



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Econometrics 120 (2004) 181–205

JOURNAL OF
Econometrics

www.elsevier.com/locate/econbase

Mobility measurement, transition matrices and statistical inference

John P. Formby^a, W. James Smith^b, Buhong Zheng^{b,*}

^a*Department of Economics, Finance and Legal Studies, University of Alabama, Tuscaloosa, AL, USA*

^b*Department of Economics, University of Colorado at Denver, Campus Box 181, P.O. Box 173364, Denver, CO 80217-3364, USA*

Accepted 30 June 2003

Abstract

This paper develops statistical inference procedures for testing income mobility with transition matrices. Both summary mobility measures and partial mobility orderings are considered. We first examine the different ways that transition matrices are constructed in the literature on mobility measurement. Different approaches lead not only to distinct interpretations of mobility but also to different sampling distributions. The large sample properties of the estimates of transition matrices allow us to derive testing procedures for both summary mobility measures and partial orders of mobility across income regimes. The tests are illustrated by applying them to income mobility in the U.S. and Germany using the Panel Study of Income Dynamics and German Socio-Economic Panel data.

© 2003 Elsevier B.V. All rights reserved.

JEL classification: C40; D31; D60

Keywords: Mobility measurement; Partial mobility orderings; Transition matrix; Statistical inference; Asymptotic distribution

1. Introduction

The concepts and measurement of income distribution and income mobility focus on two quite distinct attributes of an income-generating regime. The former is concerned with how incomes are distributed among individuals over a given period of time, the latter with how individuals' incomes change over time. Economists have long recognized that relying on the size distribution of income alone is insufficient to describe the well-being of a society. It is widely agreed that income mobility must also be weighed

* Corresponding author. Tel.: +1-303-556-4413; fax: +1-303-556-3547.

E-mail address: buhong.zheng@cudenver.edu (B. Zheng).

when comparing income-generating regimes in different societies. Kuznets (1966, p. 203), for example, argues that two societies with identical size distributions of income “... may differ greatly in meaning because of different degrees of internal mobility,” and the society with more mobility enjoys greater social welfare. Rosen (1985, p. 79) goes further to argue that if there is sufficient income mobility, one need not be overly concerned about how unequally incomes are distributed. Thus, an income regime with a higher level of income inequality may well be preferred because it can lay claim to greater income mobility.

Prais (1955) is credited with proposing the first measure of income mobility (although interest in the quantification of mobility goes back much earlier, e.g., Ginsberg, 1929). The literature on mobility is now substantial with a significant number of summary measures of mobility proposed.¹ Yet, income mobility is much less clearly defined than income inequality. As a consequence, there is less consensus on the measurement of mobility than on the measurement of inequality. Some researchers (e.g., Prais, 1955; Shorrocks, 1978a; Sommers and Conlisk, 1979) view mobility as a reranking phenomenon, in which individuals switch income positions. In this approach, mobility is a purely relative concept. In the view of other researchers (e.g., Fields and Ok, 1996, 1999), mobility arises as soon as individuals move away from their initial income levels. In this approach, mobility is best characterized as an absolute concept. Most researchers measure mobility by simply invoking a specific definition of mobility. Alternatively, Atkinson (1983), Shorrocks (1978b), Chakravarty et al. (1985), Dardanoni (1993), and Formby et al. (2003) argue that the measurement of mobility should be undertaken by first exploring mobility’s implications for social welfare.

Despite the lack of agreement on the meaning of mobility, mobility measures have been increasingly applied to empirical data to describe income dynamics. The surge in applications has been facilitated by the increasing availability of panel data, that are, of course, necessary for any systematic empirical study of mobility. Panel data are random samples, and it is important to employ statistical inference when using these samples. This importance has long been recognized by researchers. For example, in the pathfinding paper that initiated the study of mobility measurement, Prais (1955, p. 63) expressed the need to assess “... the statistical errors in the estimation of the transition matrix” and believed that such a practice was “rather important.” Despite the importance placed by Prais, it is fair to say that the use of statistical inference of mobility has been largely neglected. Only recently have serious attempts to address this issue been mounted. Several researchers (e.g., Trede, 1999; Schluter, 1998; Maasoumi and Trede, 2001; Biewen, 2002) have taken up this issue and begun to devise statistical inference procedures for the measurement of mobility. These newly proposed procedures test mobility measures based upon transition matrices as well as the inequality-reducing mobility measures proposed by Shorrocks (1978b) and Maasoumi and Zandvakili (1986).

The purpose of this paper is to further develop statistical inference procedures for testing mobility measures with transition matrices. We depart from the existing literature in two respects.

¹ See Bartholomew (1996), Maasoumi (1997), and Fields and Ok (1999) for a review of these measures.

First, we consider the various ways that a transition matrix can be defined and derive the large sample properties of the estimates of each of these matrices. A transition matrix documents the movement of individuals across different income classes or occupational categories. The income classes may be defined for different generations (intergenerational mobility) or within the same generation (intragenerational mobility) with class boundaries predefined or endogenously determined from the data. All previous inference results on mobility apply to occupational mobility where the boundaries are naturally defined, or to income mobility where boundaries are assumed to be exogenously determined. We demonstrate that the asymptotic variance–covariance structures of transition matrices corresponding to different assumptions are quite different from one other.

Second, we test partial mobility orderings. The earlier literature on mobility measurement largely focuses on summary mobility measures, while the more recent literature primarily addresses partial ordering conditions. This evolution mirrors similar developments in income inequality and poverty measurement. The literature of partial mobility orderings argues that income mobility is a multi-dimensional concept, and, as a result, no single measure can capture all of its characteristics. Thus, instead of seeking summary measures, researchers have derived dominance conditions similar to Lorenz curves in the measurement of income inequality. Partial mobility orderings can be used to draw much broader conclusions than a single or even several summary measures but may be unable to rank order all income-generating regimes. Virtually, all studies that seek to derive welfare implications from mobility analysis rely upon the partial ordering approach to rank income-generating regimes.

The remainder of the paper is organized as follows. Section 2 briefly reviews the literature on mobility measurement with particular emphasis on transition matrices. Both summary measures and partial mobility orderings are examined. Section 3 discusses the different ways a transition matrix can be constructed and estimated, and derives the large sample properties for the estimates of different transition matrices. Section 4 establishes inference procedures for testing summary mobility measures and partial mobility orderings. Section 5 illustrates the inference procedures by testing for differences in income mobility in the U.S. and Germany between 1985 and 1990. Section 6 concludes.

2. Mobility measures and partial mobility orderings

Consider a joint distribution between two income variables $x \in [0, \infty)$ and $y \in [0, \infty)$ with a continuous *c.d.f.* $K(x, y)$. Clearly, the function $K(x, y)$ completely captures the movement between x and y . This movement may be intergenerational if x is, say, a father's income and y is his son's income; it is intragenerational if x and y are the same individual's income at two points in time. For ease of reference, unless otherwise stated, we consider intragenerational mobility between two points in time in the remainder of the paper. The marginal distributions of x and y are denoted as $F(x)$ and $G(y)$, i.e., $F(x) \equiv K(x, \infty)$ and $G(y) \equiv K(\infty, y)$. For convenience, we also

assume that functions F , G and K are strictly monotone, and the first two moments of x and y exist and are finite.

In the mobility measurement literature, the movement between x and y is described by a transition matrix, which is a transformation from a continuous *c.d.f.* of an income regime. To form such a transition matrix from $K(x, y)$, one first needs to determine the number of and boundaries between income classes. Suppose, there are m classes in each income distribution and the boundaries of these classes are, respectively, $0 < \zeta_1 < \zeta_2 < \dots < \zeta_{m-1} < \infty$ and $0 < \xi_1 < \xi_2 < \dots < \xi_{m-1} < \infty$ (the different ways for determining these boundaries are discussed in the next section). The resulting transition matrix is denoted $P = \{p_{ij}\}$, and each element p_{ij} is a conditional probability that an individual moves to class j of income y given that she was initially in class i of income x , i.e.,

$$p_{ij} = \frac{\Pr(\zeta_{i-1} \leq x < \zeta_i \text{ and } \xi_{j-1} \leq y < \xi_j)}{\Pr(\zeta_{i-1} \leq x < \zeta_i)}, \quad (1)$$

where $\zeta_0 = \xi_0 = 0$ and $\zeta_m = \xi_m = \infty$. The probability that an individual falls into income class i of x is denoted π_i , i.e., $\pi_i = \Pr(\zeta_{i-1} \leq x < \zeta_i)$. Clearly, π_i can also be interpreted as the proportion of people in income class i of x and p_{ij} the proportion of the people in the i th class of x that moves to class j of y .²

2.1. Summary mobility measures

Using the transition matrix P rather than the *c.d.f.* $K(x, y)$, a mobility measure can be defined as a function $M(P)$, which maps P into a scalar. We say that a society with matrix P is more mobile than matrix \tilde{P} , denoted as $P \succcurlyeq_M \tilde{P}$, according to a measure M , if and only if $M(P) > M(\tilde{P})$. The following table documents several commonly used summary measures.³

In Table 1, M_1 measures the average probability across all classes that an individual will leave her initial class in the succeeding period; it is also interpreted as the normalized distance of P away from the identity matrix I (Bartholomew, 1996, p. 83). M_2 relies on the second largest eigenvalue (λ_2) of P which can be regarded as the distance between P and perfect mobility, or as a correlation coefficient between the initial and ending income classes. M_3 uses the product of all eigenvalues ($1 = |\lambda_1| > |\lambda_2| > \dots > |\lambda_m|$) as a measure of mobility. M_4 is identical to M_1 if $\pi_i = 1/m$ for all i . M_5 is the average number of income classes crossed by all individuals.

² In the literature on mobility measurement, the income distribution, $\pi = (\pi_1, \pi_2, \dots, \pi_m)$, is often assumed to be at or near a steady-state. This ensures that mobility will leave the static income distribution unchanged ($P\pi = \pi$). However, as emphasized by several researchers this assumption is unrealistic. An income regime may spend much more time away from its steady-state than at the steady-state. In fact, if the transition matrix is changing across time, the income distribution may never reach the steady state. In this paper we do not treat π as a steady-state distribution. Instead, π is simply the initial income distribution.

³ A more detailed discussion of these measures can be found in Bartholomew (1996).

