transition, Industrialization, and Social Mobility Trends in Postrevolution China." *American Journal of Sociology* 124(6):1810-47.

Figure 1: The Comparison of Mobility Table Frequency Distribution and the Copula Density

(a)

(b)

Figure 2: Illustration of the Density of the Normal, Clayton, Frank, and Gumbel Copula

Functions

Figure 3: Descriptive Information of Family Income and Adult Income

Data Source: the NLSY79

Figure 4: Test of the Uniformity Assumption

Data Source: the NLSY79

Figure 5: Estimates of the Copula Functions

(a)

(b)

(c)

(d)

Note: (a) and (b) are estimated parametrically, where the copula function of (a) is the Survival BB8 (parameter 1 = 2.73, parameter 2 = 0.6) and the copula function of (b) is the Survival BB8 (parameter 1 = 1.31, parameter 2 = 0.9). (c) and (d) are estimated non-parametrically using the kernel functions.

Data Source: the NLSY79

Figure 6: Estimated Probabilities of Children's Income Quantiles

(a) Original Data

(b) Residualized Data

Data Source: the NLSY79

Figure 7: Comparison between Males and Females

Note: (a) and (b) concern the original data and (c) and (d) concern the residualized data.
Data Source: the NLSY79

Table 1: Comparisons between Analytical Frameworks

Research Questions	Mobility Tables Analysis	IGE-based Economic Modeling	Copula Functions
<i>Distinction between Structural and Circulation Mobility</i>	(Questionable) Independence of Odds Ratio from the Marginal Distribution	Null	Cumulative Distribution Transformation
<i>Recognition of Mobility Pattern</i>	Design Matrices	Null	Parametric/Nonparametric Copula Function Identification
<i>Causal Interpretation/Conditioning</i>	Complicated, such as the Stereotype Ordered Regression	Controlling for Covariates in a Linear Model	Residualization
<i>Comparison of Mobility Patterns between Groups</i>	Conditional Association Models	Interaction Analysis in a Linear Model	The Cramér–von Mises Test

Appendix

Table A1: Descriptive Statistics of the NLSY79

Variable	Mean	Std. Dev.
<i>Family Characteristics</i>		
Family Income, Child Age 14-19 in 1979	7.942	0.706
Education attainment of mother	11.062	3.071
Education attainment of father	11.037	3.847
Mother born in U.S.(%)	89.29	
Father born in U.S.(%)	88.41	
Residing in the South at age 14(%)	36.85	
Residing in urban areas at age 14(%)	78.9	
<i>Adult Characteristics</i>		
Income above 30	10.05	1.559
Age in 1979	16.683	1.632
Male(%)	50.47	
Race(%)		
Hispanic	16.69	
Black	26.22	
Others	57.09	
N	5539	

Table A2: Results of the Ordinary Least Squares Models Predicting Family Income

Education attainment of mother	.053 (-0.003) ***
Education attainment of father	.051 (-0.003) ***
Mother born in U.S.	-.099(-0.038) ***
Father born in U.S.	-.172(-0.04) ***
Residing in the South at age 14	-.122(-0.016) ***
Residing in urban areas at age 14	-0.026(-0.019)
Intercept	7.068(-0.04) ***
N	7048
R-squared	0.203

Unstandardized coefficients with standard errors in parentheses

*** p<.001, ** p<.01, * p<.05

Table A3: Results of the Ordinary Least Squares Models Predicting Adult Income

Age in 1979	0.259 (-0.011)***
Male	0.524(-0.036)***
Race (Hispanic)	
Black	-0.297(-0.052)***
Others	0.214(-0.048)***
Intercept	5.357(-0.186)***
N	7032
R-squared	0.117

Standard errors are in parentheses

*** p<.001, ** p<.01, * p<.05