The University of Sydney Business School
The University of Sydney

BUSINESS ANALYTICS WORKING PAPER SERIES

Two-Sample Nonparametric Estimation of Intergenerational Income Mobility

Irina Murtazashvili, Di Liu, Artem Prokhorov

Abstract

We estimate intergenerational income mobility in the USA and Sweden. To measure the degree to which income status is transmitted from one generation to another we propose a nonparametric estimator, which is particularly relevant for cross-country comparisons. Our approach allows intergenerational mobility to vary across observable family characteristics. Furthermore, it fits situations when data on fathers and sons come from different samples. Finally, our estimator is consistent in the presence of measurement error in fathers' long-run economic status. We find that family background captured by fathers' education matters for intergenerational income persistence in the USA more than in Sweden suggesting that the character of inequality in the two countries is rather different.

Keywords: Intergeneration Income Persistence; GMM estimation

JEL Classification: D3, C14

April 2013

BA Working Paper No: 07/2013
http://sydney.edu.au/business/business_analytics/research/working_papers

Two-Sample Nonparametric Estimation of Intergenerational Income Mobility*

Irina Murtazashvili†
Drexel University

Di Liu‡
Concordia University

Artem Prokhorov§
Concordia University

April 2, 2013

Abstract

We estimate intergenerational income mobility in the USA and Sweden. To measure the degree to which income status is transmitted from one generation to another we propose a nonparametric estimator, which is particularly relevant for cross-country comparisons. Our approach allows intergenerational mobility to vary across observable family characteristics. Furthermore, it fits situations when data on fathers and sons come from different samples. Finally, our estimator is consistent in the presence of measurement error in fathers' long-run economic status. We find that family background captured by fathers' education matters for intergenerational income persistence in the USA more than in Sweden suggesting that the character of inequality in the two countries is rather different.

JEL Codes: D3, C14

Key Words: Intergeneration Income Persistence; GMM estimation

*This paper has benefitted from helpful comments from and conversations with Soiliou Namoro, Jean-Francois Richard, Gary Solon, Aman Ullah, Jeffrey Wooldridge, and Victoria Zinde-Walsh. We are grateful to Markus Jäntti for providing us with the Swedish data. The third author gratefully acknowledges the financial support from the Social Sciences and Humanities Research Council of Canada.

†Corresponding Author. Department of Economics and International Business, LeBow College of Business, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104. Tel: (215)895-6007, fax: (215)895-6975, and e-mail: im99@drexel.edu.

‡Department of Economics, Concordia University, 1455 De Maisonneuve Blvd W, Montreal QC H3G1M8 Canada; email: flyingliudi@gmail.com.

§Department of Economics, Concordia University, 1455 De Maisonneuve Blvd W, Montreal QC H3G1M8 Canada; email: Artem.Prokhorov@concordia.ca.

1 Introduction

The extent of income mobility across generations has kept economists' attention for a long while. Solon (2002) provides a thorough survey of the existing literature on how fathers' long-run economic status affect those of their sons in different countries. Cross-country comparisons of intergenerational income mobility allow to study the character of inequality in a particular society. As Björklund and Jäntti (1997) point out, cross-country comparisons of income mobility produce some insights on whether cross-sectional and intergenerational inequalities are linked to each other.

Following Björklund and Jäntti (1997), our study focuses on measuring intergenerational income mobility in Sweden and the United States. A comparison among the United States and Scandinavian countries has been viewed as one of the most relevant when studying intergenerational income mobility across countries.[1] This is in part due to the finding highlighted in Gottschalk and Smeeding (1997) that the Scandinavian countries have the lowest annual income inequality in contrast to the United States that is among the countries with the highest.

Three key issues have been discussed in the literature when conducting cross-country comparisons of intergenerational mobility. First, Bratsberg, Røed, Raaum, Naylor, Jäntti, Eriksson, and Österbacka (2007) emphasize that the appropriateness of elasticities of sons' income with respect to their fathers' income as traditional measures of intergenerational income mobility depends on the functional relationship between fathers' and sons' income being linear in logs. If the functional form is nonlinear, elasticities estimated using linear models can be quite misleading. There were several attempts to address this concern. In particular, Bhattacharya and Mazumder (2011) suggest a new direct measure of intergenerational income mobility obtained by using a nonparametric model where the conditional transition probability of moving across income quantiles varies with

[1] Gustafsson (1994), Björklund and Jäntti (1997), Aaberge, Björklund, Jäntti, Palme, Pedersen, Smith, and Wennemo (2002) and Bratsberg, Røed, Raaum, Naylor, Jäntti, Eriksson, and Österbacka (2007) are just a few studies focusing on intergenerational economic status transmission in these countries.

individual-specific covariates. Murtazashvili (2012) proposes yet another alternative measure of intergenerational mobility, which is based on a random coefficient model and which allows intergenerational income mobility to vary across the distribution of families.

Second, the lack of sufficiently comparable cross-country data has also been acknowledged to cause problems for studying intergenerational mobility in different countries.[2] Björklund and Jäntti (1997) address this concern by employing a two-sample two stage least squares (TS-2SLS) approach, proposed by Angrist and Krueger (1992), to deal with the fact that fathers' and sons' incomes come from different samples.

Finally, Solon (1992) emphasizes dangers of ordinary least squares (OLS) estimation of intergenerational elasticities due to measurement error in fathers' permanent economic status. If a valid instrument is available, Solon (1992) proposes using instrumental variables (IV) estimation to obtain consistent estimates of intergenerational income mobility.

In this paper, we propose a new econometric approach to address the three main concerns arising when conducting cross-country comparisons in intergenerational income mobility. Specifically, to allow intergenerational income mobility to vary across the population of families we employ a nonparametric GMM (NPGMM) approach by Cai and Li (2008). This method permits exploiting the functional form flexibility of nonparametric estimation to introduce heterogeneity in intergenerational elasticities across the population of families as a function of observed family characteristics. In addition, the GMM nature of the approach allows for consistent estimation of intergenerational elasticities in the presence of measurement error in fathers' income. Furthermore, we develop an extension of the NPGMM estimator along the lines of Angrist and Krueger (1992) to allow observations on fathers' and sons' incomes to come from different samples. We use our approach to estimate intergenerational income mobility in the USA and Sweden. Similar to the existing studies,

[2]For example, see Aaberge, Björklund, Jäntti, Palme, Pedersen, Smith, and Wennemo (2002) for a relevant discussion.

we find that family background matters much more for labor market achievements in the United States than in Sweden.

The rest of the paper is organized as follows. Section 2 introduces the econometric model of our interest and describes our two-sample nonparametric GMM (TS-NPGMM) estimator. Section 3 describes our data. In Section 4 we estimate intergenerational income mobility in the USA and Sweden using our two-sample nonparametric (TS-NPGMM) approach. Section 5 concludes.

2 Our Econometric Model and Its Estimation

In compliance with the extensive literature on intergenerational income mobility, to measure the degree to which income status is transmitted from one generation to another, as a starting point, we employ the intergenerational income elasticity from the following model:

$$y_S = \rho y_F + \varepsilon. \tag{1}$$

Here, y_S and y_F are the natural logarithms of permanent incomes of sons and fathers, respectively, and ε is an idiosyncratic error.[3] We follow Solon (1992) and deviations of log income from generation means not to include the intercept in model (1).

In order to address one of the major concerns raised by economists that the standard elasticity of sons' income with respect to that of their fathers' does not appropriately capture the nonlinearity in the intergenerational transmission mechanism of economic status, we modify model (1) to explicitly allow for dependence between the intergenerational elasticity and some characteristic(s) of families.

[3]Here, ρ is referred to as the intergenerational income elasticity if the variances in the long-run economic statuses of fathers and sons are different, i.e., $Var(y_F) = \sigma_F^2 \neq \sigma_S^2 = Var(y_S)$. In a special case when $\sigma_F^2 = \sigma_S^2$ the slope coefficient from (1) coincides with intergenerational income correlation.