Table 1
Summary measures of mobility

Measures	Sources
$M_1(P) = \frac{m - \sum_{i=1}^m p_{ii}}{m-1}$	Prais (1955), Shorrocks (1978a)
$M_2(P) = 1 - \lambda_2 $	Sommers and Conlisk (1979)
$M_3(P) = 1 - \det(P) = 1 - \prod_{i=1}^m \lambda_i $	Shorrocks (1978a)
$M_4(P) = \frac{m-m \sum_{i=1}^m \pi_i p_{ii}}{m-1}$	Bartholomew (1982)
$M_5(P) = \frac{1}{m-1} \sum_{i=1}^m \sum_{j=1}^m \pi_i p_{ij} i - j $	Bartholomew (1982)

2.2. Partial mobility orderings

In a seminal paper, Atkinson (1983, p. 61) proposed the first dominance approach to measuring income mobility. Atkinson’s method relates mobility “... to the properties of a social welfare function defined over incomes at different dates.” Mobility per se is not directly measured. Instead, the welfare implications of mobility are explored and an indirect measure of mobility implied. Atkinson considers a utilitarian social welfare function:

$$W(x, y) = \int_0^\infty \int_0^\infty U(x, y) dK(x, y), \tag{2}$$

where $U(x, y)$ satisfies $U_{xy} \leq 0$ and other regularity conditions. For two income regimes, characterized by $K(x, y)$ and $\tilde{K}(x, y)$, that also have identical marginal distributions (i.e., $F(x) = \tilde{F}(x)$ and $G(y) = \tilde{G}(y)$) or the distribution is at the steady state, Atkinson showed that the regime with $K(x, y)$ has *greater* social welfare than the regime with $\tilde{K}(x, y)$ according to *all* $W(x, y)$ if and only if

$$K(x, y) \leq \tilde{K}(x, y) \quad \text{for all } x \text{ and } y \tag{3}$$

with strict inequality holding for some x and y . When Atkinson’s result is applied to transition matrices, the requirement of equal marginal distributions is reflected in the fact that the sums of rows and columns must be the same between the matrices. For two transition matrices P and \tilde{P} , condition (3) becomes

$$\sum_{i=1}^k \sum_{j=1}^l \pi_j p_{ij} \leq \sum_{i=1}^k \sum_{j=1}^l \pi_j \tilde{p}_{ij} \quad \text{for all } k \text{ and } l, k, l = 1, 2, \dots, m, \tag{4}$$

with at least one strict inequality holding for some k and l . The dominance condition (4) has also been characterized by Dardanoni (1993) who shows that the Atkinson condition is both necessary and sufficient for one regime to have greater (Bergson–Samuelson) social welfare. Condition (4) is denoted as $P \succcurlyeq_{AD} \tilde{P}$.

Bénabou and Ok (1999) view mobility as a mechanism that equalizes income opportunities and derive a quite different dominance condition. They represent a person’s opportunity as the expected income in the succeeding period. For example, given that $(\eta_1, \eta_2, \dots, \eta_m)$ is the income vector associated with the m income classes in both regimes, a person initially in the i th class will have an expected income of

$e_i = \sum_{j=1}^m p_{ij}\eta_j$. Bénabou and Ok show that the necessary and sufficient condition for a (size) transition matrix P to be more opportunity equalizing than \tilde{P} , denoted as $P \succ_{BO} \tilde{P}$, for all possible income distributions of x is

$$\frac{\tilde{e}_1}{e_1} \geq \frac{\tilde{e}_2}{e_2} \geq \dots \geq \frac{\tilde{e}_m}{e_m}, \quad (5)$$

with at least one strict inequality holding for some $j=1, 2, \dots, m$. Note that, in contrast to the Atkinson–Dardanoni condition, Bénabou and Ok’s condition does not require the initial distributions to be equal.

Because the core of these summary measures and ordering criteria is a transition matrix, we will review in the next section the different ways that a transition matrix may be constructed and derive the large sample properties of the estimates of these matrices. Once these properties are derived, the inference procedures for testing mobility measures and ordering conditions can be straightforwardly established.

3. Asymptotic distributions of transition matrices

The meaning of mobility is largely determined by the way a transition matrix is constructed. For occupational mobility, classes are formed by aggregating individuals by profession or skill into agreed upon definitions of occupations. In this case, the data are discrete by nature, and the boundaries between classes easily drawn. For income mobility, however, income classes are formed by grouping individual incomes measured on a continuous scale. There are no natural boundaries for these classes; consequently, researchers can quite differently document and describe income movements within an income regime. In what follows, we first review the different approaches to constructing a transition matrix. We show that each approach leads to a different estimation method and a distinct asymptotic variance–covariance structure.

3.1. Different approaches to constructing an income transition matrix

The first approach views mobility as an absolute concept and exogenously sets boundaries between income classes. The resulting transition matrix is referred to as a *size transition matrix*. Using this approach the boundaries of income classes $\{\zeta_i\}$ and $\{\xi_i\}$ are predetermined and do not depend on the particular income regime or distribution under investigation. A number of writers, including Solow (1951), McCall (1973), Hart (1976a,b, 1983) and Schluter (1998), adopt this approach and construct size transition matrices. The advantage of this type of transition matrix is that it reflects income movement between different income levels; thus both the exchange of positions of individuals and economic growth (the increasing availability of positions at high income levels) are incorporated into mobility. One can draw welfare implications of mobility directly from comparisons of transition matrices of this type. We argue that size transition matrices are necessary for applying both the Atkinson–Dardanoni condition and the Bénabou–Ok condition. Welfare implications of these dominance conditions cannot be drawn if income mobility is not associated with absolute income levels.

The second approach views mobility as a relative concept. This approach allows the same number of individuals in each class. The resulting matrix is referred to as a *quantile transition matrix*. The advantage of this approach is that the transition matrix is biostochastic, and the steady-state condition is always satisfied. The disadvantage is that only those movements that involve reranking (i.e., people switching positions) is recorded as mobility. Thus, the quantile matrix approach cannot take into account whether overall income is increasing or decreasing. Thus, the “upward mobility” accompanying economic growth, which Kuznets (1966) studied, is ignored. It follows that studies using this type of transition matrix cannot draw a complete picture of changes in social welfare between different income regimes. Both Hart (1983) and Atkinson et al. (1992) voice concerns about the use of the quantile approach for this reason.

The third and fourth approaches incorporate elements of both the absolute and relative approaches to mobility. Class boundaries are defined as percentages of mean income or median income of the initial and ending distributions. The resulting matrices are, respectively, referred to as *mean transition matrix* and *median transition matrix*.⁴ In an early study, Thatcher (1971) uses the mean transition matrix in his analysis of the UK earnings mobility. Atkinson et al. (1992) argue that Thatcher’s approach relates income mobility to both income level and the relative positions of individuals. Trede (1998) and Burkhauser et al. (1998) consider the median transition matrix in their investigations of income/earnings mobility in the United States and Germany.

3.2. Estimation of transition matrices and the asymptotic distributions

Assume a random (paired) sample of size n , $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, drawn from the joint distribution $K(x, y)$. By definition, p_{ij} is the proportion of the people in the i th class of x who move into the j th class of y , and π_i is the proportion of people who fall into the i th class of x . It follows that the general approach to estimating mobility requires that we first estimate the boundaries between classes, denoted as $\{\hat{\zeta}_i\}$ and $\{\hat{\xi}_i\}$, then count the number of people within each class and between classes, and finally to calculate the appropriate ratios between them. In short, p_{ij} and π_i can be estimated as follows:

$$\hat{p}_{ij} = \frac{(1/n) \sum_{t=1}^n I(\hat{\zeta}_{i-1} \leq x_t < \hat{\zeta}_i \text{ and } \hat{\xi}_{j-1} \leq y_t < \hat{\xi}_j)}{(1/n) \sum_{t=1}^n I(\hat{\zeta}_{i-1} \leq x_t < \hat{\zeta}_i)}, \tag{6}$$

and

$$\hat{\pi}_i = \frac{1}{n} \sum_{t=1}^n I(\hat{\zeta}_{i-1} \leq x_t < \hat{\zeta}_i), \tag{7}$$

⁴ In a recent paper investigating the U.K. income mobility, Jarvis and Jenkins (1998) construct an absolute mobility matrix with the class boundaries (both initial and ending) being fractions of mean income of the initial distribution. The inference procedure for testing this “mixed” mobility matrix can be established as a special case of the mean mobility matrix.

where $\hat{\zeta}_0 = \hat{\xi}_0 = 0$ and $\hat{\zeta}_m = \hat{\xi}_m = \infty$. $I(\cdot)$ is an indicator variable which is 1 if the condition is satisfied and 0 otherwise. Let

$$\theta_{ij} = \Pr(\zeta_{i-1} \leq x < \zeta_i \text{ and } \xi_{j-1} \leq y < \xi_j) \tag{8}$$

be the proportion of people falling into the i th class of x and the j th class of y . The estimator of θ_{ij} is

$$\hat{\theta}_{ij} = \frac{1}{n} \sum_{t=1}^n I(\hat{\zeta}_{i-1} \leq x_t < \hat{\zeta}_i \text{ and } \hat{\xi}_{j-1} \leq y_t < \hat{\xi}_j) \tag{9}$$

and $\hat{p}_{ij} = \hat{\theta}_{ij} / \hat{\pi}_i$.

To facilitate our presentation, let π be the $1 \times m$ vector $(\pi_1, \pi_2, \dots, \pi_m)$, θ be the $1 \times mm$ vectorized matrix of $\{\theta_{ij}\}$ (i.e., lay each row of the matrix next to the above row in a single line), and p be the $1 \times mm$ vectorized matrix of $\{p_{ij}\}$. The main result of this section is

Theorem 1. *Under the assumption that $K(x, y)$, $F(x)$ and $G(y)$ are continuous and differentiable, and the first two moments of $F(x)$ and $G(x)$ exist and are finite, then (i) $\hat{\theta}$ and \hat{p} are consistent estimators of θ and p , respectively, and (ii) $\hat{\theta}$ and \hat{p} are asymptotically normal with mean vectors θ and p and variance–covariance matrices Φ and Ψ , respectively.*

In what follows, we derive the variance–covariance structures Φ and Ψ for each transition matrix and prove the asymptotic normalities of $\hat{\theta}$ and \hat{p} .

3.2.1. Size transition matrix

For size transition matrices with exogenously determined class boundaries, $\{\zeta_i\}$ and $\{\xi_i\}$, no boundary estimation is needed. Therefore, $\hat{\theta}_{ij}$, $\hat{\pi}_i$ and \hat{p}_{ij} can be directly estimated as $\hat{\theta}_{ij} = (1/n) \sum_{t=1}^n I(\zeta_{i-1} \leq x_t < \zeta_i \text{ and } \xi_{j-1} \leq y_t < \xi_j)$, $\hat{\pi}_i = (1/n) \sum_{t=1}^n I(\zeta_{i-1} \leq x_t < \zeta_i)$ and $\hat{p}_{ij} = \hat{\theta}_{ij} / \hat{\pi}_i$. Clearly, the law of large numbers and the central limit theorem imply that both $\hat{\theta}_{ij}$ and $\hat{\pi}_i$ tend to normal variates and are consistent estimates of θ_{ij} and π_i , respectively. It follows from the Slutsky theorem (see, for example, Serfling, 1980, p. 19) that \hat{p}_{ij} also tends to a normal variable and is a consistent estimate of p_{ij} .

Through direct calculations, we can easily show that the asymptotic covariances among $\hat{\pi}_i$, $\hat{\pi}_j$, $\hat{\theta}_{ij}$ and $\hat{\theta}_{kl}$ are:

$$\text{cov}(\hat{\pi}_i, \hat{\pi}_j) = \begin{cases} \frac{\pi_i(1 - \pi_i)}{n} & \text{if } i = j, \\ -\frac{\pi_i\pi_j}{n} & \text{else,} \end{cases} \tag{10}$$

$$\text{cov}(\hat{\theta}_{ij}, \hat{\theta}_{kl}) = \begin{cases} \frac{\theta_{ij}(1 - \theta_{ij})}{n} & \text{if } i = k \text{ and } j = l, \\ -\frac{\theta_{ij}\theta_{kl}}{n} & \text{else,} \end{cases} \tag{11}$$

and

$$cov(\hat{\pi}_i, \hat{\theta}_{kl}) = \begin{cases} \frac{\theta_{kl}(1 - \pi_i)}{n} & \text{if } i = k, \\ -\frac{\pi_i \theta_{kl}}{n} & \text{else.} \end{cases} \tag{12}$$

Thus, we can derive the variance–covariance structures for $\hat{\pi}$, $\hat{\theta}$ and $(\hat{\pi}, \hat{\theta})$. The covariance matrix of $\hat{\theta}$ is denoted Φ_1 .

Using the well-known delta method and the covariance matrix of $(\hat{\pi}, \hat{\theta})$ derived above, we can also derive the covariance matrix for \hat{p} :

$$cov(\hat{p}_{ij}, \hat{p}_{kl}) = \begin{cases} \frac{p_{ij}(1 - p_{ij})}{n\pi_i} & \text{if } i = k \text{ and } j = l, \\ -\frac{p_{ij}p_{il}}{n\pi_i} & \text{if } i = k, \\ 0 & \text{else.} \end{cases} \tag{13}$$

Denote this covariance matrix as Ψ_1 .⁵

3.2.2. Quantile transition matrix

This section starts with the observation that, for a quantile transition matrix, class boundaries must be endogenously determined from the data in such a way that the total number of people are divided equally across classes, both in the distribution of x and in the distribution of y . Because the class boundaries are now stochastic, it follows that we need to consider the variability of class boundaries in computing the variance of \hat{p}_{ij} .

Without loss of generality, we assume that the n individuals are divided into m classes and the proportions of people in these classes are $\pi_1, \pi_2, \dots, \pi_m$ in both distributions. Denote the cumulative proportion of people in the first i classes as

$$s_i = \sum_{j=1}^i \pi_j. \tag{14}$$

Because the proportion of people in each class (π_i) is given and not stochastic, the transition matrix $\{p_{ij}\}$ can be estimated as

$$\hat{p}_{ij} = \frac{\hat{\theta}_{ij}}{\hat{\pi}_i} = \frac{(1/n) \sum_{t=1}^n I(\hat{\zeta}_{i-1} \leq x_t < \hat{\zeta}_i \text{ and } \hat{\xi}_{j-1} \leq y_t < \hat{\xi}_j)}{\pi_i}. \tag{15}$$

Denoting $\tilde{\theta}_{ij} = (1/n) \sum_{t=1}^n I(\zeta_{i-1} \leq x_t < \zeta_i \text{ and } \xi_{j-1} \leq y_t < \xi_j)$, we can rewrite $\hat{\theta}_{ij}$ as

$$\hat{\theta}_{ij} = \tilde{\theta}_{ij} + (\hat{\theta}_{ij} - \tilde{\theta}_{ij}). \tag{16}$$

As shown in Appendix A, (16) can be further expressed as

$$\hat{\theta}_{ij} = \tilde{\theta}_{ij} + a_{ij}(\hat{\zeta}_i - \zeta_i) + b_{ij}(\hat{\xi}_j - \xi_j) - c_{ij}(\hat{\zeta}_{i-1} - \zeta_{i-1}) - d_{ij}(\hat{\xi}_{j-1} - \xi_{j-1}), \tag{17}$$

where a_{ij} , b_{ij} , c_{ij} and d_{ij} are given in (A.8).

⁵ The result in (13) is also provided in Trede (1999) who utilized a result due to Christensen (1990).

Because $\tilde{\theta}_{ij}$ converges to θ_{ij} almost surely and $\hat{\zeta}_i, \hat{\zeta}_{i-1}, \hat{\zeta}_j$ and $\hat{\zeta}_{j-1}$ converge, respectively, to $\zeta_i, \zeta_{i-1}, \zeta_j$ and ζ_{j-1} almost surely (Serfling, 1980, p. 75), $\hat{\theta}_{ij}$ also converges to θ_{ij} almost surely. It follows that \hat{p}_{ij} converges almost surely to p_{ij} . Further, because $\tilde{\theta}_{ij}, \hat{\zeta}_i, \hat{\zeta}_{i-1}, \hat{\zeta}_j$ and $\hat{\zeta}_{j-1}$ are all asymptotically normal, \hat{p}_{ij} is also asymptotically normal.

From (17), one can see that the variability of $\hat{\theta}_{ij}$ comes not only from the simple counting process ($\tilde{\theta}_{ij}$) but also from the determination of class boundaries ($\hat{\zeta}_{i-1}, \hat{\zeta}_i, \hat{\zeta}_{j-1}$ and $\hat{\zeta}_j$). Thus, the variance of \hat{p}_{ij} would be quite different from what is calculated from (13) when the class boundaries are treated as exogenously determined. In practice, this difference may very likely be so substantial that it cannot be ignored.⁶

To derive the asymptotic variance of $\hat{\theta}_{ij}$, we need to use the Bahadur representation (Bahadur, 1966; Ghosh, 1971) which states the relationship between a population quantile, say ζ_i , and its sample estimate, $\hat{\zeta}_i$:

$$\hat{\zeta}_i - \zeta_i = \frac{s_i - (1/n) \sum_{t=1}^n I(x_t < \zeta_i)}{f(\zeta_i)} + o_p(n^{-1/2}), \tag{18}$$

where $f(x)$ is the density function of $F(x)$. Using this relationship, the covariance matrix for $\hat{\theta}$ and the vectorized matrix of $\{\hat{\theta}_{ij}\}$ can be readily derived.⁷ The covariance matrix is denoted as Φ_2 . The covariance matrix of the vector \hat{p} , which is the vectorized matrix of $\{\hat{p}_{ij}\}$, can also be obtained from Φ_2 using the relation $\hat{p}_{ij} = \hat{\theta}_{ij}/\pi_i$. The resulting covariance matrix is denoted as Ψ_2 .

3.2.3. Mean transition matrix

The class boundaries in the mean transition matrix case are percentages of the mean income of each distribution (x and y). Let μ_x and μ_y be the mean incomes of the two distributions and $0 < \alpha_1 < \alpha_2 < \dots < \alpha_{m-1} < \infty$ be pre-specified percentages, the corresponding class boundaries are $\{\alpha_i \mu_x\}$ and $\{\alpha_i \mu_y\}$, respectively. It follows that θ_{ij} and π_i can be estimated as

$$\hat{\theta}_{ij} = \frac{1}{n} \sum I(\alpha_{i-1} \bar{x} \leq x_t < \alpha_i \bar{x} \text{ and } \alpha_{j-1} \bar{y} \leq y_t < \alpha_j \bar{y}) \tag{19}$$

and

$$\hat{\pi}_i = \frac{1}{n} \sum I(\alpha_{i-1} \bar{x} \leq x_t < \alpha_i \bar{x}), \tag{20}$$

respectively, where \bar{x} and \bar{y} are sample means, $\alpha_0 = 0$ and $\alpha_m = \infty$. Consequently, p_{ij} is estimated as $\hat{p}_{ij} = \hat{\theta}_{ij}/\hat{\pi}_i$.

Applying the same reasoning used in the case of quantile transition matrix, we can show that

$$\hat{\theta}_{ij} \sim \tilde{\theta}_{ij} + (\alpha_i a_{ij} - \alpha_{i-1} c_{ij})(\bar{x} - \mu_x) + (\alpha_i b_{ij} - \alpha_{i-1} d_{ij})(\bar{y} - \mu_y) \tag{21}$$

⁶ For example, in our illustration below, the first element in the U.S. absolute mobility covariance matrix is 0.9195, while the first element in the relative mobility covariance matrix is 2.2448.

⁷ Appendix B provides formulae for elements of the covariance structures of the estimated quantile, mean and median transition matrices.

and

$$\hat{\pi}_i \sim \tilde{\pi}_i + [\alpha_i f(\alpha_i \mu_x) - \alpha_{i-1} f(\alpha_{i-1} \mu_x)](\bar{x} - \mu_x), \tag{22}$$

where $\tilde{\theta}_{ij} = (1/n) \sum I(\alpha_{i-1} \mu_x \leq x_t < \alpha_i \mu_x \text{ and } \alpha_{j-1} \mu_y \leq y_t < \alpha_j \mu_y)$, $\tilde{\pi}_i = (1/n) \sum I(\alpha_{i-1} \mu_x \leq x_t < \alpha_i \mu_x)$, a_{ij} , b_{ij} , c_{ij} and d_{ij} are the same as those defined in Appendix A (but replacing ζ_i and ξ_i with $\alpha_i \mu_x$ and $\alpha_i \mu_y$, respectively).

From (21) and (22), we know that $\hat{\theta}_{ij}$ and $\hat{\pi}_i$ converge almost surely to θ_{ij} and π_i , respectively. It follows that \hat{p}_{ij} will also converge almost surely to p_{ij} . Because the joint distribution of vector $(\hat{\pi}, \hat{\theta})$ is asymptotically normal, vector \hat{p} will also be asymptotically normal by the Slutsky theorem. Finally, the covariance structure of $(\hat{\pi}, \hat{\theta})$ can be derived directly from (21) and (22), and the covariance structure of \hat{p} can be derived using the delta method. The covariance matrix of $\hat{\theta}$ is denoted by Φ_3 , and the covariance matrix of \hat{p} by Ψ_3 .

3.2.4. Median transition matrix

The class boundaries of the median transition matrix are determined in a manner similar to the mean transition matrix. In fact, we only need to replace the mean income with median income at appropriate places. Thus, the derivation process is the same as in the previous section. Denoting population median as τ_x and τ_y and their sample estimates as x_d and y_d , we have

$$\hat{\theta}_{ij} \sim \tilde{\theta}_{ij} + (\alpha_i a_{ij} - \alpha_{i-1} c_{ij})(x_d - \tau_x) + (\alpha_i b_{ij} - \alpha_{i-1} d_{ij})(y_d - \tau_y) \tag{23}$$

and

$$\hat{\pi}_i \sim \tilde{\pi}_i + [\alpha_i f(\alpha_i \tau_x) - \alpha_{i-1} f(\alpha_{i-1} \tau_x)](x_d - \tau_x), \tag{24}$$

where $\tilde{\theta}_{ij}$, $\tilde{\pi}_i$, a_{ij} , b_{ij} , c_{ij} and d_{ij} are defined similarly as above. The covariance matrix of $\hat{\theta}$ is denoted Φ_4 and the covariance matrix of \hat{p} is denoted Ψ_4 . Note that in deriving these covariance structures, one must replace the terms $(x_d - \tau_x)$ and $(y_d - \tau_y)$ with the corresponding Bahadur representation (18).