Thus, we employ the following econometric model:

$$y_S = \rho(\mathbf{z}_1)y_F + \varepsilon, \tag{2}$$

where $\mathbf{z}_1$ contains some family characteristic(s) observable to researchers. While also allowing for explicit variation in intergenerational mobility across the distribution of families, both Bhattacharya and Mazumder (2011) and Murtazashvili (2012) are agnostic about any specific characteristics of the transmission mechanism of earnings from one generation to the next. Here, employing model (2), we allow observable family characteristics to better describe the intergenerational transmission mechanism of earnings at various points of the distribution of families.

Unfortunately, data limitations do not allow us to employ conventional econometric methods to estimate the functional coefficient of our interest, $\rho(\mathbf{z}_1)$. First, as it has been widely discussed in the literature, the desired permanent earnings of fathers and sons are unobserved. Instead, short-run earnings in the form of either annual earnings or even hourly earnings have been used to measure the long-run economic status of fathers and sons. Due to the measurement error in short-run earnings used to proxy long-run earnings, traditional methods used to estimate intergenerational mobility, which ignore the endogeneity of y_F, suffer from the well-known "attenuation" bias. Second, conventional econometric methods, which are able to account for endogeneity in y_F, assume that all the variables contained in the model of interest are available in the same sample. Contrary to the traditional assumption and similar to the intergenerational income mobility studies by Björklund and Jäntti (1997) and Lefranc and Trannoy (2005), we are faced with the situation when information on fathers' and sons' income comes from two different samples.

When model (1) is the model of interest, Angrist and Krueger (1992) propose a parametric method that can deal with situations when data used in the analysis come from two different samples. In essence, the authors suggest using instrumental variables, $\mathbf{z}_2$, to predict y_F and, then, employ the predicted y_F to estimate ρ from model (1). We propose a way to combine a particular existing nonparametric approach to estimation of model (2) – the method due to Cai and Li (2008) – with the parametric approach by Angrist and Krueger (1992) to incorporate situations when fathers' and sons' incomes come from different samples in order to estimate $\rho(\mathbf{z}_1)$ from model (2).

An important additional advantage of the two-sample nonparametric approach we advocate is that it produces consistent estimates of the intergenerational earnings mobility that can vary across the distribution of families when both the data used in the analysis come from two samples and the measurement error in the explanatory variable is present (provided instruments used are valid).

2.1 Our General Econometric Framework

In this section we formally describe our econometric model in a general setting. Our model of interest can be written as follows:

$$y_i = \mathbf{x}_i \mathbf{b}(\mathbf{z}_{1i}) + u_i, \tag{3}$$

where y_i is a response variable, $\mathbf{x}_i$ is a $1 \times K$ vector of endogenous explanatory variables, $\mathbf{z}_{1i}$ is a $1 \times L_1$ subvector of the vector of exogenous variables $\mathbf{z}_i = (\mathbf{z}_{1i}, \mathbf{z}_{2i})$, $\mathbf{b}(\cdot)$ is an unknown K-valued function on $\mathbb{R}^{L_1}$, with typical element $b_j(\cdot)$, $j = 1, .., K$, and u_i is an idiosyncratic error. As usual, we assume that $\mathrm{E}[u_i|\mathbf{z}_i] = 0$, i.e. $\mathbf{z}_i$ is exogenous. We will assume that $L = L_1 + L_2 \geq K$, where $L_2 = \dim(\mathbf{z}_{2i})$, so that there is at least one instrument not in $\mathbf{z}_{1i}$ for every endogenous explanatory variable in $\mathbf{x}_i$.

In the nonparametric literature, model (3) is called a varying (or functional) coefficient model.[4] In a parametric context, when $\mathbf{b}(\mathbf{z}_{1i}) \equiv \mathbf{b}_i$ model (3) is called a random coefficient model. The

[4] Our model (3) incorporates many linear and partially linear models and has been used in many applications other than intergenerational income mobility, including exchange rate forecasts (see, e.g., Hong and Lee, 2003) and US unemployment and interest rate analysis (see, e.g., Juhl, 2005).

possibility of correlation between the individual-specific coefficients $\mathbf{b}_i \equiv \mathbf{b}(\mathbf{z}_{1i})$ and $\mathbf{x}_i$ makes (3) a correlated random coefficient model.

When y, $\mathbf{x}$, and $\mathbf{z}$ are contained in the same sample, numerous estimation methods have been developed.[5] Recently, Cai and Li (2008) propose a new nonparametric GMM estimator, which combines the local linear fitting technique and the generalized method of moments. The method proposed by Cai and Li (2008) is computationally simpler than other nonparametric methods, and we focus on this method here. Situations when the pairs $(y, \mathbf{z})$ and $(\mathbf{x}, \mathbf{z})$ come from two samples cause difficulties for researchers in many empirical fields including those economists who study intergenerational mobility (see, for example, Angrist and Krueger (1992) and Arellano and Meghir (1992)). We extend the approach by Cai and Li (2008) to incorporate these situations.

2.2 Two-Sample Nonparametric GMM Estimator

In this section, we present the two-sample nonparametric GMM estimator and consider conditions under which the nonparametric estimator by Cai and Li (2008) is consistent and asymptotically normal in the two-sample context.

For a given point $\mathbf{z}_1 \in \mathbb{R}^{L_1}$ and for $\{\mathbf{z}_{1i}\}$ in a neighborhood of $\mathbf{z}_1$, assuming that $\{b_j(\cdot)\}$ are twice continuously differentiable, we exploit Taylor expansions to approximate $b_j(\mathbf{z}_{1i})$ by a linear function $b_{j0} + \mathbf{b}_{j1}(\mathbf{z}_{1i} - \mathbf{z}_1)$, where $b_{j0} = b_j(\mathbf{z}_1)$ and $\mathbf{b}_{j1} = \frac{\partial b_j(\mathbf{z}_1)}{\partial \mathbf{z}_1}$. So, (3) can be approximated locally by

$$y_i \simeq \mathbf{w}_i \beta + u_i, \tag{4}$$

where $\mathbf{w}_i = \{\mathbf{x}_i, \mathbf{x}_i \otimes (\mathbf{z}_{1i} - \mathbf{z}_1)\}$ is a $1 \times K(1 + L_1)$ vector, and $\beta = (b_{10}, ..., b_{K0}, \mathbf{b}'_{11}, ..., \mathbf{b}'_{K1})'$ is a $K(1 + L_1) \times 1$ vector of parameters. Thus, for any vector function $\mathbf{Q}(\mathbf{z}_i)$, we can rewrite the

[5] See, for example, Zhang, Lee, and Song (2002) and Cai, Das, Xiong, and Wu (2006), among others.

standard orthogonality conditions $\mathrm{E}[u_i|\mathbf{z}_i] = 0$ in the following form:

$$\mathrm{E}[\mathbf{Q}(\mathbf{z}_i)u_i|\mathbf{z}_i] = 0. \tag{5}$$

If we observe all the variables y_i, $\mathbf{x}_i$, and $\mathbf{z}_i$ in one sample, the local approximation of conditions (5) will produce the (one-sample) nonparametric GMM (NPGMM) estimator of Cai and Li (2008). We are interested in estimating model (3) using data that come from two different samples. To emphasize the distinction between the two samples we use superscripts (1) and (2). The difficulty in estimation arises due to the fact that the first data set contains only $\{y_i^{(1)}, \mathbf{z}_i^{(1)}\}$, $i = 1, ..., N_1$, while the second data set contains only $\{\mathbf{x}_l^{(2)}, \mathbf{z}_l^{(2)}\}$, $l = 1, ..., N_2$.

The fundamental difficulty here is that, due to the data structure, calculation of model residuals is infeasible. The dependent variable, the instruments and the independent variables are not available in the same sample and so error-based objective functions, such as the sum of squared residuals, cannot be used. In the setting of nonparametric estimation, this also means that traditional methods of data-driven bandwidth selection and variance estimation are infeasible because they are based on residuals. However, under certain conditional moment assumptions it is still possible to obtain a consistent nonparametric estimator based on averages from the two samples.

In the context of two samples, condition (5) can be approximated by the locally weighted moment condition (6).

$$\mathrm{E}[\mathbf{Q}(\mathbf{z}_i^{(1)})y_i^{(1)}K_{h_1}(\mathbf{z}_{1i}^{(1)} - \mathbf{z}_1) - \mathbf{Q}(\mathbf{z}_l^{(2)})\mathbf{w}_l^{(2)}\beta K_{h_2}(\mathbf{z}_{1l}^{(2)} - \mathbf{z}_1)] = 0, \tag{6}$$

where where $K_{h_j}(\cdot)$ is a bounded symmetric kernel function in $\mathbb{R}^{L_1}$, $j = 1, 2$, h_1 and h_2 are bandwidths and the dimension of $\mathbf{Q}(\cdot)$ must be at least $K(1 + L_1)$. Though there are many possibilities for $\mathbf{Q}(\cdot)$, we follow Cai and Li (2008) in using the following form

$$\mathbf{Q}(\mathbf{z}_i) = (\mathbf{z}_{2i}, \mathbf{z}_{2i} \otimes (\mathbf{z}_{1i} - \mathbf{z}_1)/h_2)' : \quad L_2(1 + L_1) \times 1$$

Clearly, for such a choice of $\mathbf{Q}(\mathbf{z}_i)$, a necessary identification condition is $L_2 \geq K$.

Define the following local averages

$$\mathbf{S}_2 = \frac{1}{N_2} \sum_{l=1}^{N_2} \mathbf{Q}(\mathbf{z}_l^{(2)}) \mathbf{w}_l^{(2)} K_{h_2}(\mathbf{z}_{1l}^{(2)} - \mathbf{z}_1) \tag{7}$$

$$\mathbf{T}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} \mathbf{Q}(\mathbf{z}_i^{(1)}) K_{h_1}(\mathbf{z}_{1i}^{(1)} - \mathbf{z}_1) y_i^{(1)} \tag{8}$$

Then, the two-sample nonparametric GMM (TS-NPGMM) estimator we propose can be written as follows

$$\hat{\beta} = (\mathbf{S}_2' \mathbf{S}_2)^{-1} \mathbf{S}_2' \mathbf{T}_1. \tag{9}$$

Implicitly, the estimator in (9), as well as its components $\mathbf{S}_2$ and $\mathbf{T}_1$, are functions of $\mathbf{z}_1$. In essence it is a nonparametric estimator of $\mathbf{b}(\mathbf{z}_1)$ and of its first-order derivatives $\nabla b_j(\mathbf{z}_1)$, where $j = 1, ..., K$, obtained combining the local linear fitting techniques and the generalized method of moments. While other smoothing methods are available, this approach has some rather attractive statistical properties in comparison with other smoothing methods (see, e.g., Fan and Gijbels, 1996, for details).

We now impose some regularity conditions sufficient for consistency and asymptotic normality of (9).

Assumptions:

(1) $\{y_i^{(1)}, \mathbf{x}_i^{(1)}, \mathbf{z}_i^{(1)}, u_i^{(1)}\}$ and $\{y_i^{(2)}, \mathbf{x}_i^{(2)}, \mathbf{z}_i^{(2)}, u_i^{(2)}\}$ are two independent samples from the population, for which observations are independent across i and only $\{y_i^{(1)}, \mathbf{z}_i^{(1)}\}$ and $\{\mathbf{x}_i^{(2)}, \mathbf{z}_i^{(2)}\}$ are observed. Further, $\mathrm{E}||\mathbf{z}_{2i}^{(j)\prime}\mathbf{x}_i^{(j)}||^2 < \infty$, $\mathrm{E}||\mathbf{z}_{2i}^{(j)\prime}\mathbf{z}_{2i}^{(j)}||^2 < \infty$, and $\mathrm{E}|u_i^{(j)}|^2 < \infty$, where $||\mathbf{A}||^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}')$, and $j = 1, 2$.

(2) For each $\mathbf{z}_1$, $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\mathbf{z}_1) = \mathrm{E}[\mathbf{z}_{2i}'\mathbf{x}_i|\mathbf{z}_{1i} = \mathbf{z}_1] > 0$, $f(\mathbf{z}_1) > 0$, where $f(\mathbf{z}_1)$ is the density function of $\mathbf{z}_{1i}$, and $\mathbf{b}(\mathbf{z}_1)$ and $f(\mathbf{z}_1)$ are both twice continuously differentiable at $\mathbf{z}_{1i} \in \mathbb{R}^{L_1}$.

(3) The kernel $K(\cdot)$ is a symmetric, non-negative and bounded second-order kernel function having a compact support; $h_j \to 0$, $h_2/h_1 \to 1$ and $N_j h_j^{L_1} \to \infty$ as $N_j \to \infty$, $j = 1, 2$. Also, $\lim_{N_2 \to \infty} \frac{N_2}{N_1} = k$ for some constant, k.

(4) A. $\mathrm{E}(u_i|\mathbf{z}_i) = 0$ and $\mathrm{E}[\pi(\mathbf{z}_i)\pi(\mathbf{z}_i)'|\mathbf{z}_{1i} = \mathbf{z}_1]$ has a full rank for all $\mathbf{z}_1$, where $\pi(\mathbf{z}_i) = \mathrm{E}[\mathbf{x}_i'|\mathbf{z}_i]$.

B. $\mathrm{E}[\mathbf{x}_i^{(1)}|\mathbf{z}] = \mathrm{E}[\mathbf{x}_l^{(2)}|\mathbf{z}] = \mathrm{E}[\mathbf{x}|\mathbf{z}]$.

C. Finally, the density of $\mathbf{z}_1$ is identical for both samples and equal to $f(\mathbf{z}_1)$.

Assumptions 1–3 are essentially identical to those in Cai and Li (2008) except for modifications due to the presence of two samples. An important difference is that now we require the bandwidths used in the two samples to satisfy a condition that ensures that no extra terms appear in the asymptotic bias. Assumption 4 is necessary and sufficient for model identification – it makes sure NPGMM works in the two-sample setting. This assumption is similar to the one used by Angrist and Krueger (1992). Their two-sample IV estimator is a parametric two-sample estimator based on equality of unconditional expectations in the two samples while our nonparametric estimator is based on equality of conditional expectations in the context of the two-sample data structure.

A key part of Assumption 4 is that the conditional expectation function for $\mathbf{x}_i$ given $\mathbf{z}_i$ is the same for the two samples. If we could observe $\mathbf{x}_i$ in both samples then given a value of $\mathbf{z}_i = \mathbf{z}_i^{(1)} = \mathbf{z}_l^{(2)}$, the moment condition $\mathrm{E}[\mathbf{x}_i^{(1)}|\mathbf{z}_i^{(1)}] = \mathrm{E}[\mathbf{x}_l^{(2)}|\mathbf{z}_l^{(2)}] = \mathrm{E}[\mathbf{x}|\mathbf{z}]$ must hold. Under this assumption, sample equivalents of the quantities contained in moment conditions (6) have the same probability limit as the one-sample NPGMM estimator. Thus the probability limit of our estimator should be the same as that of Cai and Li (2008).