3.3. Estimation of the covariance matrices Φ and Ψ

The covariance structures derived above are generally not known and must be estimated from the sample. For the size transition matrix, the estimation of Φ_1 and Ψ_1 is straightforward. For the other three types of transition matrices, consistent estimates can also be obtained by finding a consistent estimate for each element.

For the quantile transition matrix, we first need to estimate density functions $f(\zeta_i)$ and $g(\xi_i)$ for $i = 1, 2, \dots, m - 1$. Silverman (1986) presents several methods of density estimation. Among these approaches, kernel estimation is the most popular because the consistency of the estimation is well established in the literature and its application is relatively straightforward. In performing the computation using the kernel method, one needs to choose a kernel function and a window width function. In empirical studies, the Epanechnikov kernel function has often been used. Silverman (1986, p. 42) recommends this method.

The estimation of the coefficients a_{ij} , b_{ij} , c_{ij} and d_{ij} requires the estimation of the bivariate density functions such as $\int_0^{\xi_j} k(\zeta_i, y) dy$ and $\int_0^{\xi_i} k(x, \zeta_j) dx$. In the literature, the kernel method has also been extended to the estimation of bivariate density $k(x, y)$. To estimate $\int_0^{\xi_j} k(\zeta_i, y) dy$, we suggest the use of the following approximation method. First divide the interval $[0, \xi_j]$ into, say v , equal sub-intervals $[\zeta_{j(t-1)}, \zeta_{jt}]$, $t=1, 2, \dots, v$, with $\zeta_{j0}=0$ and $\zeta_{jv}=\xi_j$. Then estimate $k(x, y)$ at each point (ζ_i, ζ_{jt}) using the standard bivariate kernel method. Finally, $\int_0^{\xi_j} k(\zeta_i, y) dy$ is estimated as $(\xi_j/v) \sum_{t=1}^v \hat{k}(\zeta_i, \zeta_{jt})$. In empirical studies, one can also use the multivariate Epanechniko kernel function (Silverman, 1986, p. 76) to estimate $k(x, y)$.

All elements involved in both the mean transition matrix and median transition matrix can be similarly estimated. It is easy to see that these estimations are consistent. Therefore, all four transition matrices and their variance–covariance structures can be consistently estimated.

4. Testing mobility measures and partial mobility orderings

Using the results developed in Section 3, we are now in a position to establish the testing procedures for various mobility measures and partial ordering criteria.

4.1. Summary mobility measures

Trede (1999) provides inference procedures for testing summary mobility measures using occupational (size) transition matrices. His results can be directly generalized to other types of transition matrices using the various covariance matrices derived in the previous section. In general, a mobility measure can be regarded as a function of vector θ (M_4 and M_5 when π is the initial income distribution rather than the equilibrium distribution) or p (M_1 , M_2 and M_3). The use of the well-known delta method leads directly to the variance formulae of each estimation.

4.2. Partial mobility orderings

To test the Atkinson–Dardanoni mobility dominance condition, we need to first express their condition as a comparison between two vectors. By introducing an $mm \times mm$ matrix $H = T \otimes T$ where T is the $m \times m$ lower triangular matrix of 1s, we can define a vector $\beta_{AD} = HP$. Consequently, the Atkinson–Dardanoni mobility dominance ($P \succcurlyeq_{AD} \tilde{P}$) between matrices P and \tilde{P} is

$$\tilde{\beta}_{AD} \geq \beta_{AD}, \quad (25)$$

and the asymptotic covariance matrix of $\tilde{\beta}_{AD}$ is $\Omega_{AD} = H\Phi_1H'$.

For two income regimes with the same initial income distribution, the dominance relationship between vectors β_{AD} and $\tilde{\beta}_{AD}$ can be tested in several ways (e.g., the union–intersection approach as used in Bishop et al. (1992) or the intersection–union method as proposed by Howes (1994)). Alternatively, a version of the general Wald statistic can be used to test for significant differences. In this paper, we follow Kodde

and Palm (1986) and Wolak (1989) and use the Wald procedure to test the following sets of hypotheses:

$$H_0 : P =_{AD} \tilde{P} \text{ versus } H_1 : P \succneq_{AD} \tilde{P} \tag{26}$$

and

$$H_0 : P \succneq_{AD} \tilde{P} \text{ versus } H_1 : P \not\succeq_{AD} \tilde{P}. \tag{27}$$

Assume two samples drawn independently from the two regimes with β_{AD} and $\tilde{\beta}_{AD}$. If we denote $\Delta\beta_{AD} = \beta_{AD} - \tilde{\beta}_{AD}$, then the covariance matrix of $\Delta\hat{\beta}_{AD}$ is $\Omega_{AD} = H(\Phi_1 + \tilde{\Phi}_1)H'$. The critical step in using the Wald test is to solve the following minimization problem:

$$\min_{\gamma \geq 0} (\Delta\hat{\beta}_{AD} - \gamma)' \Omega_{AD}^{-1} (\Delta\hat{\beta}_{AD} - \gamma). \tag{28}$$

Denoting the solution to this minimization question as $\bar{\gamma}$, we can compute the following two Wald test statistics:

$$c_1 = (\Delta\hat{\beta}_{AD})' \Omega_{AD}^{-1} (\Delta\hat{\beta}_{AD}) - (\Delta\hat{\beta}_{AD} - \bar{\gamma})' \Omega_{AD}^{-1} (\Delta\hat{\beta}_{AD} - \bar{\gamma}) \tag{29}$$

and

$$c_2 = (\Delta\hat{\beta}_{AD} - \bar{\gamma})' \Omega_{AD}^{-1} (\Delta\hat{\beta}_{AD} - \bar{\gamma}). \tag{30}$$

Next, compare c_1 or c_2 with the lower bound and upper bound of the critical value for a pre-selected significance level (Kodde and Palm (1986) provide a table of these values). If c_1 or c_2 lies below the lower bound then H_0 is accepted; if c_1 or c_2 falls above the upper bound then H_0 is rejected. If c_1 or c_2 falls between the lower bound and the upper bound, then a Monte Carlo simulation is required to complete the inference. See Wolak (1989) for details and Fisher et al. (1998) for an illustration.

Denote $\phi = (\tilde{e}_1/e_1, \tilde{e}_2/e_2, \dots, \tilde{e}_m/e_m)$, where \tilde{e}_i and e_i are defined in (5), the Bénabou–Ok condition ($P \succneq_{BO} \tilde{P}$) can be expressed as

$$\beta_{BO} = R\phi' \geq 0, \tag{31}$$

where the $(m - 1) \times m$ matrix R has 1 on its main diagonal, -1 on the diagonal above and 0 elsewhere. The covariance matrix of $\hat{\beta}_{BO}$ is derived as follows. First, derive the covariance matrix of the estimates of $\varepsilon = (e_1, e_2, \dots, e_m, \tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_m)$. Because $\varepsilon = R_1 \bar{p}'$, where $\bar{p} = (p, \tilde{p})$ and $R_1 = T_1 \otimes I \otimes \eta$ with T_1 being a 1×2 matrix of 1s, I the $m \times m$ identity matrix and η the vector of income levels, the covariance matrix of $\hat{\varepsilon}$ is

$$\Sigma_1 = R_1 \begin{pmatrix} \Psi & 0 \\ 0 & \tilde{\Psi} \end{pmatrix} R_1'. \tag{32}$$

Next, introduce an $(m - 1) \times mm$ matrix

$$R_2 = \begin{pmatrix} \frac{1}{e_1} & -\frac{1}{e_2} & 0 & \dots & 0 & 0 & -\frac{\tilde{e}_1}{e_1^2} & \frac{\tilde{e}_2}{e_2^2} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{e_2} & -\frac{1}{e_3} & \dots & 0 & 0 & 0 & -\frac{\tilde{e}_2}{e_2^2} & \frac{\tilde{e}_3}{e_3^2} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{1}{e_{m-1}} & -\frac{1}{e_m} & 0 & 0 & 0 & \dots & -\frac{\tilde{e}_{m-1}}{e_{m-1}^2} & \frac{\tilde{e}_m}{e_m^2} \end{pmatrix}. \tag{33}$$

The covariance matrix of $\hat{\phi}$ is then

$$\Sigma_2 = R_2 \Sigma_1 R_2'. \quad (34)$$

As a consequence, the covariance matrix of $\hat{\beta}_{BO}$ is

$$\Omega_{BO} = R \Sigma_2 R'. \quad (35)$$

5. An illustration: earnings mobility in the U.S. and Germany

Several recent studies (e.g., Burkhauser and Holtz-Eakin, 1993; Burkhauser and Poupore, 1997; Burkhauser et al., 1997, 1998; Schluter, 1996, 1997, 1998; Trede, 1999) compare income and earnings mobility between the U.S. and Germany for the decade of the 1980s. Data are drawn from the Panel Study of Income Dynamics (PSID) and the German Socio-Economic Panel (GSOEP). The PSID utilizes a representative sample of about 5000 families with interviews conducted annually since 1968 (see Hill, 1992 for a detailed description). The GSOEP draws upon a sample of about 6000 families (see Burkhauser et al., 1995) and was started in 1984. Using summary measures of mobility, these studies find “similarities” in mobility between the two countries, although “inequality was greater in the U.S. than in Germany during the 1980s” (Burkhauser et al., 1998).

In this illustration, we test both relative mobility and absolute mobility between the U.S. and German individual earnings over the 1985–1990 period. Testing relative mobility allows us to examine the commonly invoked (relative) notion of mobility and demonstrate the importance of statistical inference; testing absolute mobility allows us to determine whether or not a broader definition of mobility may lead to different conclusions on mobility. Testing absolute mobility also allows us to draw welfare implications.