Next we provide a formal proof of this intuition. Specifically, we establish consistency and asymptotic normality for the TS-NPGMM estimator. For ease of reference, we adopt the following notation. Let $\mu_2(K) = \int \mathbf{u}\mathbf{u}' K(\mathbf{u})d\mathbf{u}$ and $\mu = \int K^2(\mathbf{u})d\mathbf{u}$. Define

$$\mathbf{R}_j = \frac{1}{N_j}\sum_{i=1}^{N_j} K_{h_j}(\mathbf{z}_{1i}^{(j)} - \mathbf{z}_1)\mathbf{Q}(\mathbf{z}_i^{(j)})\sum_{k=1}^{K} R_k(\mathbf{z}_{1i}^{(j)}, \mathbf{z}_1)x_{ik}^{(j)}, \tag{10}$$

where $R_k(\mathbf{z}_{1i}, \mathbf{z}_1) = b_k(\mathbf{z}_{1i}) - b_{k0} - \mathbf{b}_{k1}(\mathbf{z}_{1i} - \mathbf{z}_1) - \frac{1}{2}(\mathbf{z}_{1i} - \mathbf{z}_1)' \frac{\partial^2 b_k(\mathbf{z}_1)}{\partial \mathbf{z}_1^2}(\mathbf{z}_{1i} - \mathbf{z}_1)$,

$$\mathbf{B}_j = \frac{1}{N_j}\sum_{i=1}^{N_j} K_{h_j}(\mathbf{z}_{1i}^{(j)} - \mathbf{z}_1)\mathbf{Q}(\mathbf{z}_i^{(j)})\frac{1}{2}\sum_{k=1}^{K}(\mathbf{z}_{1i}^{(j)} - \mathbf{z}_1)' \frac{\partial^2 b_k(\mathbf{z}_1^{(j)})}{\partial \mathbf{z}_1^2}(\mathbf{z}_{1i}^{(j)} - \mathbf{z}_1)x_{ik}^{(j)}, \tag{11}$$

and

$$\mathbf{T}_j^* = \frac{1}{N_j}\sum_{i=1}^{N_j} K_{h_j}(\mathbf{z}_{1i}^{(j)} - \mathbf{z}_1)\mathbf{Q}(\mathbf{z}_i^{(j)})u_i^{(j)}, \tag{12}$$

where $j = 1, 2$. Denote the first-sample analogue of $\mathbf{S}_2$ by $\mathbf{S}_1$. Clearly, $\mathbf{S}_1$, $\mathbf{R}_1$, $\mathbf{B}_1$ and $\mathbf{T}_j^*$ are not feasible because $\mathbf{x}_i^{(1)}$, $y_i^{(2)}$, and $u_i^{(j)}$, $j = 1, 2$, are not observed. However, it turns out that for the Cai and Li (2008) asymptotic results to apply, we will need assumptions on these quantities.

Let $\mathbf{H}_j = \text{diag}\{\mathbf{I}_K, h_j \mathbf{I}_{KL_1}\}$, $j = 1, 2$, where $\mathbf{I}_m$ is a $m \times m$ identity matrix.

Theorem. Under Assumptions 1–4, we have

$$\sqrt{N_2 h_2^{L_1}}\left[\mathbf{H}_2(\hat{\beta} - \beta) - \frac{h_2^2}{2}\begin{pmatrix}\mathbf{B}_b(\mathbf{z}_1) \\ \mathbf{0}\end{pmatrix} + o_p(h_2^2)\right] \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{\Psi}), \tag{13}$$

where $\mathbf{\Psi} = f^{-2}(\mathbf{z}_1)(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{\Phi}\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}$ with $\mathbf{S} = \mathbf{S}(\mathbf{z}_1) = \text{diag}\{\mathbf{\Omega}, \mathbf{\Omega} \otimes \mu_2(K)\}$ and $\mathbf{\Phi}$ being the limiting covariance matrix of $\sqrt{N_2 h_2^{L_1}}(\mathbf{T}_1 - \mathbf{S}_2\beta)$. In addition, $\mathbf{B}_b(\mathbf{z}_1) = \int \mathbf{D}(\mathbf{u}, \mathbf{z}_1)K(\mathbf{u})d\mathbf{u} =$

$(\mathrm{tr}(\nabla^2 b_j(\mathbf{z}_1)\mu_2(K)))$, $\mathbf{D}(\mathbf{u},\mathbf{z}_1) = \begin{pmatrix} \mathbf{u}'\nabla^2 b_1(\mathbf{z}_1)\mathbf{u} \\ \vdots \\ \mathbf{u}'\nabla^2 b_K(\mathbf{z}_1)\mathbf{u} \end{pmatrix}$, and $\nabla^2 b_j(\mathbf{z}_1) = \frac{\partial^2 b_j(\mathbf{z}_1)}{\partial \mathbf{z}_1 \partial \mathbf{z}_1'}$.

Proof: see Appendix A.

In a parametric setting, Inoue and Solon (2010) show that the two-stage least squares version of the two-sample estimator of Angrist and Krueger (1992) is more efficient. Similarly, in our case, estimator (9) may not deliver the minimal asymptotic variance due to the suboptimal choice of $\mathbf{Q}(\mathbf{z})$. While finding the optimal instruments is feasible, we prefer the computationally simple form for $\mathbf{Q}(\mathbf{z})$ from Cai and Li (2008).

We assume $\mathbf{z}_1$ to have the same support in the two samples. Assumption 4 implies that the density of $\mathbf{z}_1$ for the two samples is identical and equal to $f(\mathbf{z}_1)$, which is an even stronger assumption than the same support. If the densities are different, e.g., if they differ in their support, it may cause problems for consistency and nonparametric identification of $\mathbf{b}(\mathbf{z}_1)$. We leave such cases to future work.

2.3 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to study the finite sample performance of the TS-NPGMM estimator. First, we graphically illustrate that the TS-NPGMM approach can successfully uncover the true functional form. Second, we present numerical results illustrating convergence at various sample sizes.

2.3.1 Graphical Illustrations

We consider the following data generating process:

$$Y_i = (0.5 + 0.25U_i^{c^2} + 0.5U_i^c) + (1 + e^{0.1U_i^c} + U_i^c)X_i + s\epsilon_i, \tag{14}$$

where $U_i^c \sim N(0,1)$ truncated at ± 2, $X_i = (Z_i + \tau\epsilon_i)/\sqrt{1+\tau^2}$, $\epsilon_i \sim N(0,1)$, and $(Z_i, \epsilon_i)' \sim N(0_{2\times 1}, I_2)$. This is one of the DGPs considered by Su, Murtazashvili, and Ullah (2013). Similarly, we use τ to control the degree of endogeneity and choose s to ensure the signal-noise ratio is 1 when we generate observations on Y_i.

$\{Y_i, U_i, Z_i\}_{i=1}^{N_1}$ and $\{X_j, U_j, Z_j\}_{j=1}^{N_2}$ are two independent samples drawn from a population, subject to (14). We do not observe X_i in the first sample and Y_j in the second sample. As a benchmark, we also consider a hypothetical setting where we can observe $\{X_k, Y_k, U_k, Z_k\}_{k=1}^{N}$ in one sample. We are interested in estimating the two functional coefficients: $g_1(u) = 0.5 + 0.25u^2 + 0.5u$ and $g_2(u) = 1 + e^{0.1u} + u$. For both the two and one-sample NPGMM estimators, we use the standardized Epanechnikov kernel $k(u) = \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}u^2)\mathbb{I}(|u| \leq \sqrt{5})$ for smoothing and the following simple rule of thumb for bandwidth: $h = s_U\, n^{-1/5}$, where s_U is the standard errors of U.

Figure 1 provides graphical representations of the two and one-sample NPGMM approaches based on $N_1 = N_2 = 3,000$ observations and 500 replications. We have two interesting observations from Figure 1. First, both of these estimators successfully recover the true functional coefficients. Second, the TS-NPGMM approach appears to have a slightly larger bias than the one from the one-sample NPGMM method. We conjecture that the bias may be caused by a violation of the first equality of Assumption 4B in finite samples.

2.3.2 Numerical Assessment

Next, we consider the behavior of TS-NPGMM and NPGMM as the sample size increases. We do so using a grid of $S = 25$ equally spaced points on the interval $[-2, 2]$. We evaluate the estimates of $g_1(u)$ and $g_2(u)$ on the grid and calculate the mean absolute deviation (MAD) and mean squared error (MSE) for each estimator as follows:

$$MAD_l = \frac{1}{S\,R}\sum_{r=1}^{R}\sum_{s=1}^{S}|\hat{g_l}^{(r)}(u_s) - g_l(u_s)|,$$
$$MSE_l = \frac{1}{S\,R}\sum_{r=1}^{R}\sum_{s=1}^{S}[\hat{g_l}^{(r)}(u_s) - g_l(u_s)]^2,$$

where $\hat{g_l}^{(r)}(u_s)$, $l = 1, 2$, is an estimate of $g_l(u_s)$ evaluated at grid point s in the r-th replication. This is done over $R = 500$ replications.