5.1. Mobility comparisons with quantile transition matrices

To test relative mobility, we divide each population into five income classes with equal number of people in each class. The resulting transition matrices are quintile matrices with $\pi = (0.2, 0.2, 0.2, 0.2, 0.2)$. The estimated matrices are:

$$\hat{P}_{US} = \begin{bmatrix} 0.500 & 0.275 & 0.141 & 0.058 & 0.026 \\ 0.311 & 0.377 & 0.197 & 0.080 & 0.035 \\ 0.111 & 0.233 & 0.411 & 0.193 & 0.052 \\ 0.053 & 0.079 & 0.199 & 0.481 & 0.188 \\ 0.025 & 0.036 & 0.052 & 0.188 & 0.699 \end{bmatrix}$$

and

$$\hat{P}_{GM} = \begin{bmatrix} 0.618 & 0.224 & 0.094 & 0.035 & 0.029 \\ 0.240 & 0.430 & 0.190 & 0.094 & 0.047 \\ 0.072 & 0.239 & 0.401 & 0.226 & 0.062 \\ 0.041 & 0.076 & 0.262 & 0.481 & 0.140 \\ 0.029 & 0.031 & 0.053 & 0.164 & 0.723 \end{bmatrix}.$$

We first test summary mobility measures between the two countries. Table 2 reports estimates of summary mobility measures, associated standard errors and test statistics. All eigenvalues for both transition matrices are real and positive. Examining the hypothesis $H_0: U.S. =_M \text{ Germany}$ versus $H_1: U.S. \succcurlyeq_M \text{ Germany}$, we reject H_0 at the 1 percent significance level for all test statistics except that of M_2 . Thus, we conclude that the U.S. has more (relative) earning mobility than Germany, according to all of the measures we considered except one.⁸

Because both \hat{P}_{US} and \hat{P}_{GM} are quintile matrices, the equal-initial-distribution and steady-state condition required by the Atkinson–Dardanoni condition is automatically satisfied. Clearly, the two countries’ income levels corresponding to $\pi=(0.2, 0.2, 0.2, 0.2, 0.2)$ will not be the same, and, hence, the equal-income-level assumption required by the Bénabou–Ok condition is not fulfilled. Thus, we will test only the Atkinson–Dardanoni condition; the Bénabou–Ok condition will be tested below with size transition matrices. Condition (4) requires that the matrix of $\{\sum_{i=1}^k \sum_{j=1}^l \pi_j(p_{ij} - \tilde{p}_{ij})\}$ be nonpositive with some strictly negative elements. The sample estimate of this

Table 2
Summary indices of relative mobility

	M_1	M_2	M_3	M_4	M_5
U.S.	0.639 (0.0075)	0.286 (0.0013)	0.994 (0.0011)	0.639 (0.0075)	0.179 (0.0026)
Germany	0.588 (0.0101)	0.284 (0.0011)	0.986 (0.0025)	0.588 (0.0101)	0.164 (0.0035)
<i>z</i> -value	4.088 ^a	1.192	2.728 ^a	4.088 ^a	3.520 ^a

The numbers in parentheses are standard errors.

^aDifference is significant at the 1 percent level.

⁸ This conclusion is generally in line with previous findings of [Burkhauser et al. \(1998\)](#) and others. It is useful to note that although some mobility indices for the two countries are fairly close to each other, the differences may be statistically significant. This observation suggests the necessity of using statistical inference in describing the “similarities” between the U.S. and Germany in income and earnings mobility.

matrix is

$$\begin{bmatrix} -0.024 & -0.013 & -0.004 & 0.001 & 0 \\ -0.009 & -0.010 & 0.001 & 0.003 & 0 \\ -0.002 & 0.000 & 0.005 & 0.005 & 0 \\ 0.001 & 0.000 & 0.000 & -0.005 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Suppose we are interested in testing the hypothesis $H_0: \text{U.S.} =_{AD} \text{Germany}$ versus $H_1: \text{U.S.} \succcurlyeq_{AD} \text{Germany}$. By solving the minimization problem of (28) using some MATLAB functions, we obtain the test statistic $c_1 = 69.086$. Because c_1 is greater than the upper bound of the critical value of 31.353—from Kodde and Palm (1986)—at the 1 percent significance level (the degrees of freedom are 16—there are 16 inequality conditions), we reject the null hypothesis that the U.S. and Germany have the same amount of (relative) individual earnings mobility as evaluated by the Atkinson–Dardanoni criterion.

5.2. Mobility comparisons with size transition matrices

We now consider the size transition matrices of the U.S. and Germany in order to draw welfare implications from mobility comparisons. In constructing the transition matrices, we use five earnings classes in 1990 U.S. dollars. The class boundaries are \$0, \$10,000, \$20,000, \$35,000, \$50,000 and ∞ . The representative earnings level of each class is chosen to be the middle point of that class; i.e., earnings vector η is (\$5,000, \$15,000, \$27,500, \$42,500, \$80,000). German earnings can be converted into U.S. dollars using purchasing power parity or simply normalized to U.S. mean earnings. We choose the latter approach. The distributions of people within these classes of the two countries in 1985 are estimated as

$$\hat{\pi}_{US} = (0.268, 0.261, 0.263, 0.125, 0.082)$$

and

$$\hat{\pi}_{GM} = (0.160, 0.235, 0.429, 0.125, 0.051),$$

respectively. The estimated transition matrices over the period are

$$\hat{P}_{US} = \begin{bmatrix} 0.444 & 0.367 & 0.154 & 0.027 & 0.008 \\ 0.193 & 0.438 & 0.300 & 0.050 & 0.020 \\ 0.055 & 0.136 & 0.550 & 0.200 & 0.056 \\ 0.033 & 0.054 & 0.218 & 0.401 & 0.294 \\ 0.028 & 0.030 & 0.040 & 0.201 & 0.702 \end{bmatrix}$$

and

$$\hat{P}_{GM} = \begin{bmatrix} 0.515 & 0.311 & 0.155 & 0.014 & 0.005 \\ 0.165 & 0.434 & 0.352 & 0.037 & 0.011 \\ 0.041 & 0.105 & 0.733 & 0.110 & 0.011 \\ 0.033 & 0.012 & 0.174 & 0.637 & 0.144 \\ 0.021 & 0.017 & 0.046 & 0.201 & 0.715 \end{bmatrix}.$$

Clearly, the requirement of equal initial distributions between the two countries in the Atkinson–Dardanoni condition is not satisfied. Thus, instead of asking broadly about which country is more mobile in earnings, we address the following question: will U.S. earnings be equally mobile if the German transition pattern is imposed on the U.S. initial distribution?

Table 3 reports estimates of summary mobility measures, associated standard errors, and test statistics. All eigenvalues of both transition matrices are real and positive. Examining the hypothesis $H_0: U.S. =_M \text{Germany}$ versus $H_1: U.S. \succcurlyeq_M \text{Germany}$, all test statistics are greater than the critical value at the 1 percent significance level. Thus, the null hypothesis that the U.S. has the same (absolute) mobility level with both the U.S. transition matrix and the German transition matrix is rejected.

The Atkinson–Dardanoni condition examines whether the U.S. transition matrix is equally welfare enhancing as the German matrix. The sample estimate of the difference matrix $\{\sum_{i=1}^k \sum_{j=1}^l \pi_j(p_{ij} - \tilde{p}_{ij})\}$ is

$$\begin{bmatrix} -0.019 & -0.004 & -0.004 & -0.001 & 0 \\ -0.012 & 0.004 & -0.010 & -0.003 & 0 \\ -0.008 & 0.016 & -0.046 & -0.015 & 0 \\ -0.008 & 0.021 & -0.035 & -0.034 & 0 \\ -0.008 & 0.023 & -0.034 & -0.033 & 0 \end{bmatrix}.$$

Table 3
Summary indices of absolute mobility

	M_1	M_2	M_3	M_4	M_5
U.S.	0.617 (0.0081)	0.273 (0.0125)	0.990 (0.0014)	0.643 (0.0075)	0.160 (0.0027)
Germany	0.491 (0.0121)	0.218 (0.0173)	0.954 (0.0056)	0.521 (0.0113)	0.125 (0.0040)
z-value	8.562 ^a	2.591 ^a	6.181 ^a	8.904 ^a	7.335 ^a

The numbers in parentheses are standard errors.

^aDifference is significant at the 1 percent level.

Suppose we want to test the hypothesis $H_0: \text{U.S.} =_{AD} \text{Germany}$ versus $H_1: \text{U.S.} \succ_{AD} \text{Germany}$. The test statistic is 136.406 which is greater than the upper bound of the critical value of 36.935—again from Kodde and Palm (1986)—at the 1 percent significance level (the degrees of freedom are 20). Thus, we reject the null hypothesis that the U.S. mobility pattern and the German mobility pattern are equivalent in enhancing social welfare as evaluated by the Atkinson–Dardanoni condition.

Finally, we test the Bénabou–Ok condition. Because we put the U.S. and Germany on the same income scale (by converting the German marks into U.S. dollars), both countries have the same income level for each income class. Thus the equal-income-level assumption required by the Bénabou–Ok condition is satisfied although the initial distributions are not the same. Table 4 lists the expected incomes under each transition matrix and their comparisons. The Bénabou–Ok condition requires the last column of Table 4 to be nonpositive. If we test $H_0: \text{U.S.} =_{BO} \text{Germany}$ versus $H_1: \text{U.S.} \succ_{BO} \text{Germany}$, the test statistic is 4.242 which is smaller than the lower bound of the critical value (5.412) at the 1 percent significance level. Thus the hypothesis that the U.S. mobility pattern and the German mobility pattern are equally effective in equalizing expected earnings distributions cannot be rejected.

To sum up, the two processes are not (statistically) significantly different in equalizing expected income distributions according to the Bénabou–Ok condition. However, under the Atkinson–Dardanoni condition, the U.S. (absolute) mobility process yields (statistically) significantly greater social welfare than the German (absolute) mobility processes.

6. Summary and conclusion

Economists have long recognized that the measurement of income distribution alone is inadequate for evaluating social welfare. An important additional consideration is the degree of mobility present in alternative income regimes. Since the work of (1955), various summary measures of mobility and partial mobility ordering criteria have been proposed. While each summary mobility measure captures a specific intuitive characteristic of mobility, partial mobility orderings allow researchers to assess the welfare implications of mobility.

The increasing attention focused on the measurement of mobility using sample data points to the need for the development of appropriate statistical tests for differences

Table 4
Testing the Bénabou–Ok dominance condition

Classes	$e_{US}(\$)$	$e_{GM}(\$)$	$r = E_{GM}/E_{US}$	$r_i - r_{i+1}$
0–10,000	13,730	12,513	0.9114	–0.0893
10,000–20,000	19,494	19,508	1.0007	0.1028
20,000–35,000	30,614	27,487	0.8979	–0.0219
35,000–50,000	47,539	43,724	0.9197	–0.0953
$\geq 50,000$	66,366	67,363	1.0150	

in mobility. Schluter (1998), Trede (1999), Maasoumi and Trede (2001), and Biewen (2002) recently establish inference procedures for several summary indices of mobility. This paper also takes up the issue of statistical inference for the measurement of income mobility and provides general testing procedures for a variety of mobility measures. Like most earlier studies, we focus our attention on mobility indices derived from transition matrices. We deviate from previous studies by taking into account the various ways a transition matrix can be constructed. Different methods of construction lead not only to distinct interpretations of mobility but to different asymptotic distributions as well. We find that the variance–covariance structure of the most commonly used transition matrix, the quantile transition matrix, is much more complicated than simple intuition would suggest, and is radically different from the covariance matrix of the size transition matrix or occupational transition matrix. We also review several partial mobility ordering conditions and provided appropriate inference procedures. The test procedures we develop are all asymptotically normal, and the various variance–covariance structures consistently estimable.