We consider nine sample sizes for $N_1 = N_2$. We also look at unequal sizes but the substantive results are unchanged. We report the corresponding MSE and MAD for TS-NPGMM and NPGMM in Table 1. Clearly, for both estimators, as the sample size increases, the MSE and MAD decrease quite quickly but remain substantially larger than their single-sample analogues. This is an interesting result that has to do with the relative bias and efficiency of the two-sample versus one-sample estimator and needs further exploration.

3 Data Description

Following Björklund and Jäntti (1997), we use the Panel Study of Income Dynamics (PSID) and the Swedish Level of Living Survey(SLLS) to obtain data for the United States and Sweden, respectively. Both, the PSID and SLLS are longitudinal (but not necessarily annual) surveys conducted since 1968.

Our US sample of individuals – we will refer to these individuals as fathers – taken from 1968 contains 1,613 male heads of household of age between 27 and 68 who had at least one child (daughter or son). This sample is obtained from the Survey Research Center (SRC) component of the PSID. The independent sample of the US individuals – we will refer to these individuals as sons – taken from the 1988 SRC contains 467 individuals. Sons are restricted to those individuals who were born between 1951 and 1959 and who were the oldest sons from multiple-son families. Only those fathers and sons who reported positive annual earnings for 1967 and 1987, respectively, are included into the samples. The US fathers report their 1967 annual income, education and occupation, while the US sons report their annual income in 1987, as well as occupation and education of their actual fathers.

Our Swedish sample of fathers (called so similarly to our US sample of fathers) is taken from 1968 wave of the SLLS and contains 565 individuals who are either native Swedes or moved to Sweden before the age of 16. The independent sample of Swedish sons (called so similarly to our US sample of sons) is taken from the 1991 wave of the SLLS and contains 324 individuals who were born between 1952 and 1961. The Swedish fathers report their 1967 annual income, occupation and the highest education level attained. The Swedish sons provide information on their 1990 annual income, as well as their actual fathers' education and occupation.

Panel A of Table 2 presents summary statistics for the fathers and sons from the US samples, while Panel B of Table 2 reports those for the Swedish samples. Table 3 presents more detailed information on education and occupation of fathers. Here, we consider both types of fathers present in our analysis: (1) individuals observed directly in the samples of fathers (we can think of them as pseudo-fathers), and (2) individuals who were not directly observed but described by their sons in the samples of sons (we can think of them as actual fathers). As Table 3 suggests, the two distributions of the fathers' educational and occupational characteristics appear to be reasonably consistent.

4 Intergenerational Income Mobility in the US and Sweden

In this section, we apply our TS-NPGMM approach developed in Section 2 to estimation of income mobility across generations in the USA and Sweden while allowing intergenerational income mobility to vary across the distribution of families. The last (but not the least) question we consider before applying our TS-NPGMM approach to estimation of intergenerational income mobility is what should be used as exogenous variables, $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2)$, in our model (2).

4.1 Our Exogenous Variables

The existing literature on intergenerational mobility has long emphasized that the simplicity of the empirical model presented by equation (1) should not be taken at the face value. Solon (1999) points out that, though seemingly simple, model (1) is still capable of showing that "intergenerational transmission occurs through a multitude of processes."[6] A large empirical body of research maintains that the child's earnings are likely to depend on other aspects of family background, and, thus, this strand of the literature incorporates factors other than parents' earnings into the intergenerational income equation to account for the influence of those factors on intergenerational mobility. Specifically, father's education, occupation, race, union status, industry and country of residence are a few fathers' characteristics that have been argued to affect intergenerational mobility.[7]

Furthermore, studies that use IV methods often employ fathers' characteristics other than income as instrumental variables for fathers' income. In particular, Lefranc and Trannoy (2005) use fathers' education, occupation, and indicators for living in Paris and in a rural area as IVs when

[6] Solon (1999), page 1765.
[7] See Black and Devereux (2011) for a recent survey of relevant studies.

studying intergenerational earnings mobility in France. Solon (1992) uses fathers' education to instrument fathers' income when estimating intergenerational income mobility in the USA. Björklund and Jäntti (1997) also employ fathers' education but add fathers' occupation dummies as instruments for fathers' income to compare intergenerational income mobility in the United States and Sweden. Studying the USA, Zimmerman (1992) uses the Duncan index of the prestige of fathers' occupation as an instrument for fathers' earnings.

We follow Björklund and Jäntti (1997) and use fathers' education and fathers' occupation dummies as a set of exogenous variables. Our choice of the exogenous variables is driven in part by data availability. In fact, fathers' education and occupation are the only two characteristics of fathers available in our data set in addition to fathers' earnings. We use fathers' education as z_1 and fathers' occupation as z_2.[8]

While our choice of the exogenous variables fits comfortably in the existing literature on intergenerational mobility, we cannot but mention that, when used as instrumental variables, fathers' occupation might also have an independent direct effect on sons' income. However, there is a wide range of literature claiming that it is unlikely to be so. In particular, studies by Sewell and Hauser (1975), Kiker and Condon (1981), Datcher (1982), Corcoran, Gordon, Laren, and Solon (1992), Checchi, Ichino, and Rustichini (1999), and Lefranc and Trannoy (2005) are a small sample of studies that maintain that fathers' occupational status is correlated with sons' earnings only through its correlation with fathers' income. Thus, we follow this wide literature and employ a full set of fathers' occupational dummies as instrumental variables for fathers' income in our analysis.[9]

[8] We could also employ fathers' education as z_2 and fathers' occupation as z_1. However, we believe the case when fathers' education is chosen as z_1 is more interesting and relevant, as occupational status can vary substantially across countries. In fact, Table 3 suggests that, while similar, occupational categories used in the US and Swedish surveys are not identical. To have a better chance of conducting a direct cross-country companion we choose fathers' education as z_1 and fathers' occupation as z_2.

[9] Solon (1992) shows that if the instrumental variable used has a direct effect on sons' income then, under the typical assumptions used by labor economists, the traditional IV approach will produce an upper bound on the intergenerational income elasticity.

4.2 Our Empirical Findings

We use the second-order Epanechnikov kernel in our estimation, and devise a data-driven bandwidth selection method based on matching and cross-validation, which takes account of our data structure – we describe the method in Appendix B. Table 4 reports the TS-NPGMM estimates of the intergenerational income elasticity as a function of fathers' education for some values of fathers' educational attainment measured in years for both the USA and Sweden. In addition, Table 4 reports the standard errors of the TS-NPGMM estimates. The residuals used for calculating the standard errors are based on the surrogate sample obtained by matching as described in Appendix B.

Our empirical results indicate that for the median levels of fathers' education, which are 12 and 6 years of education in our US and Swedish samples, respectively, the corresponding intergenerational income elasticities are about 0.42 and 0.23. Our TS-NPGMM estimates of intergenerational income mobility at the median levels of fathers' education are similar to those reported by Björklund and Jäntti (1997), which are 0.42 and 0.28, respectively.

Figure 2 presents a visual summary of the results reported in Table 4 extending them to the entire range of fathers' educational attainment. Figure 2 clearly suggests that the shapes of the US and Swedish intergenerational income mobilities are rather distinct. The intergenerational income elasticity in Sweden increases in fathers' education for those families whose fathers have at least some high school education. A minor exception to that observation arises over roughly a one-year interval between 12 and 13 years of fathers' education. Further, we observe a decrease in intergenerational elasticity for families with fathers' education below about 9 years. In fact, the lowest intergenerational elasticity in Sweden is observed in those families whose fathers' education is about 9 years. The US intergenerational elasticity has a much more dramatic shape than that in Sweden. The maximum intergenerational elasticity is observed for families where fathers have education of about 11 years. The two local minimums of intergenerational elasticities in the USA are achieved in those families whose fathers have education of about 8 and 14 years.

More importantly, when comparing the two parts of Figure 2, we note that it is not only the elasticity at the median level of fathers' education that is lower in Sweden than in the USA. The Swedish intergenerational elasticity is lower than that in the USA at almost all levels of fathers' education (except for very high educational achievements of fathers). Part A of Figure 2 clearly shows a nonlinear functional relationship between intergenerational income mobility and fathers' educational attainment for the USA. According to Part B of Figure 2, while also nonlinear, this relation is much closer to being linear for Sweden than for the USA especially when only levels of fathers' education above 9 years are taken into consideration. Furthermore, we have no grounds for serious concerns about the least mobile subpopulation of the Swedish society characterized by highly educated fathers as long as we assume that better educated individuals, on average, earn more than less educated persons. On the contrary, in the USA, the least mobile individuals are the ones whose fathers have less than a high school degree.

To summarize, our empirical findings suggest that family background measured by fathers' education matters less for Sweden than for the US – the range of the elasticities we estimate is lower for Sweden. Interestingly, this is exactly the result from the previous studies summarized, for example, by Björklund, Jäntti, and Lindquist (2009) who write that "family background is more important for labor market achievement in the United States than in most other rich countries."[10] Moreover, our empirical results clearly indicate that, overall, the Swedish society is more mobile than the US one across (almost) the entire distribution of families.

[10] Björklund, Jäntti, and Lindquist (2009), page 671.

5 Conclusion

Our study focuses on measuring intergenerational income mobility in Sweden and the United States to assess the character of the inequality in these countries. Our choice of the USA and Sweden is driven by a previous finding that the Scandinavian countries have the lowest annual income inequality in contrast to the US that is among the countries with the highest inequality.

We propose a new nonparametric approach to address the three main concerns arising when conducting cross-country comparisons in intergeneration income mobility – constancy of intergenerational income mobility across the distribution of families, measurement error in fathers' log-run economic status, and several sources of data used in analysis. First, we allow for a flexible nonparametric functional dependence between intergenerational income mobility and observable family background characteristics (represented by fathers' education in our analysis). Second, we exploit an instrumental variable approach to account for measurement error in fathers' permanent income (using fathers' occupation as an instrument for fathers' income). Third, we extend a two-sample parametric approach, proposed by Angrist and Krueger (1992), to deal with the fact that fathers' and sons' incomes come from two different samples.

We develop the theoretical foundation for our two-sample nonparametric GMM estimator and provide a small Monte Carlo study illustrating its finite sample behavior. Further, we apply our approach to estimation of intergenerational income mobility in the USA and Sweden. We find that the character of inequality in the two countries in rather different. Similar to previous studies, we find that family background matters much more for labor market achievements in the US society than in the Swedish one. More importably, we report the exact shapes of intergenerational income mobility in the two countries, which is of high relevance for any policy discussions.

Appendix

A Proof of Theorem

First, notice $\hat{\beta}-\beta = (\mathbf{S}_2'\mathbf{S}_2)^{-1}\mathbf{S}_2'(\mathbf{T}_1-\mathbf{S}_2\beta)$. Then, $\mathbf{H}_2\hat{\beta} = \mathbf{H}_2(\mathbf{S}_2'\mathbf{S}_2)^{-1}\mathbf{H}_2\mathbf{H}_2^{-1}\mathbf{S}_2'\mathbf{T}_1 = (\tilde{\mathbf{S}}_2'\tilde{\mathbf{S}}_2)^{-1}\tilde{\mathbf{S}}_2'\mathbf{T}_1$, where $\tilde{\mathbf{S}}_2 = \mathbf{S}_2\mathbf{H}_2^{-1}$, and we can express $\mathbf{T}_1$ as $\mathbf{T}_1 = \mathbf{S}_1\beta + \mathbf{R}_1 + \mathbf{B}_1 + \mathbf{T}_1^*$. Then, $\mathbf{T}_1 - \mathbf{S}_2\beta = (\mathbf{S}_1 - \mathbf{S}_2)\beta + \mathbf{R}_1 + \mathbf{B}_1 + \mathbf{T}_1^*$, and we can write

$$\mathbf{H}_2(\hat{\beta} - \beta) - (\tilde{\mathbf{S}}_2'\tilde{\mathbf{S}}_2)^{-1}\tilde{\mathbf{S}}_2'[(\tilde{\mathbf{S}}_1\mathbf{H}_1 - \tilde{\mathbf{S}}_2\mathbf{H}_2)\beta + \mathbf{R}_1 + \mathbf{B}_1] = (\tilde{\mathbf{S}}_2'\tilde{\mathbf{S}}_2)^{-1}\tilde{\mathbf{S}}_2'\mathbf{T}_1^*. \tag{15}$$

The proof of Theorem 2 from Cai and Li (2008) shows that $\sqrt{N_j h_j^{L_1}}\mathbf{T}_j^*$, where $j = 1, 2$, is asymptotically normal with zero mean and finite variance.Further, by Proposition 1 of Cai and Li (2008), $\mathbf{B}_j = O_p(h_j^2)$ and $\mathbf{R}_j = o_p(h_j^2)$, where $j = 1, 2$. These results imply that the last term on the left-hand side of (15), which contains $\mathbf{B}_1$, contributes to the asymptotic bias, while the term containing $\mathbf{R}_1$ is negligible in probability. Condition $h_1/h_2 \to 1$ of Assumption 3 assures that $\mathbf{R}_1 = o_p(h_2^2)$ and $\mathbf{B}_1 = O_p(h_2^2)$. Then, to establish consistency of (9), we are left with determining the behavior of $(\tilde{\mathbf{S}}_1\mathbf{H}_1 - \tilde{\mathbf{S}}_2\mathbf{H}_2)$. Notice that, by Proposition 1 of Cai and Li (2008), $\tilde{\mathbf{S}}_1 - \tilde{\mathbf{S}}_2 = o_p(1)$ and

$$\sqrt{N_2 h_2^{L_1}}(\tilde{\mathbf{S}}_1\mathbf{H}_1 - \tilde{\mathbf{S}}_2\mathbf{H}_2) =$$

$$\sqrt{N_2 h_2^{L_1}}(\tilde{\mathbf{S}}_1 - f(\mathbf{z}_1)\mathbf{S})\mathbf{H}_1 - \sqrt{N_2 h_2^{L_1}}(\tilde{\mathbf{S}}_2 - f(\mathbf{z}_1)\mathbf{S})\mathbf{H}_2 - \sqrt{N_2 h_2^{L_1}}f(\mathbf{z}_1)\mathbf{S}(\mathbf{H}_2 - \mathbf{H}_1), \tag{16}$$

where $\tilde{\mathbf{S}}_j \to f(\mathbf{z}_1)\mathbf{S}$ for both samples due to Proposition 1 of Cai and Li (2008) and Assumption 4. Condition $h_2/h_1 \to 1$ of Assumption 3 guarantees that the last term of (16) is negligible in probability. Condition $\lim_{N_2\to\infty} \frac{N_2}{N_1} = k$ of Assumption 3 allows to rewrite the first term of (16) as $\sqrt{k}\sqrt{N_1 h_1^{L_1}}(\tilde{\mathbf{S}}_1 - f(\mathbf{z}_1)\mathbf{S})\mathbf{H}_1$. Then, the first two terms of (16) are also negligible in probability due to Proposition 1 of Cai and Li (2008). Thus, the consistency of 2S-NPGMM is established, and the order of the bias term in expression (15) is h_2^2.