To illustrate the inference procedures we test for differences in several summary mobility measures and partial mobility ordering conditions, and compare earnings mobility in the U.S. and Germany between 1985 and 1990. The data are drawn from the PSID and GSOEP. We examine the question of whether the U.S. and Germany are equally mobile in terms of individual earnings. We test the hypotheses both in terms of quantile transition matrix and the size transition matrix. Quantile transition matrices allow us to test the robustness of the established conclusions in the literature, while the size matrices allow us to draw valid welfare implications from mobility comparisons. For the mobility comparisons with the quantile matrices, all but one summary measure indicate that the U.S. individual earnings is more (relatively) mobile than Germany. Testing the same pair of hypotheses, the partial ordering condition (the Atkinson–Dardanoni criterion) indicates that the U.S. is more (relatively) mobile than Germany in earnings. For size transition matrices, all summary measures indicate that the U.S. is more (absolutely) mobile. The Atkinson–Dardanoni dominance condition also reveals that the U.S. earnings mobility pattern is more welfare enhancing than Germany. Only for the Bénabou–Ok condition, do we find a failure to reject the hypothesis that the U.S. mobility pattern and the Germany mobility pattern are essentially equivalent in equalizing expected earnings distributions.

Acknowledgements

We thank two anonymous referees, an Associate Editor and seminar participants at various conferences and universities for useful comments and suggestions on earlier versions of the paper. The usual caveats apply.

Appendix A

Denoting $\tilde{\theta}_{ij} = (1/n) \sum_{t=1}^n I(\zeta_{i-1} \leq x_t < \zeta_i \text{ and } \xi_{j-1} \leq y_t < \xi_j)$, we can rewrite $\hat{\theta}_{ij}$ as

$$\hat{\theta}_{ij} = \tilde{\theta}_{ij} + (\hat{\theta}_{ij} - \tilde{\theta}_{ij}). \quad (\text{A.1})$$

Further, since $I(\zeta_{i-1} \leq x_t < \zeta_i \text{ and } \zeta_{j-1} \leq y_t < \zeta_j) = I(\zeta_{i-1} \leq x_t < \zeta_i) \cdot I(\zeta_{j-1} \leq y_t < \zeta_j)$ and $I(\zeta_{i-1} \leq x_t < \zeta_i) = I(x_t < \zeta_i) - I(x_t < \zeta_{i-1})$, we have

$$\begin{aligned} \tilde{\theta}_{ij} &= \frac{1}{n} \sum_{t=1}^n I(x_t < \zeta_i) I(y_t < \zeta_j) - \frac{1}{n} \sum_{t=1}^n I(x_t < \zeta_i) I(y_t < \zeta_{j-1}) \\ &\quad - \frac{1}{n} \sum_{t=1}^n I(x_t < \zeta_{i-1}) I(y_t < \zeta_j) + \frac{1}{n} \sum_{t=1}^n I(x_t < \zeta_{i-1}) I(y_t < \zeta_{j-1}). \end{aligned} \tag{A.2}$$

Similarly, we can express $\hat{\theta}_{ij}$ as a combination of four terms such as $(1/n) \sum_{t=1}^n I(x_t < \hat{\zeta}_i) I(y_t < \hat{\zeta}_j)$. Thus the second term $(\hat{\theta}_{ij} - \tilde{\theta}_{ij})$ in the right side of (16) can be rewritten as a combination of four terms of difference such as $(1/n) \sum_{t=1}^n I(x_t < \hat{\zeta}_i) I(y_t < \hat{\zeta}_j) - (1/n) \sum_{t=1}^n I(x_t < \zeta_i) I(y_t < \zeta_j)$.

Note that $\sum_{t=1}^n I(x_t < \hat{\zeta}_i) I(y_t < \hat{\zeta}_j)$ is the number of the data values (x_t, y_t) falling below $(\hat{\zeta}_i, \hat{\zeta}_j)$ and $\sum_{t=1}^n I(x_t < \zeta_i) I(y_t < \zeta_j)$ is the number of the data values (x_t, y_t) falling below (ζ_i, ζ_j) . Thus, $(1/n) \sum_{t=1}^n I(x_t < \hat{\zeta}_i) I(y_t < \hat{\zeta}_j) - (1/n) \sum_{t=1}^n I(x_t < \zeta_i) I(y_t < \zeta_j)$ is nothing but the (signed) number of the data values falling between $(\hat{\zeta}_i, \hat{\zeta}_j)$ and (ζ_i, ζ_j) and can be approximated by

$$n[K(\hat{\zeta}_i, \hat{\zeta}_j) - K(\zeta_i, \zeta_j)], \tag{A.3}$$

when n is large. Further, applying the Taylor series expansion, we have

$$K(\hat{\zeta}_i, \hat{\zeta}_j) - K(\zeta_i, \zeta_j) \sim (\hat{\zeta}_i - \zeta_i) \int_0^{\hat{\zeta}_j} k(\zeta_i, y) dy + (\hat{\zeta}_j - \zeta_j) \int_0^{\hat{\zeta}_i} k(x, \zeta_j) dx, \tag{A.4}$$

where $k(x, y)$ is the density function of $K(x, y)$. Thus,

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n I(x_t < \hat{\zeta}_i) I(y_t < \hat{\zeta}_j) - \frac{1}{n} \sum_{t=1}^n I(x_t < \zeta_i) I(y_t < \zeta_j) \\ &\sim (\hat{\zeta}_i - \zeta_i) \int_0^{\hat{\zeta}_j} k(\zeta_i, y) dy + (\hat{\zeta}_j - \zeta_j) \int_0^{\hat{\zeta}_i} k(x, \zeta_j) dx. \end{aligned} \tag{A.5}$$

Similarly, we can approximate the other terms of $(\hat{\theta}_{ij} - \tilde{\theta}_{ij})$ and consequently,

$$\begin{aligned} \hat{\theta}_{ij} &= \tilde{\theta}_{ij} + \left\{ (\hat{\zeta}_i - \zeta_i) \int_0^{\hat{\zeta}_j} k(\zeta_i, y) dy + (\hat{\zeta}_j - \zeta_j) \int_0^{\hat{\zeta}_i} k(x, \zeta_j) dx \right\} \\ &\quad - \left\{ (\hat{\zeta}_i - \zeta_i) \int_0^{\hat{\zeta}_{j-1}} k(\zeta_i, y) dy + (\hat{\zeta}_{j-1} - \zeta_{j-1}) \int_0^{\hat{\zeta}_i} k(x, \zeta_{j-1}) dx \right\} \end{aligned}$$

$$\begin{aligned}
 & - \left\{ (\hat{\zeta}_{i-1} - \zeta_{i-1}) \int_0^{\hat{\zeta}_j} k(\zeta_{i-1}, y) dy + (\hat{\zeta}_j - \zeta_j) \int_0^{\hat{\zeta}_{i-1}} k(x, \zeta_j) dx \right\} \\
 & + \left\{ (\hat{\zeta}_{i-1} - \zeta_{i-1}) \int_0^{\hat{\zeta}_{j-1}} k(\zeta_{i-1}, y) dy + (\hat{\zeta}_{j-1} - \zeta_{j-1}) \int_0^{\hat{\zeta}_{i-1}} k(x, \zeta_{j-1}) dx \right\}.
 \end{aligned}
 \tag{A.6}$$

Or equivalently,

$$\hat{\theta}_{ij} = \tilde{\theta}_{ij} + a_{ij}(\hat{\zeta}_i - \zeta_i) + b_{ij}(\hat{\zeta}_j - \zeta_j) - c_{ij}(\hat{\zeta}_{i-1} - \zeta_{i-1}) - d_{ij}(\hat{\zeta}_{j-1} - \zeta_{j-1}), \tag{A.7}$$

where

$$\begin{aligned}
 a_{ij} &= \int_0^{\hat{\zeta}_j} k(\zeta_i, y) dy - \int_0^{\zeta_{j-1}} k(\zeta_i, y) dy, \\
 b_{ij} &= \int_0^{\hat{\zeta}_i} k(x, \zeta_j) dx - \int_0^{\zeta_{i-1}} k(x, \zeta_j) dx, \\
 c_{ij} &= \int_0^{\hat{\zeta}_j} k(\zeta_{i-1}, y) dy - \int_0^{\zeta_{j-1}} k(\zeta_{i-1}, y) dy, \\
 d_{ij} &= \int_0^{\hat{\zeta}_i} k(x, \zeta_{j-1}) dx - \int_0^{\zeta_{i-1}} k(x, \zeta_{j-1}) dx.
 \end{aligned}
 \tag{A.8}$$

Appendix B

This appendix provides formulae of variances for estimates of the quantile transition matrix, the mean transition matrix and the median transition matrix. The formulae for the size transition matrix are provided in the body of the paper (Eqs. (10) through (13)). To facilitate the inference testing using various transition matrices, we have written SAS and FORTRAN programs to perform the calculation of the covariance matrices. These programs as well as the programs used in Section 5 are available from the authors upon request (and are downloadable from our department website at www.cudenver.edu/econ).