Second, observe that $\sqrt{N_2 h_2^{L_1}}\mathbf{H}_2(\hat{\beta}-\beta) = \sqrt{N_2 h_2^{L_1}}(\tilde{\mathbf{S}}_2'\tilde{\mathbf{S}}_2)^{-1}\tilde{\mathbf{S}}_2'(\mathbf{T}_1 - \mathbf{S}_2\beta)$. Then,

$$\sqrt{N_2 h_2^{L_1}}\left[\mathbf{H}_2(\hat{\beta}-\beta) - \frac{h_2^2}{2}\begin{pmatrix}\mathbf{B}_b(\mathbf{z}_1)\\ \mathbf{0}\end{pmatrix} + o_p(h_2^2)\right] \xrightarrow{d} \mathrm{N}(\mathbf{0}, (\tilde{\mathbf{S}}_2'\tilde{\mathbf{S}}_2)^{-1}\tilde{\mathbf{S}}_2'\boldsymbol{\Phi}\tilde{\mathbf{S}}_{\mathbf{2}}(\tilde{\mathbf{S}}_2'\tilde{\mathbf{S}}_2)^{-1}), \qquad (17)$$

where $\boldsymbol{\Phi}$ is the limiting covariance matrix of $\sqrt{N_2 h_2^{L_1}}(\mathbf{T}_1 - \mathbf{S}_2\beta)$. Using Proposition 1 from Cai and Li (2008) the asymptotic variance of the left-hand side of (17) becomes $f^{-2}(\mathbf{z}_1)(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\boldsymbol{\Phi}\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}$, where $\mathbf{S} = \mathbf{S}(\mathbf{z}_1) = \mathrm{diag}\{\boldsymbol{\Omega}, \boldsymbol{\Omega} \otimes \mu_2(K)\}$. *QED.*

B Bandwidth Selection

Bandwidth selection is not straightforward for two-sample models since the data for y_i and $\mathbf{x}_i$ in such models are not contained in the same sample. Therefore, the calculation of the residuals is infeasible and such standard data-driven tools of residual-based bandwidth selection as least squares cross-validation cannot be applied directly. Instead, we propose using matching in the following three-step bandwidth selection rule:

Step 1: We match $\mathbf{x}_j^{(2)}$ with $y_i^{(1)}$ by matching $\mathbf{z}_j^{(2)}$ to $\mathbf{z}_i^{(1)}$. That is, for a given value of $\mathbf{z}_j^{(2)}$, we look for such i that $\mathbf{z}_i^{(1)} = \mathbf{z}_j^{(2)}$. The $y_i^{(1)}$ corresponding to that $\mathbf{z}_i^{(1)}$ is the value of y matched to $\mathbf{x}_j^{(2)}$. Specifically, in our empirical analysis, we match sons' income with fathers' income by choosing equal values of fathers' education, reported in both samples. Not surprisingly, there are more than one such matching observations for any value of fathers' education. That is, for each value $\mathbf{x}_j^{(2)}$ we have several matched values of $y_i^{(1)}$.

Step 2: We take an average of the subsample of $y_i^{(1)}$ matched to $\mathbf{x}_j^{(2)}$. This produces a single value of the matched y – we denote it by $\bar{y}_j^{(2)}$ – and a surrogate full sample $\{\bar{y}_j^{(2)}, \mathbf{x}_j^{(2)}, \mathbf{z}_j^{(2)}\}$.

Step 3: We apply the standard leave-one-out cross-validation technique to the full sample $\{\bar{y}_j^{(2)}, \mathbf{x}_j^{(2)}, \mathbf{z}_j^{(2)}\}$.

The bandwidth selection rule we propose is based on Assumption 4B. Under Assumption 4B, $E(y_i^{(1)}|\mathbf{z}) = E(y_j^{(2)}|\mathbf{z})$. Thus, given $\mathbf{z}$, we expect to observe the same values of y in both samples. In essence, we estimate the value of y for the sample that does not contain it with the average of the matched values of y from the sample where y is actually observed. Once we have the full sample, the selection rule becomes a standard cross-validation exercise.

Table 1: Simulation Results for TS-NPGMM and NPGMM

Sample Size	TS-NPGMM				NPGMM			
	$\hat{g}_1(u)$		$\hat{g}_2(u)$		$\hat{g}_1(u)$		$\hat{g}_2(u)$	
	MAD	MSE	MAD	MSE	MAD	MSE	MAD	MSE
500	0.445	0.684	0.680	2.275	0.217	0.092	0.229	0.103
800	0.349	0.318	0.496	0.789	0.174	0.057	0.178	0.061
1,000	0.320	0.273	0.441	0.617	0.159	0.046	0.164	0.052
1,500	0.262	0.177	0.363	0.396	0.134	0.033	0.134	0.034
2,000	0.227	0.126	0.323	0.308	0.116	0.025	0.120	0.027
2,500	0.203	0.099	0.287	0.231	0.106	0.020	0.109	0.022
3,000	0.194	0.092	0.271	0.203	0.098	0.017	0.101	0.018
3,500	0.173	0.067	0.245	0.156	0.091	0.015	0.094	0.016
4,000	0.166	0.064	0.236	0.147	0.087	0.013	0.089	0.014

Table 2: Summary Characteristics of the US and Swedish Samples

Variables	Mean	Stand Deviation	Min	Max
Panel A: US samples[1]				
Father's age in 1967	45.1	10.9	27	68
Father's earnings in 1967[2]	28,311	19,432	442	222,744
Father's log earnings in 1967	10.03	0.74	6.09	12.31
Father's education	11.45	4.11	0	18
Son's age in 1987	32.4	2.4	28	36
Son's earnings in 1987	28,598	19,352	1,200	210,000
Son's log earnings in 1987	10.06	0.67	7.09	12.25
Reported father's education	11.7	3.4	0	18
Panel B: Swedish samples[3]				
Father's age in 1967	42.8	7.6	25	60
Father's earnings in 1967[4]	16,821	8,903	5,678	76,687
Father's log earnings in 1967	9.6	0.4	8.6	11.2
Father's education	8.1	2.9	6	16
Son's age in 1990	34.4	3.1	30	39
Son's earnings in 1990	1,797	637	550	5,550
Son's log earnings in 1990	7.4	0.3	6.3	8.6
Reported father's education	8.1	3.1	6	16

[1] Panel A is based on 467 sons and 1,613 fathers.
[2] Father's 1967 earnings are in 1987 dollars.
[3] Panel B is based on 324 sons and 565 fathers.
[4] Father's 1967 earnings are in 1990 Swedish krona.

Table 3: Education and Occupation Characteristics of Fathers

	Fathers' own	Sons'
	report of fathers' characteristics	
Panel A: US Samples[1]		
Fraction with education higher than compulsory	0.94	0.97
Fraction with given occupation:		
1 Professional, technical and kindred workers	0.16	0.15
2 Managers, officials and proprietors	0.11	0.08
3 Self-employed businessmen	0.07	0.03
4 Clerical and sales workers	0.11	0.11
5 Craftsmen, foremen, and kindred workers	0.23	0.25
6 Operatives and kindred workers	0.17	0.19
7 Laborers and service workers, farm laborers	0.09	0.08
8 Farmers and farm managers	0.04	0.09
Panel B: Swedish Samples[2]		
Fraction with education higher than compulsory	0.62	0.63
Fraction with given occupation:		
1 Higher-grade professional	0.09	0.11
2 Lower-grade professional	0.12	0.08
3 Non-manual workers and lower-grade technicians	0.14	0.15
4 Small proprietors with employees	0.06	0.07
5 Small proprietors without employees	0.04	0.05
6 Farmers, self-employed in primary agricultural production and other workers	0.11	0.13
7 Skilled manual workers	0.22	0.18
8 Semi-skilled manual workers	0.17	0.20

[1] Panel A is based on 467 sons and 1,613 fathers.
[2] Panel B is based on 324 sons and 565 fathers.

Table 4: TS-NPGMM Estimates of Intergenerational Income Mobility as a Function of Fathers' Education

Fathers' Education	Income Elasticity	Standard Error
Panel A: US Samples		
7	0.304	0.021
10	0.371	0.026
12	0.422	0.024
14	0.264	0.026
15	0.290	0.027
16	0.321	0.023
18	0.336	0.016
Panel B: Swedish Samples		
6	0.227	0.026
9	0.145	0.026
10	0.207	0.027
12	0.296	0.023
16	0.478	0.014

Notes: Panel A is based on 467 sons and 1,613 fathers. Panel B is based on 324 sons and 565 fathers. Fathers' income is instrumented with a full set of fathers' occupational dummies.