For the quantile transition matrix, the covariance between $n^{1/2}\hat{\theta}_{ij}$ and $n^{1/2}\hat{\theta}_{kl}$ is

$$\begin{aligned}
 & cov(\tilde{\theta}_{ij}, \tilde{\theta}_{kl}) + a_{kl} cov(\tilde{\theta}_{ij}, \hat{\zeta}_k) + b_{kl} cov(\tilde{\theta}_{ij}, \hat{\zeta}_l) + c_{kl} cov(\tilde{\theta}_{ij}, \hat{\zeta}_{k-1}) \\
 & + d_{kl} cov(\tilde{\theta}_{ij}, \hat{\zeta}_{l-1}) + a_{ij} cov(\tilde{\theta}_{kl}, \hat{\zeta}_i) + b_{ij} cov(\tilde{\theta}_{kl}, \hat{\zeta}_j) \\
 & + c_{ij} cov(\tilde{\theta}_{kl}, \hat{\zeta}_{i-1}) + d_{ij} cov(\tilde{\theta}_{kl}, \hat{\zeta}_{j-1}) \\
 & + a_{ij} a_{kl} \frac{\min(s_i, s_k) - s_i s_k}{f(\zeta_i) f(\zeta_k)} + a_{ij} b_{kl} \frac{\min(s_i, s_l) - s_i s_l}{f(\zeta_i) g(\zeta_l)}
 \end{aligned}$$

$$\begin{aligned}
 & - a_{ij}c_{kl} \frac{\min(s_i, s_{k-1}) - s_i s_{k-1}}{f(\zeta_i)f(\zeta_{k-1})} - a_{ij}d_{kl} \frac{\min(s_i, s_{l-1}) - s_i s_{l-1}}{f(\zeta_i)g(\zeta_{l-1})} \\
 & + b_{ij}a_{kl} \frac{\min(s_j, s_k) - s_j s_k}{g(\zeta_j)f(\zeta_k)} + b_{ij}b_{kl} \frac{\min(s_j, s_l) - s_j s_l}{g(\zeta_j)g(\zeta_l)} \\
 & - b_{ij}c_{kl} \frac{\min(s_j, s_{k-1}) - s_j s_{k-1}}{g(\zeta_j)f(\zeta_{k-1})} - b_{ij}d_{kl} \frac{\min(s_j, s_{l-1}) - s_j s_{l-1}}{g(\zeta_j)g(\zeta_{l-1})} \\
 & - c_{ij}a_{kl} \frac{\min(s_{i-1}, s_k) - s_{i-1} s_k}{f(\zeta_{i-1})f(\zeta_k)} - c_{ij}b_{kl} \frac{\min(s_{i-1}, s_l) - s_{i-1} s_l}{f(\zeta_{i-1})g(\zeta_l)} \\
 & + c_{ij}c_{kl} \frac{\min(s_{i-1}, s_{k-1}) - s_{i-1} s_{k-1}}{f(\zeta_{i-1})f(\zeta_{k-1})} + c_{ij}d_{kl} \frac{\min(s_{i-1}, s_{l-1}) - s_{i-1} s_{l-1}}{f(\zeta_{i-1})g(\zeta_{l-1})} \\
 & - d_{ij}a_{kl} \frac{\min(s_{j-1}, s_k) - s_{j-1} s_k}{g(\zeta_{j-1})f(\zeta_k)} - d_{ij}b_{kl} \frac{\min(s_{j-1}, s_l) - s_{j-1} s_l}{g(\zeta_{j-1})g(\zeta_l)} \\
 & + d_{ij}c_{kl} \frac{\min(s_{j-1}, s_{k-1}) - s_{j-1} s_{k-1}}{g(\zeta_{j-1})f(\zeta_{k-1})} + d_{ij}d_{kl} \frac{\min(s_{j-1}, s_{l-1}) - s_{j-1} s_{l-1}}{g(\zeta_{j-1})g(\zeta_{l-1})}, \tag{B.1}
 \end{aligned}$$

where $s_i = \sum_{j=1}^i \pi_j$,

$$cov(\tilde{\theta}_{ij}, \tilde{\theta}_{kl}) = \begin{cases} \theta_{ij}(1 - \theta_{ij}) & \text{if } i = k \text{ and } j = l, \\ -\theta_{ij}\theta_{kl} & \text{else,} \end{cases} \tag{B.2}$$

$$cov(\tilde{\theta}_{ij}, \hat{\zeta}_k) = \begin{cases} -\frac{a_{kl}\theta_{ij}(1 - s_k)}{f(\zeta_k)} & \text{if } i \leq k, \\ \frac{a_{kl}\theta_{ij}s_k}{f(\zeta_k)} & \text{else,} \end{cases} \tag{B.3}$$

$$cov(\tilde{\theta}_{ij}, \hat{\zeta}_l) = \begin{cases} -\frac{b_{kl}\theta_{ij}(1 - s_l)}{g(\zeta_l)} & \text{if } i \leq l, \\ \frac{b_{kl}\theta_{ij}s_l}{g(\zeta_l)} & \text{else.} \end{cases} \tag{B.4}$$

In particular, the asymptotic variance of $n^{1/2}\hat{\theta}_{ij}$ is

$$\begin{aligned}
 & \theta_{ij}(1 - \theta_{ij}) + \frac{a_{ij}^2 s_i(1 - s_i)}{f^2(\zeta_i)} + \frac{b_{ij}^2 s_j(1 - s_j)}{g^2(\zeta_j)} + \frac{c_{ij}^2 s_{i-1}(1 - s_{i-1})}{f^2(\zeta_{i-1})} \\
 & + \frac{d_{ij}^2 s_{j-1}(1 - s_{j-1})}{g^2(\zeta_{j-1})} - \frac{2a_{ij}\theta_{ij}(1 - s_i)}{f(\zeta_i)} - \frac{2b_{ij}\theta_{ij}(1 - s_j)}{g(\zeta_j)} + \frac{2a_{ij}b_{ij}(s_{ij} - s_i s_j)}{f(\zeta_i)g(\zeta_j)} \\
 & - \frac{2a_{ij}c_{ij}s_{i-1}(1 - s_i)}{f(\zeta_i)f(\zeta_{i-1})} - \frac{2a_{ij}d_{ij}(s_{i(j-1)} - s_i s_{j-1})}{f(\zeta_i)g(\zeta_{j-1})} - \frac{2b_{ij}c_{ij}(s_{(i-j)j} - s_{i-1} s_j)}{f(\zeta_{i-1})g(\zeta_j)} \\
 & - \frac{2b_{ij}d_{ij}s_{j-1}(1 - s_j)}{g(\zeta_{j-1})g(\zeta_j)} + \frac{2c_{ij}d_{ij}(s_{(i-1)(j-1)} - s_{i-1} s_{j-1})}{f(\zeta_{i-1})g(\zeta_{j-1})}, \tag{B.5}
 \end{aligned}$$

where $s_{ij} = \Pr(x < \zeta_i \text{ and } y < \xi_j)$. The asymptotic variance of $n^{1/2} \hat{p}_{ij}$ is simply the asymptotic variance of $n^{1/2} \hat{\theta}_{ij}$ divided by π_i^2 .

For the mean transition matrix, we need to first estimate the asymptotic covariance of $(\hat{\pi}, \hat{\theta})$. Denote this covariance matrix as A , the asymptotic covariance matrix of \hat{p} is $\Gamma A \Gamma'$ where $\Gamma = [\partial p / \partial (\pi, \theta)]$ which is an $mm \times (m + 1)m$ matrix. Matrix A contains covariances between $\hat{\theta}_{ij}$ and $\hat{\theta}_{kl}$, between $\hat{\pi}_i$ and $\hat{\pi}_j$, and between $\hat{\pi}_i$ and $\hat{\theta}_{kl}$. The covariance between $n^{1/2} \hat{\theta}_{ij}$ and $n^{1/2} \hat{\theta}_{kl}$ is

$$\begin{aligned} \epsilon(i, j, k, l) + \bar{a}_{kl}(\mu_x^{ij} - \theta_{ij}\mu_x) + \bar{b}_{kl}(\mu_y^{ij} - \theta_{ij}\mu_y) + \bar{a}_{ij}(\mu_x^{kl} - \theta_{kl}\mu_x) + \bar{a}_{ij}\bar{a}_{kl}\sigma_x^2 \\ + \bar{a}_{ij}\bar{b}_{kl}\rho_{xy} + \bar{b}_{ij}(\mu_y^{kl} - \theta_{kl}\mu_y) + \bar{a}_{kl}\bar{b}_{ij}\rho_{xy} + \bar{b}_{ij}\bar{b}_{kl}\sigma_y^2, \end{aligned} \tag{B.6}$$

where $\epsilon(i, j, k, l) = \theta_{ij}(1 - \theta_{ij})$ if $i = j$ and $k = l$, and $-\theta_{ij}\theta_{kl}$ otherwise, $\bar{a}_{ij} = \alpha_i a_{ij} - \alpha_{i-1} c_{ij}$ and $\bar{b}_{ij} = \alpha_i b_{ij} - \alpha_{i-1} d_{ij}$, $\mu_x^{ij} = \int_{\zeta_{i-1}}^{\zeta_i} \int_{\xi_{j-1}}^{\xi_j} x \, dK(x, y)$, $\mu_y^{ij} = \int_{\zeta_{i-1}}^{\zeta_i} \int_{\xi_{j-1}}^{\xi_j} y \, dK(x, y)$, σ_x^2 and σ_y^2 are variances of x and y and ρ_{xy} is covariance between x and y .

The asymptotic covariance between $n^{1/2} \hat{\pi}_i$ and $n^{1/2} \hat{\pi}_j$ is

$$\epsilon(i, j) + h_j(\mu_x^i - \pi_i\mu_x) + h_i(\mu_x^j - \pi_j\mu_x) + h_i h_j \sigma_x^2, \tag{B.7}$$

where $\epsilon(i, j) = \pi_i(1 - \pi_i)$ if $i = j$ and $-\pi_i\pi_j$ otherwise, $h_i = \alpha_i f(\alpha_i\mu_x) - \alpha_{i-1} f(\alpha_{i-1}\mu_x)$ and $\mu_x^i = \int_{\zeta_{i-1}}^{\zeta_i} x \, dF(x)$.

The asymptotic covariance between $n^{1/2} \hat{\pi}_i$ and $n^{1/2} \hat{\theta}_{kl}$ is

$$\begin{aligned} v(i, k, l) + \bar{a}_{kl}(\mu_x^i - \pi_i\mu_x) + \bar{b}_{kl}(\mu_y^i - \pi_i\mu_y) + h_i(\mu_x^{kl} - \theta_{kl}\mu_x) \\ + h_i \bar{a}_{kl} \sigma_x^2 + h_i \bar{b}_{kl} \rho_{xy}, \end{aligned} \tag{B.8}$$

where $v(i, k, l) = \theta_{kl}(1 - \pi_k)$ if $i = k$ and $-\pi_i\theta_{kl}$ otherwise, and $\mu_y^i = \int_{\zeta_{i-1}}^{\zeta_i} \int_0^\infty y \, dK(x, y)$.