Figure 1: Monte Carlo Simulations for $g_1(u)$ and $g_2(u)$ with 500 replications

(a) TS-NPGMM ($N_1 = 3,000$, $N_2 = 3,000$) (b) NPGMM (sample size N= 3,000)

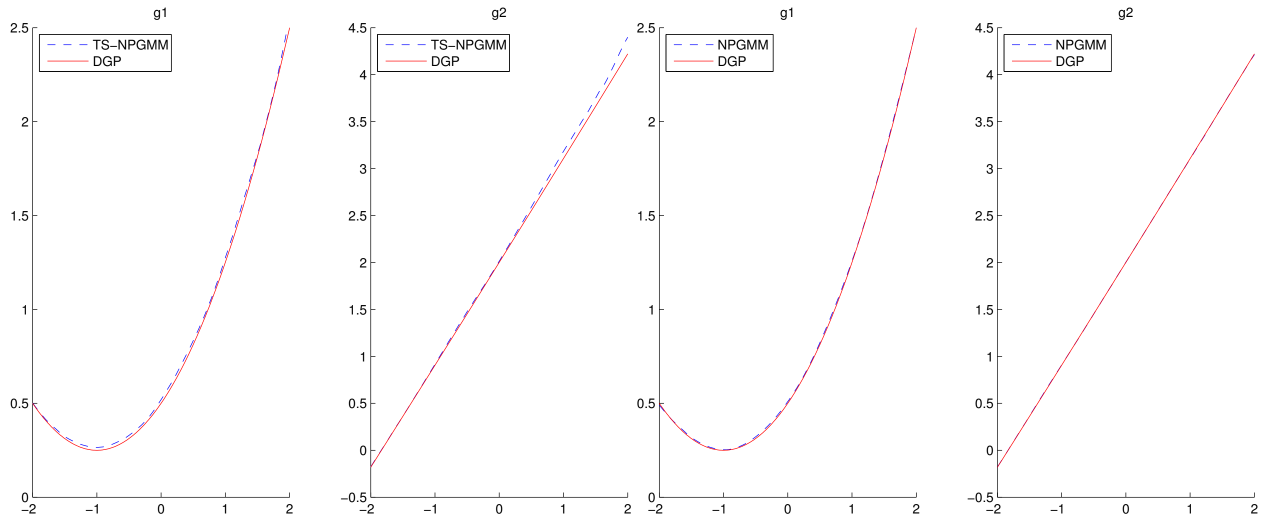
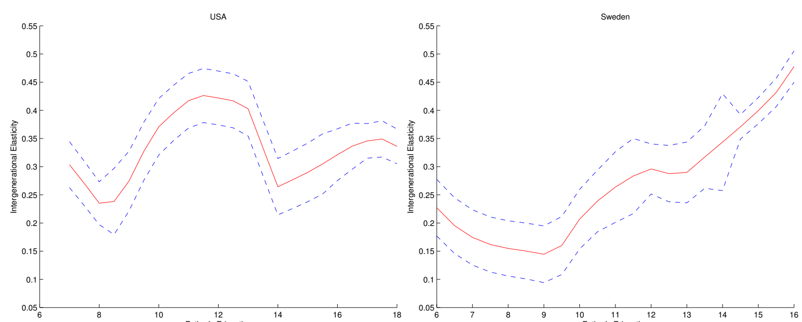


Figure 2: TS-NPGMM estimates of income elasticity as a function of fathers' education

(a) US (b) Sweden



Notes: Horizontal Axis - father's education. Vertical Axis - intergenerational income elasticity. TS-NPGMM estimate, solid line; 95% confidence interval, dashed line.

References

Aaberge, R., A. Björklund, M. Jäntti, M. Palme, P. Pedersen, N. Smith, and T. Wennemo (2002): "Income Inequality and Income Mobility in the Scandinavian Countries Compared to the United States," *Review of Income and Wealth*, 48, 443–469.

Angrist, J. D., and A. B. Krueger (1992): "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples," *Journal of the American Statistical Association*, 87(418), 328–336.

Arellano, M., and C. Meghir (1992): "Female Labour Supply and On-the-Job Search: An Empirical Model Estimated Using Complementary Data Sets," *The Review of Economic Studies*, 59(3), 537–559.

Bhattacharya, D., and B. Mazumder (2011): "A Nonparametric Analysis of Black-White Differences in Intergenerational Income Mobility in the United States," *Quantitative Economics*, 2, 335–379.

Björklund, A., and M. Jäntti (1997): "Intergenerational Income Mobility in Sweden Compared to the United States," *The American Economic Review*, 87(5), 1009–1018.

Björklund, A., M. Jäntti, and M. Lindquist (2009): "Family Background and Income DUring the Rise of the Welfare State: Brother Correlations in Income for Swedish Men Born 1932-1968," *Journal of Public Economics*, 93, 671–680.

Black, S. E., and P. J. Devereux (2011): "Recent Developments in Intergenerational Mobility," in *Handbook of Labor Economics*, vol. 4B, pp. 1487–1541. Elsevier.

Bratsberg, B., K. Røed, O. Raaum, R. Naylor, M. Jäntti, T. Eriksson, and E. Österbacka (2007): "Nonlinearities in Intergenerational Earnings Mobility: Consequences for Cross-Country Comparisons," *The Economic Journal*, 117(519), C72–C92.

Cai, Z., M. Das, H. Xiong, and X. Wu (2006): "Functional Coefficient Instrumental Variables Models," *Journal of Econometrics*, 133(1), 207–241.

Cai, Z., and Q. Li (2008): "Nonparametric Estimation of Varying Coefficient Dynamic Panel Data Models," *Econometric Theory*, 24(05), 1321–1342.

Checchi, D., A. Ichino, and A. Rustichini (1999): "More Equal but Less Mobile? Education Financing and Intergenerational Mobility in Italy and in the US," *Journal of Public Economics*, 74, 351–393.

Corcoran, M., R. Gordon, D. Laren, and G. Solon (1992): "The Association between Men's Economic Status and Their Family and Community Origins," *Journal of Human Resources*, 27, 575–601.

Datcher, L. P. (1982): "Effects of Community and Family Background on Achievement," *The Review of Economics and Statistics*, 64, 32–41.

Fan, J., and I. Gijbels (1996): *Local Polynomial Modeling and its Applications*. Chapman and Hall, London.

Gottschalk, P., and T. M. Smeeding (1997): "Cross-National Comparisons of Earnings and Income Inequality," *Journal of Economic Literature*, 32, 633–686.

Gustafsson, B. (1994): "The Degree and Pattern of Income Immobility in Sweden," *Review of Income and Wealth*, 40, 67–86.

Hong, Y., and T.-H. Lee (2003): "Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models," *Review of Economics and Statistics*, 85(4), 1048–1062.

Inoue, A., and G. Solon (2010): "Two-Sample Instrumental Variables Estimators," *Review of Economics and Statistics*, 92(3), 557–561.

Juhl, T. (2005): "Functional-coefficient Models under Unit Root Behaviour," *The Econometrics Journal*, 8(2), 197–213.

Lefranc, A., and A. Trannoy (2005): "Intergenerational Earnings Mobility in France: Is France more Mobile than the US?," *Annales d'Economie et de Statistiques*, 78, 57–75.

Murtazashvili, I. (2012): "An Alternative Measure of Intergenerational Income Mobility Based on a Random Coefficient Model," *Journal of Applied Econometrics*, 27, 1161–1173.

Sewell, W. H., and R. M. Hauser (1975): *Education, Occupation, and Earnings: Achievement in the Early Career*. Academic Press, New York.

Solon, G. (1992): "Intergenerational Income Mobility in the United States," *The American Economic Review*, 82(3), 393–408.

——— (1999): "Intergenerational Mobility in the Labor Market," in *Handbook of Labor Economics*, vol. Volume 3, Part A, pp. 1761–1800. North Holland.

——— (2002): "Cross-Country Differences in Intergenerational Earnings Mobility," *The Journal of Economic Perspectives*, 16, 59–66.

Su, L., I. Murtazashvili, and A. Ullah (2013): "Local Linear GMM Estimation of Functional Coefficient IV Models with an Application to Estimating the Rate of Return to Schooling," *Journal of Business and Economic Statistics*, forthcoming.

Zhang, W., S.-Y. Lee, and X. Song (2002): "Local Polynomial Fitting in Semivarying Coefficient Model," *Journal of Multivariate Analysis*, 82(1), 166–188.

Zimmerman, D. (1992): "Regression toward Mediocrity in Economic Stature," *American Economic Review*, 82, 409–429.