For the median transition matrix, we also need to first estimate the asymptotic covariance of $(\hat{\pi}, \hat{\theta})$. Denote this covariance matrix as Δ , the asymptotic covariance matrix of \hat{p} is $\Gamma \Delta \Gamma'$ where $\Gamma = [\partial p / \partial (\pi, \theta)]$ which is an $mm \times (m + 1)m$ matrix. Matrix Δ contains covariance between $\hat{\theta}_{ij}$ and $\hat{\theta}_{kl}$, between $\hat{\pi}_i$ and $\hat{\pi}_j$, and between $\hat{\pi}_i$ and $\hat{\theta}_{kl}$. The asymptotic covariance between $n^{1/2} \hat{\theta}_{ij}$ and $n^{1/2} \hat{\theta}_{kl}$ is

$$\begin{aligned} \epsilon_d(i, j, k, l) - \frac{\bar{a}_{kl}(\epsilon_x^{ij} - \frac{1}{2} \theta_{ij})}{f(\tau_x)} - \frac{\bar{b}_{kl}(\epsilon_y^{ij} - \frac{1}{2} \theta_{ij})}{g(\tau_y)} - \frac{\bar{a}_{ij}(\epsilon_x^{kl} - \frac{1}{2} \theta_{kl})}{f(\tau_x)} + \frac{\frac{1}{4} \bar{a}_{ij} \bar{a}_{kl}}{f^2(\tau_x)} \\ + \frac{\bar{a}_{ij} \bar{b}_{kl} \epsilon_{xy}}{f(\tau_x)g(\tau_y)} - \frac{\bar{b}_{ij}(\epsilon_y^{kl} - \frac{1}{2} \theta_{kl})}{g(\tau_y)} + \frac{\bar{a}_{kl} \bar{b}_{ij} \epsilon_{xy}}{f(\tau_x)g(\tau_y)} + \frac{\frac{1}{4} \bar{b}_{ij} \bar{b}_{kl}}{g^2(\tau_y)}, \end{aligned} \tag{B.9}$$

where $\epsilon(i, j, k, l) = \theta_{ij}(1 - \theta_{ij})$ if $i = j$ and $k = l$, and $-\theta_{ij}\theta_{kl}$ otherwise, $\bar{a}_{ij} = \alpha_i a_{ij} - \alpha_{i-1} c_{ij}$ and $\bar{b}_{ij} = \alpha_i b_{ij} - \alpha_{i-1} d_{ij}$, $\epsilon_x^{ij} = \int_{\zeta_{i-1}}^{\zeta_i} \int_{\xi_{j-1}}^{\xi_j} I(x < \tau_x) \, dK(x, y)$, $\epsilon_y^{ij} = \int_{\zeta_{i-1}}^{\zeta_i} \int_{\xi_{j-1}}^{\xi_j} I(y < \tau_y) \, dK(x, y)$, and $\epsilon_{xy} = \int_0^{\tau_x} \int_0^{\tau_y} I(x < \tau_x) I(y < \tau_y) \, dK(x, y) - \frac{1}{4}$.

The asymptotic covariance between $n^{1/2}\hat{\pi}_i$ and $n^{1/2}\hat{\pi}_j$ is

$$\varepsilon_d(i, j) - \frac{h_j^d(\varepsilon_x^i - \frac{1}{2}\pi_i)}{f(\tau_x)} - \frac{h_i^d(\varepsilon_x^j - \frac{1}{2}\pi_j)}{f(\tau_x)} + \frac{\frac{1}{4}h_i^d h_j^d}{f^2(\tau_x)}, \tag{B.10}$$

where $\varepsilon_d(i, j) = \pi_i(1 - \pi_i)$ if $i = j$ and $-\pi_i\pi_j$ otherwise, $h_i^d = \alpha_i f(\alpha_i \tau_x) - \alpha_{i-1} f(\alpha_{i-1} \tau_x)$ and $\varepsilon_x^i = \int_{\xi_{i-1}}^{\xi_i} I(x < \tau_x) dF(x)$.

The asymptotic covariance between $n^{1/2}\hat{\pi}_i$ and $n^{1/2}\hat{\theta}_{kl}$ is

$$\begin{aligned} v_d(i, k, l) - \frac{\bar{a}_{kl}(\varepsilon_x^i - \frac{1}{2}\pi_i)}{f(\tau_x)} - \frac{\bar{b}_{kl}(\varepsilon_y^i - \frac{1}{2}\pi_i)}{g(\tau_y)} - \frac{h_i^d(\varepsilon_x^{kl} - \frac{1}{2}\theta_{kl})}{f(\tau_x)} \\ + \frac{\frac{1}{4}h_i^d \bar{a}_{kl}}{f^2(\tau_x)} + \frac{h_i^d \bar{b}_{kl} \varepsilon_{xy}}{f(\tau_x)g(\tau_y)}, \end{aligned} \tag{B.11}$$

where $v_d(i, k, l) = \theta_{kl}(1 - \pi_k)$ if $i = k$ and $-\pi_i\theta_{kl}$ otherwise, and $\varepsilon_y^i = \int_{\xi_{i-1}}^{\xi_i} \int_0^\infty I(y < \tau_y) dK(x, y)$.

References

Atkinson, A.B., 1983. The measurement of economic mobility. In: *Social Justice and Public Policy*, MIT Press, Cambridge, MA.

Atkinson, A.B., Bourguignon, F., Morrison, C., 1992. *Empirical Studies of Earnings Mobility*. Harwood Academic Publishers, Chur.

Bahadur, R., 1966. A note on quantiles in large samples. *Annals of Mathematical Statistics* 37, 577–580.

Bartholomew, D.J., 1982. *Stochastic Models for Social Processes*, 3rd ed. Wiley, London.

Bartholomew, D.J., 1996. *The Statistical Approach to Social Measurement*. Academic Press, San Diego.

Bénabou, R., Ok, E.A., 1999. Mobility as progressivity: ranking income process according to equality of opportunity. mimeo, New York University.

Biewen, M., 2002. Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics* 108, 317–342.

Bishop, J.A., Formby, J.P., Thistle, P., 1992. The convergence of south and non-south income distribution, 1969–1979. *American Economic Review* 82, 262–272.

Burkhauser, R.V., Holtz-Eakin, D., 1993. Changes in the distribution of wage earnings in the United States and Germany during the 1980s. In: Burkhauser, R.V., Wagner, G.G. (Eds.), *Proceedings of the 1993 International Conference of German Socio-economic Panel study Users (Vierteljahresheft zur Wirtschaftsforschung)*, pp. 27–35.

Burkhauser, R.V., Poupore, J.G., 1997. A cross-national comparison of permanent inequality in the United States and Germany. *Review of Economics and Statistics* 79, 10–17.

Burkhauser, R.V., Burtrica, B.A., Daly, M.C., 1995. The Syracuse University PSID-GSOEP equivalent data file: a product of cross-national research, cross-national studies in aging program policy project paper 26. The Maxwell School, Syracuse, New York.

Burkhauser, R.V., Holtz-Eakin, D., Rhody, S., 1997. Labor earnings mobility and inequality in the United States and Germany during the growth years of the 1980s. *International Economic Review* 38, 775–794.

Burkhauser, R.V., Holtz-Eakin, D., Rhody, S., 1998. Mobility and inequality in the 1980s: a cross-national comparison of the United States and Germany. In: Jenkins, S.P., Kapteyn, A., Van Praag, B. (Eds.), *The Distribution of Welfare and Household Production: International Perspectives*. Cambridge University Press, Cambridge.

Chakravarty, S.J., Dutta, B., Weymark, J.A., 1985. Ethical indices of income mobility. *Social Choice and Welfare* 2, 1–21.

Christensen, R., 1990. *Long-linear Models*. Springer, New York.

- Dardanoni, V., 1993. On measuring social mobility. *Journal of Economic Theory* 61, 372–394.
- Fields, G., Ok, E.A., 1996. The meaning and measurement of income mobility. *Journal of Economic Theory* 71, 349–377.
- Fields, G., Ok, E.A., 1999. The measurement of income mobility: an introduction to the literature. In: Silber, J. (Ed.), *Handbook of Inequality Measurement*. Kluwer Academic Publishers, Dordrecht.
- Fisher, G., Willson, D., Xu, K., 1998. A empirical analysis of term premiums using significance tests for stochastic dominance. *Economics Letters* 60, 195–203.
- Formby, J.P., Smith, W.J., Zheng, B., 2003. Economic growth, welfare and the measurement of social mobility. In: Amiel Y., Bishop, J. (Eds.), *Research in Economic Inequality*, Vol. 9, JAI Press, Amsterdam.
- Ghosh, J.K., 1971. A new proof of the Bahadur representation of quantiles and an application. *Annals of Mathematical Statistics* 42, 1957–1961.
- Ginsberg, M., 1929. Interchange between social classes. *Economic Journal* 39, 554–565.
- Hart, P.E., 1976a. The dynamics of earnings, 1963–1973. *Economic Journal* 86, 551–565.
- Hart, P.E., 1976b. The comparative statics and dynamics of income distributions. *Journal of the Royal Statistical Society* 139, 108–125.
- Hart, P.E., 1983. The size mobility of earnings. *Oxford Bulletin of Economics and Statistics* 45, 181–193.
- Hill, M.S., 1992. *The Panel Study of Income Dynamics: a User's Guide*. Sage Publications, Beverley Hills, CA.
- Howes, S., 1994. Testing for dominance: inferring population rankings from sample data, World Bank Discussion Paper.
- Jarvis, S., Jenkins, S., 1998. How much income mobility is there in Britain? *Economic Journal* 108, 428–443.
- Kodde, D.A., Palm, F.C., 1986. Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* 50, 1243–1248.
- Kuznets, S.S., 1966. *Modern Economic Growth: Rate, Structure, and Spread*. Yale University, New Haven.
- Maasoumi, E., 1997. On mobility. In: Giles, D., Ullah, A. (Eds.), *The Handbook of Economic Statistics*. Marcel Dekker, New York.
- Maasoumi, E., Trede, M., 2001. Comparing income mobility in Germany and the United States using generalized entropy mobility measures. *Review of Economics and Statistics* 83, 551–559.
- Maasoumi, E., Zandvakili, S., 1986. A class of generalized measures of mobility with applications. *Economics Letters* 22, 97–102.
- McCall, J.J., 1973. *Income Mobility, Racial Discrimination and Economic Growth*. D.C. Heath, Lexington.
- Prais, S.J., 1955. Measuring social mobility. *Journal of the Royal Statistical Society Series A, Part I*, 118, 56–66.
- Rosen, H.S., 1985. *Public Finance*. Irwin, Homewood, IL.
- Schluter, C., 1996. Income mobility in Germany: evidence from panel data. DARP Discussion Paper 17, London School of Economics.
- Schluter, C., 1997. On the non-stationarity of German income mobility. DARP Discussion Paper 30, London School of Economics.
- Schluter, C., 1998. Statistical inference with mobility indices. *Economics Letters* 59, 157–162.
- Serfling, R.J., 1980. *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- Shorrocks, A.F., 1978a. The measurement of mobility. *Econometrica* 46, 1013–1024.
- Shorrocks, A.F., 1978b. Income inequality and income mobility. *Journal of Economic Theory* 19, 376–393.
- Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.
- Solow, R., 1951. On the dynamics of the income distribution, Ph.D. thesis, Harvard University.
- Sommers, P.S., Conlisk, J., 1979. Eigenvalue immobility measures for Markov chains. *Journal of Mathematical Sociology* 6, 253–276.
- Thatcher, A.R., 1971. Year-to-year variations in the earnings of individuals. *Journal of the Royal Statistical Society* 134, 374–382.
- Trede, M., 1998. Making mobility visible: a graphical device. *Economics Letters* 59, 77–82.
- Trede, M., 1999. Statistical inference for measures of income mobility. *Jahrbücher für Nationalökonomie und Statistik* 218, 473–490.
- Wolak, F.A., 1989. Testing inequality constraints in linear econometric models. *Journal of Econometrics* 41, 205–235.